Essays on the Econometric Modelling of Consumer Preferences for Tourism

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

David Boto García

Department of Economics

Programa de Doctorado

Economía: Instrumentos del Análisis Económico

Oviedo, 2020

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: *Ensayos sobre la modelización econométrica de las preferencias turísticas del consumidor*	Inglés: *Essays on the econometric modelling of consumer preferences for tourism*

2.- Autor	
Nombre: DAVID BOTO GARCÍA	DNI:
Programa de Doctorado: ECONOMÍA: INSTRUMENTOS DEL ANÁLISIS ECONÓMICO	
Órgano responsable: UNIVERSIDAD DE OVIEDO	

RESUMEN (en español)

Dada la gran relevancia del sector turístico en la economía mundial, y en España en particular, el objetivo de la tesis es estudiar las preferencias de los individuos por las actividades turísticas. Aunque existen numerosas investigaciones en el campo del turismo sobre las decisiones de los individuos, los factores que explican las preferencias turísticas no han sido tan estudiados desde una perspectiva microeconómica. En este sentido, la tesis pone el foco en la modelización económica de las decisiones turísticas y en la heterogeneidad en preferencias. Por lo tanto, el principal objetivo es ofrecer una caracterización *microeconométrica* que esté fundamentada en la teoría económica.

La tesis se estructura en tres capítulos o ensayos. Los dos primeros utilizan datos de preferencias reveladas obtenidas a partir de encuestas. En el capítulo 1 se estudian los determinantes de la duración de la estancia de los turistas. Más concretamente, se modeliza i) la decisión de pernoctar (frente a ser un excursionista), y ii) la duración de la estancia de aquellos catalogados como turistas. Para ello, se hace uso de una amplia base de datos de más de 19.000 individuos obtenida a partir de encuestas realizadas durante el periodo 2010-2016 en Asturias (España). Se estima un modelo de conteo tipo valla. Se comparan dos distribuciones alternativas de la heterogeneidad inobservada (gamma y lognormal).

En el capítulo 2 se modeliza el papel de los atributos de las regiones en la elección de destino por parte de los turistas. Se analizan viajes turísticos por motivo naturaleza o deporte usando datos mensuales de alrededor de 7.000 individuos entre febrero de 2015 y septiembre de 2017. Los datos proceden del Instituto Nacional de Estadística. Se estudia el efecto de la distancia al origen, el diferencial de temperaturas, los precios, los kilómetros disponibles para esquiar, la superficie de áreas naturales protegidas y la presencia de costa, entre otras. Se estima un modelo de parámetros aleatorios con componentes de error que controla por heterogeneidad inobservada a nivel de individuo y de región. Se permite que las utilidades marginales vengan explicadas por un vector de características del individuo y se calculan relaciones marginales de sustitución y elasticidades.

Los datos de preferencias reveladas tienen la ventaja de que reflejan las elecciones turísticas realizadas. Sin embargo, este tipo de datos tiene la limitación de i) limitar el análisis a las variables que se hayan recogido en la encuesta, y ii) solo proporciona información de las decisiones realizadas, pero no de las alternativas consideradas. Con el objetivo de estudiar las preferencias de los individuos por diferentes conjuntos de bienes turísticos, en el capítulo 3 se lleva a cabo un experimento de elección discreta. Este experimento permite identificar las preferencias individuales en un ambiente controlado e incentivado. Un total de 262 individuos participaron en el estudio. Se estima un modelo Multinomial Logit de clases latentes que identifica diferentes grupos de individuos con diferentes gustos. Seguidamente, se derivan relaciones marginales de sustitución en la forma de 'disponibilidad a pagar' por los atributos que caracterizan las alternativas.

RESUMEN (en Inglés)

Given the great relevance of the tourism sector in the world and in Spain in particular, the purpose of the thesis is to study individual preferences for tourism-related activities. Although there is ample research about tourist decisions, there is a lack of understanding of the drivers of tourism-related preferences from a microeconomic viewpoint. In this regard, the thesis places emphasis on the economic modelling of tourist choices and taste heterogeneity. Therefore, the main aim is to provide a microeconometric theoretically-consistent characterization of individual preferences for tourism.

The thesis is structured in three chapters or essays. The first two make use of revealed preferences obtained from survey data. In Chapter 1, I study the determinants of tourists' length of stay. Specifically, I model i) the decision to stay overnight at a tourist destination (versus being a same-day visitor) and ii) how long to stay conditional on being a tourist. I make use of a rich and unexploited dataset of more than 19,000 individuals obtained from surveys during the period 2010-2016 in Asturias (Spain). I estimate a Hurdle Zero Truncated Count Data model. In doing so, two alternative distributions for unobserved heterogeneity (gamma and lognormal) are compared.

In Chapter 2, I model the role of place-based attributes on tourist choice of destination. I analyse nature-based recreational trips within Spain, using monthly data for almost 7,000 individuals between February 2015 and September 2017 from the Spanish National Statistics Institute. I study the effect of distance to origin, relative temperatures, prices, kilometres available for ski, the protected natural surface and the presence of coast, among others. A Random Parameter Logit with Error Components model, which controls for unobserved heterogeneity at the individual and the regional level, is estimated. The marginal utilities are allowed to depend on a vector of taste shifters, and marginal rates of substitution and elasticities are computed and interpreted.

Data from revealed preferences surveys have the advantage of reflecting actual tourism-related choices. However, this approach has the drawback of i) limiting the scope of the analysis to the variables gathered in the survey, and ii) only providing information on the decisions finally made but not on the considered alternatives. For studying individual preferences for alternative tourism bundles of goods, in Chapter 3 I conduct a Discrete Choice Experiment. This allows me to elicit individual preferences in an incentive compatible controlled setting. A total of 262 individuals participated in the experiment. I estimate a Latent Class Multinomial Logit by which different segments with different tastes are identified. I subsequently derive Marginal Rates of Substitution in the form of Willingness to Pay estimates.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN ECONOMIA: INSTRUMENTOS DEL ANÁLISIS ECONÓMICO

Essays on the Econometric Modelling of Consumer Preferences for Tourism

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

David Boto García

Department of Economics

Thesis advisors

José Baños and Antonio Alvarez

Oviedo, 2020

AGRADECIMIENTOS

A mis padres y abuelos, por enseñarme que el único camino para conseguir tus objetivos es el trabajo y la constancia, y por haberme apoyado en todo momento.

A José Baños, por tu infinita paciencia conmigo; por todas las horas compartidas; por tus buenos consejos; por saber pararme y encauzarme cuando me iba por las ramas; por tu apoyo cuando las cosas no salían bien; por tus aportaciones técnicas; y por transmitirme el querer hacer las cosas siempre lo mejor posible. Por todo el camino andado desde el TFG en el año 2015, y por todos los proyectos que están por venir. Gracias, Jose.

A Antonio Alvarez, por haberme inspirado el rigor, la autoexigencia, y la importancia de dominar la econometría para un correcto análisis; por tus gestiones para ponerme en contacto con otros investigadores o datos; y por tu tiempo y esfuerzo en ayudarme a, seguramente más lentamente de lo deseado, mejorar en mi forma de trabajar y de contar las cosas. Gracias, Antonio.

A Clara, porque qué bonito es pintar contigo el paraíso. Te quiero.

A mis amigos y amigas, sabéis quiénes sois, por haber estado ahí siempre, en los días buenos pero sobre todo en los malos.

A Marta y a Emma, por todas esas horas en el seminario 3 de trabajo y conversaciones, con sus días mejores y sus días peores, en las que sin lugar a dudas se generaron muchas externalidades positivas. El camino habría sido más difícil sin vosotras.

A todas las personas con las que he trabajado o estoy trabajando en algún artículo, porque con todos vosotros he aprendido cosas y he pasado grandes momentos intercambiando ideas o impresiones: José Baños, Antonio Alvarez, Marta Escalonilla, Emma Zapico, Alessandro Bucciol, Federico Perali, Petr Mariel, Matías Mayor, Sara Suárez, Pablo de la Vega, Juan Prieto, Luca, Zarri, Martina Manfrè, Álvaro Muñiz y Carlos Varela.

A Luis Valdés, Eduardo del Valle e Inés Sustacha, por haberme cedido los microdatos del SITA usados en el capítulo 1 y haber resuelto todas las dudas que tenía respecto a la recogida de datos o la definición de las variables; y por haberme aconsejado en el diseño experimental del capítulo 3.

A Petr Mariel, por haberme invitado a diferentes cursos y seminarios en Bilbao sobre discrete choice modelling; por haberme enseñado a trabajar en R; por tu tiempo y paciencia; por haberme enseñado a hacer un diseño experimental y la importancia de las simulaciones y el análisis de los datos; y por tus consejos y comentarios siempre acertados.

A Federico Perali, por invitarme a visitar en dos ocasiones la Universidad de Verona y haberme acogido con tanto afecto; por haberme ayudado con el diseño del cuestionario; por todas las conversaciones sobre la toma de decisiones dentro del hogar; y por tu trato siempre cercano.

A Alessandro Bucciol, por haber visto algo en aquel estudiante español que un día se presentó en tu despacho para haberme permitido poder trabajar contigo; por tu carácter afable y cercano; y por todo lo que he aprendido.

A Martina Menon, Vittoria Levatti, Chiara Nardi, Luca Zarri, Simone Quercia, Josepa Miquel-Florensa, Jaume Rosselló, Juan Luís Eugenio y Beatriz Tovar, por haber dedicado parte de su tiempo a atenderme, a resolver algunas dudas; por darme sus impresiones acerca de versiones preliminares de los diferentes capítulos de la tesis; y por haberme aconsejado sobre errores habituales en la realización de experimentos.

A Leví Pérez, Esteban García Canal, Rosa Rodríguez y Fermín Rodríguez, por haber colaborado desinteresadamente en la cesión de aulas y espacios para la realización de las sesiones experimentales del capítulo 3.

Al Ministerio de Educación, por haberme concedido una beca FPU para financiar estos cuatro años de doctorado y sin la cual todo habría sido mucho más difícil.

Finalmente, a los miembros del Departamento de Economía de la Universidad de Oviedo, por todos los comentarios, críticas y sugerencias realizadas en los diferentes seminarios que presenté, que han servido para mejorar el contenido de la tesis.

ACKNOWLEDGEMENTS

To my parents and my grandparents, for teaching me that the only straight way for reaching your goals is hard work, effort, and constancy, and for their support during the whole process.

To José Baños, for your huge patience with me; for all the hours spent together; for your great advices; for knowing how to stop me when I beat about the bush, for your support when things do not go as desired, for your teachings, for encouraging me to do always my best. For all the walking path since 2015, and for all the incoming projects. Thank you, Jose.

To Antonio Alvarez, for having inspired me rigor, self-commitment, good practises and the importance of econometrics for a proper economic analysis; for your efforts to put me in touch with other researchers or get new data; for the time spent in helping me to, possibly slower than desired, improve my way of doing research and writing. Thank you, Antonio.

To Clara. I love you.

To all my friends, you know who you are, for having always been there, during happy days but especially during sad days.

To Marta and Emma, for all the time spent in seminario 3 working and talking, with our happy and sad days, but in which for sure we generated many positive externalities. The way would have been harder without you ladies.

To all the people with whom I have worked during this years in different projects, because I have learned from all of you, and I have had really good time interchanging ideas and doing research: José Baños, Antonio Alvarez, Marta Escalonilla, Emma Zapico, Alessandro Bucciol, Federico Perali, Petr Mariel, Matías Mayor, Sara Suárez, Pablo de la Vega, Juan Prieto, Luca Zarri, Martina Manfrè, Álvaro Muñiz and Carlos Varela.

To Luis Valdés, Eduardo del Valle and Inés Sustacha, for having shared with me the microdata from the Tourist Information System of Asturias used in chapter 1; for having solved all my doubts about the data; and for your suggestions regarding the experimental design for chapter 3.

To Petr Mariel, for having invited me to different courses, seminars and talks in Bilbao about discrete choice modelling; for having taught me how to work in R; for your time and patience; for your help and support with the experimental design; for inspiring me the importance of simulations and data analysis; and for all your always useful comments and suggestions.

To Federico Perali, for having invited me to visit twice the University of Verona; for your friendly and supportive welcome; for having helped me so much with the questionnaire design; for all our talks about household economics.

To Alessandro Bucciol, for having seen something in that Spanish PhD candidate that one day appeared at your office that encouraged you to accept me to work with you; for your kind and friendly character; for all I have learned working with you.

To Martina Menon, Vittoria Levatti, Chiara Nardi, Luca Zarri, Simone Quercia, Josepa Miquel-Florensa, Jaume Rosselló, Juan Luís Eugenio and Beatriz Tovar, for *having spent part of your time with me, solving some of my queries, giving me some feedback about preliminary versions of the essays, and for having instructed me about usual mistakes to avoid when conducting economic experiments.*

To Leví Pérez, Esteban García Canal, Rosa Rodríguez and Fermín Rodríguez, for your selfless collaboration in allowing me to make use of rooms for the experimental sessions of chapter 3 in the different towns the experiment was conducted.

To the Spanish Ministry of Education, for having conceded me a scholarship to finance these four years and without which everything would have been much difficult.

Finally, to all the members of the Department of Economics at the University of Oviedo, for all the comments and suggestions raised in the different seminars I presented during these years, which have helped to improve the content of the thesis.

CONTENTS

INTRODUCCION

El turismo es actualmente un sector económico que no deja de crecer. En el año 2010, hubo un total de 950 millones de llegadas de turistas en el mundo. En 2018, esta cifra se incrementó hasta los 1.400 millones de llegadas. Se estima que el sector turístico supone un 10,4% del PIB mundial y genera 319 millones de puestos de trabajo. Durante los últimos ocho años, este sector ha crecido a mayor ritmo que la economía global (UWNTO, 2019), siendo el segundo sector más importante en la economía mundial tras las manufacturas en 2018. España se sitúa entre los destinos turísticos más importantes del mundo, siendo el segundo país más visitado. De acuerdo con las estimaciones del Instituto Nacional de Estadística (INE, 2019), España recibió un total de 82,6 millones de llegadas de turistas en 2018, lo que supone un 11,7% del PIB y un 12,8% del empleo total.

A pesar de la relevancia del turismo internacional, el turismo interno ha incrementado su importancia en España en la última década. En el año 2018 los flujos domésticos representaron dos tercios de la demanda agregada. Se realizaron más de 100 millones de viajes dentro de España por motivo ocio o vacacional, lo que representa un gasto agregado de más de 28 mil millones de euros. Sin embargo, existen pocos estudios que analicen los factores que incitan a los individuos a viajar a una región o a otra. En este sentido, los viajes internos han sido menos analizados que los viajes internacionales.

Existe una amplia evidencia empírica que muestra que el turístico es un motor del crecimiento económico, tanto para los países desarrollados como para los países en desarrollo (Lee y Chang, 2008; Faber y Gaubert, 2019). Para el caso español, algunos estudios muestran que existe una relación estable de largo plazo entre el crecimiento económico y la actividad turística (Capó-Parrilla et al., 2007), lo que da lugar a importantes efectos multiplicadores en el largo plazo.

Dada la creciente relevancia económica del sector turístico en el mundo y en España en particular, resulta necesario estudiar las preferencias turísticas de los consumidores. Aunque existe bastante evidencia empírica sobre las decisiones de los turistas, es preciso ahondar en el análisis de las preferencias de los consumidores por el turismo desde una perspectiva microeconómica. En este sentido, la tesis pone el énfasis en la modelización de las decisiones turísticas desde una perspectiva económica. Por lo tanto, el principal objetivo es ofrecer un análisis *microeconométrico* que sea consistente con la teoría económica.

Numerosas investigaciones han puesto el foco en el estudio de los flujos turísticos agregados. Sin embargo, está ampliamente reconocido en economía que los datos a nivel individual ofrecen una mejor caracterización del comportamiento del consumidor, ya que explotan la variabilidad de sección cruzada y son menos sensibles a sesgos de agregación (Blundell et al., 1993; McGuckin, 1995). Por esta razón, la tesis hace uso de datos individuales para ofrecer nueva evidencia empírica acerca de por qué los individuos toman diferentes decisiones. En el análisis se presta especial atención a la modelización econométrica de la heterogeneidad en preferencias.

El marco teórico se construye a partir del modelo de demanda basado en características de Lancaster (Lancaster, 1966). Esta teórica plantea que son las características de los bienes lo que produce utilidad. De esta manera, los consumidores escogen entre conjuntos de bienes basándose en sus características. Bajo el supuesto habitual de maximización de utilidad, la elección entre conjuntos de bienes se usa para inferir las preferencias por las características. Sin embargo, los individuos tienen distintas preferencias por esas características. Dicho de otro modo, las utilidades marginales son heterogéneas en la población. Parte de esta heterogeneidad puede vincularse a las características de los individuos, como la edad o la renta. En este sentido, el uso de factores socioeconómicos como fuente de heterogeneidad en el gusto por los bienes tiene una larga tradición en la teoría económica (Pollak and Wales, 1981). Otra fuente importante de esta heterogeneidad es inobservada y específica de cada individuo. Por tanto, la tesis presta especial atención a la modelización econométrica de la heterogeneidad no observada en preferencias.

Entender los factores que hacen que los individuos decidan viajar a un sitio o a otro resulta relevante desde una perspectiva de política económica. Los gestores regionales pueden estar interesados en conocer qué características atraen a potenciales visitantes a sus regiones. Entender las fuentes de la heterogeneidad en las preferencias turísticas puede también contribuir a un mejor desarrollo de campañas publicitarias y decisiones de política pública orientadas a atraer visitantes hacia las regiones.

La tesis está estructurada en tres capítulos. Los dos primeros utilizan datos de 'preferencias reveladas' (PR) obtenidos por medio de encuestas. Este tipo de datos tiene la ventaja de que refleja las decisiones y compras realizadas en mercados reales, siendo el tipo de datos más usado en los análisis empíricos. Sin embargo, tiene algunos inconvenientes. En primer lugar, dado que los investigadores habitualmente utilizan datos secundarios obtenidos a partir de encuestas realizadas por oficinas estadísticas oficiales, el análisis se tiene que limitar a las variables recogidas en la encuesta. En segundo lugar, los datos de preferencias reveladas habitualmente dan lugar a problemas de colinealidad y falta de suficiente variabilidad en las variables de interés, lo que dificulta la identificación de los efectos. Finalmente, los datos de preferencias reveladas solo recogen información sobre las elecciones realizadas. Para el propósito de modelizar la elección entre conjuntos de bienes dadas sus características, el investigador se enfrenta a la dificultad de tener que definir el conjunto de bienes entre los que se escoge. Esto implica tomar una muestra del universo de posibles alternativas consideradas, con la consiguiente incertidumbre acerca de si las alternativas relevantes para el consumidor han sido consideradas en el análisis. Por esta razón, las elecciones discretas a partir de datos de preferencias reveladas llevan asociadas un riesgo de dependencia del menú de opciones consideradas.

Como alternativa a los datos de preferencias reveladas, los investigadores han hecho uso de lo que se conoce como aproximación de preferencias declaradas (PD). Esto consiste en analizar las preferencias en contextos de elección hipotéticos, adecuadamente diseñados e implementados, esto es, experimentos. Entre los distintos tipos de estudios basados en preferencias declaradas, los experimentos de elección discreta tienen una amplia tradición en márquetin, economía del transporte, economía de la salud y economía del medio ambiente. Los experimentos de elección discreta permiten evaluar las preferencias sobre los bienes a partir de las elecciones realizadas entre distintos conjuntos cuyas características son convenientemente definidas. El investigador establece el número de alternativas y los atributos sobre las que se realiza la elección. Por lo tanto, las preferencias se identifican a partir de las elecciones realizadas dado un conjunto de bienes. Este procedimiento es más flexible y conveniente para el propósito de modelizar e identificar los complejos procesos de sustituibilidad que se producen entre los atributos que caracterizan a los bienes[1]. No obstante, como en cualquier conjunto de datos obtenido a partir de preferencias declaradas, los experimentos de elección discreta tienen el problema de que se basan en elecciones en contextos hipotéticos, con el consiguiente 'sesgo hipotético'. Esto se refiere a que lo que los individuos declaran que harían en una cierta situación puede ser diferente a lo que realmente hacen en tal caso. De cualquier manera, existe cierta evidencia de que los experimentos de elección discreta combinados con lo que se conoce como *'cheap talk'* pueden ofrecer estimaciones fiables de las preferencias individuales (Carlsson et al., 2005).

A lo largo de la tesis se hace uso de estos dos tipos de datos con el propósito de modelizar las decisiones turísticas. Es importante destacar que la tesis no los combina. Se llevan a cabo análisis separados, ya que los datos proceden de diferentes muestras. Por lo tanto, la combinación de datos de preferencias declaradas y reveladas está fuera del alcance de la tesis. El objetivo es ofrecer una caracterización *microeconométrica* de las preferencias del consumidor por el turismo explotando las ventajas y desventajas de cada tipo de datos.

La tesis está compuesta por tres capítulos. En el capítulo 1 se estudian los determinantes de la duración de la estancia de los turistas. Más concretamente, se modeliza i) la decisión de pernoctar en un destino turístico (frente a ser un excursionista o visitante de un solo día), y ii) la duración de la estancia condicional en ser un turista. Se hace uso de una rica base de datos de más de 19.000 individuos obtenida a partir de encuestas realizadas durante el periodo 2010-2016 en Asturias (España). Se estima un modelo de conteo tipo valla y se comparan dos especificaciones alternativas para la heterogeneidad no observada (gamma y lognormal).

En el capítulo 2, el foco se pone en el papel que juegan los atributos de las regiones en la elección de destino turístico. Se analizan los viajes recreacionales dentro de España por motivo naturaleza o deporte. Se usan datos mensuales para casi 7.000 individuos recogidos entre febrero de 2015 y septiembre de 2017 procedentes del Instituto Nacional de Estadística. Se estudia el efecto de la distancia al origen, la ratio de temperaturas entre origen y destino, los precios, los kilómetros disponibles para esquiar, la superficie de áreas naturales protegidas y la presencia de costa, entre otras. Se estima un modelo Logit de parámetros aleatorios que incorpora componentes del error. Este modelo controla por heterogeneidad no observada a nivel individual y regional. El modelo permite que las utilidades marginales dependan de un vector de características que

[1] Es importante dejar clara la diferencia entre los estudios de Evaluación Contingente y los estudios basados en experimentos de elección discreta. Mientras que los primeros preguntan directamente a los individuos cuánto dinero estarían dispuestos a pagar por una cierta política, los segundos *infieren* las preferencias de los consumidores basándose en sus elecciones utilizando la teoría microeconómica.

modulan esa intensidad en las preferencias. También se derivan e interpretan relaciones marginales de sustitución y elasticidades.

Finalmente, con el propósito de estudiar las preferencias individuales por diferentes conjuntos de bienes turísticos (paquetes vacacionales), en el capítulo 3 se lleva a cabo un experimento de elección discreta. Esto permite identificar las preferencias individuales de una manera incentivada y en un entorno controlado. Para ello, se recluta a una muestra de parejas reales procedentes de cuatro regiones en el norte de España. Un total de 262 individuos participaron en el experimento. Se estima un modelo de clases latentes que permite identificar diferentes segmentos con diferentes gustos. Seguidamente, se calculan relaciones marginales de sustitución en términos de disponibilidad a pagar junto con la pérdida de bienestar (variación compensada) que se derivaría del establecimiento de una tasa turística diaria por persona.

El capítulo 1 es bastante específico, ya que analiza una decisión turística relevante (número de pernoctaciones) considerando la elección de Asturias como destino como dada. El segundo es más amplio, ya que engloba una modelización microeconómica de la elección de destino considerando las 17 Comunidades Autónomas españolas. Finalmente, el capítulo 3 posiblemente añade el mayor valor al trabajo, ya que estudia la elección de destino vacacional usando datos primarios recogidos por mí mismo a partir de un experimento. Como se ha introducido anteriormente, los participantes son parejas reales reclutadas de la población general. Esto contrasta con la práctica habitual de reclutar estudiantes universitarios, quienes en algunos casos puede que no estén familiarizados con las decisiones que se les plantea realizar. En nuestro caso, la muestra está caracterizada por un alto interés y frecuencia en viajes por motivo recreacional. Por lo tanto, es una muestra relevante para los propósitos de la tesis.

INTRODUCTION

Tourism is nowadays a fast-growing industry. In 2010, there were 950 million arrivals in the world. By 2018, this figure increased to 1.4 billion arrivals. The tourism sector is estimated to account for 10.4% of global GDP and generates 319 million jobs. During the past eight years, the sector has grown at a higher rate than the global economy (UWNTO, 2019), being the second most important industry in the economy behind manufacturing in 2018. Among the most important tourist destinations, Spain stands as the second most visited country in the world. According to the Spanish National Statistics Institute (INE, 2019), Spain received a total of 82.6 million arrivals in 2018 (11.7% of GDP, 12.8% of total employment).

Despite the relevance of international tourism, the Spanish domestic travel market has gained importance in the last decade. In 2018, it accounts for over two thirds of the aggregate demand. More than 100 million domestic trips were made for leisure purposes, representing an aggregate expenditure of more than 28 billion euros. However, there are relatively few studies that examine, at the micro level, the factors that pull people to travel to one region or another. In this sense, the economic modelling of domestic trips has been overlooked in favour of analysing the international market.

A large body of empirical evidence shows that tourism is a driver of economic growth, both for developed and developing countries (Lee and Chang, 2008; Faber and Gaubert, 2019). For the case of Spain, some studies have proved the existence of a long-run stable relationship between economic growth and tourism activities (e.g. Capó-Parrilla et al., 2007), which translates into important long-run multiplier effects.

Given the growing economic relevance of the tourism industry in the world and in Spain in particular, it seems necessary to study consumer preferences for tourism. Although there is ample research about tourist decisions, there is a lack of understanding of the drivers of tourism-related preferences from a microeconomic viewpoint. In this regard, the thesis places the emphasis on the economic modelling of tourist choices. Therefore, the main aim is to provide a microeconometric theoretically consistent analysis of individual preferences for tourism.

A large body of literature has concentrated on the study of aggregate tourism flows. However, it is acknowledged in economics that micro-level data provides a better characterization of consumer behaviour, since it exploits cross-sectional variability and is less sensitive to aggregation biases (e.g. Blundell et al., 1993; McGuckin, 1995). Accordingly, the thesis makes use of individual-level data to provide further empirical evidence about why individuals make different choices. In doing so, special attention is devoted to the econometric modelling of heterogeneity.

The framework is built upon Lancaster's product characteristics approach (Lancaster, 1966). This theory posits that goods have objective characteristics that produce utility. Consumers thus choose among bundles of goods based on characteristics. Under a utility maximizing framework, the choice among bundles of goods can be used to infer preferences for characteristics. However, individuals differ in their reaction to those characteristics. That is, marginal utilities are heterogeneous in the population. Part of this heterogeneity can be linked to individual characteristics, such as age or income. In this vein, the use of demographic features as a source of heterogeneity in the taste for goods has a long tradition in economic theory (e.g. Pollak and Wales, 1981). Another important source of heterogeneity is unobserved and individual-specific. Accordingly, the thesis devotes attention to the econometric modelling of taste heterogeneity.

Understanding the factors that make individuals to travel to one place or another seems to be policy relevant. Regional decision makers might be concerned about what place-based attributes pull prospective visitors to their regions. Disentangling the sources of heterogeneity in consumer preferences for tourism may thus help in the development of promotional campaigns and policy strategies aimed at attracting visitors to regions.

The thesis is composed of three essays. The first two make use of 'revealed preferences' (RP) obtained from survey data. This sort of data has the advantage of reflecting actual decisions and purchases from real markets, being the most used data in empirical analysis. However, it has some drawbacks. First, since researchers normally use secondary data obtained from surveys conducted by official information services, it limits the scope of the analysis to the variables gathered in the survey. Second, RP data usually suffer from collinearity and lack of enough variation in the relevant variables, hindering the separate identification of effects. Finally, in general RP data gathers information on the decisions finally made. For the specific context of modelling choices among bundles of goods given their characteristics, the researcher faces the challenge of defining the choice set. This implies sampling from the universe of possible alternatives, with the subsequent uncertainty about whether the relevant alternatives have been considered in the analysis. Consequently, choices from RP data involve the risk of menu dependence.

As an alternative to revealed preferences data, researchers have made use of the so-called 'stated preferences' (SP) approach. This consists on examining preferences in carefully designed and well-implemented hypothetical choice environments (experiments). Among the different types of SP, Discrete Choice Experiments have a long tradition in marketing, transportation, health and environmental economics. DCE allow preferences over goods to be evaluated based on choices made among goods whose characteristics are properly defined. The researcher clearly delineates the number and attributes of the alternatives in the choice set, so preferences are identified from choices conditional on the choice set. This procedure is more flexible and better suited for the purpose of modelling and identifying complex trade-offs among the attributes[2]. Nevertheless, as happens with all types of SP data, DCE have the problem of relying on declared choices in hypothetical situations, being thus subject to hypothetical bias (i.e. what individuals declare they would choose in a given situation is perhaps different from what they actually do). In any case, there is some evidence that DCE combined with 'cheap talks' can provide reliable estimates of individual preferences (e.g. Carlsson et al., 2005).

[2] Please note the important distinction between Contingent Valuation studies and Choice Experiments. While the former directly asks individuals how much they are willing to pay for a policy change, the latter infers preferences based on choices using microeconomic theory.

Throughout the thesis, I make use of the two types of data for modelling tourism decisions. Importantly, the thesis does not combine them. I conduct separate analysis since my data come from different samples. Therefore, the combination of preferences from the two types of data is beyond the scope of the dissertation. Instead, the aim is to provide a microeconometric characterization of consumer preferences for tourism exploiting the advantages and disadvantages of each kind of data.

The thesis is structured in three chapters. In Chapter 1, the determinants of tourists' length of stay are examined. More specifically, this first essay models i) the decision to stay overnight at a tourist destination (versus being a same-day visitor), and ii) how long to stay conditional on being a tourist. I make use of a rich dataset of more than 19,000 individuals obtained from surveys during the period 2010-2016 in Asturias (Spain). A Hurdle Zero-Truncated count data model is estimated using and comparing two alternative distributions for unobserved heterogeneity (gamma and lognormal).

In Chapter 2, the focus is on the role of place-based attributes on tourist choice of destination. I analyse nature-based recreational trips within Spain, using monthly data for almost 7,000 individuals between February 2015 and September 2017 from the Spanish National Statistics Institute. This article studies the effect of distance to origin, relative temperatures, prices, kilometres available for ski, the protected natural surface, and the presence of coast, among others. The estimated model is a Random Parameter Logit with Error Components that controls for unobserved heterogeneity at the individual and the regional level. The model allows the marginal utilities to depend on a vector of taste shifters, and marginal rates of substitution and elasticities are computed and interpreted.

Finally, for the purpose of studying individual preferences for alternative tourism bundles of goods (vacation packages), in Chapter 3 I conduct a Discrete Choice Experiment. This allows me to elicit individual preferences in an incentive-compatible controlled setting. A sample of real-life couples from four northern Spanish cities is recruited for this purpose. A total of 262 individuals participated in the experiment. I estimate a Latent Class Model that allows the identification of segments with different tastes. I subsequently derive Marginal Rates of Substitution in the form of Willingness to Pay estimates together with the loss in welfare (compensating variation) arising from a daily tourism tax per person.

The first chapter is quite specific, since it analyses a relevant tourism decision (length of stay) taking the choice of the destination as given. The second is much broader, since it encompasses the microeconomic modelling of destination choice considering the 17 Spanish Autonomous Communities. Finally, the third possibly adds the greatest value to the work, since it examines the choice of a vacation destination using primary data obtained from an experiment. As introduced before, participants are real-life couples recruited from the general population. This contrasts with the usual practise of recruiting university students, who in some settings might not be familiar with the task being confronted with. In this case, the sample is characterized by a high interest and frequency in recreation and travelling, thereby constituting a relevant sample for the study purposes.

REFERENCES

Blundell, R., Pashardes, P. and Weber, G. (1993). “What do we learn about consumer demand patterns from micro data?” *American Economic Review*, 83(3), 570-597.

Capó-Parrilla, J., Riera, A. and Rosselló-Nadal, J. (2007). "Tourism and long-term growth: a Spanish perspective". *Annals of Tourism Research*, 34(3), 709-726.

Carlsson, F., Frykblom, P. and Lagerkvist, C.J. (2005). “Using cheap talk as a test of validity in choice experiments”. *Economics Letters*, 89(2), 147-152.

Faber, B. and Gaubert, C. (2019). "Tourism and economic development: evidence from Mexico's coastline". *American Economic Review*, 109(6), 2245-2293.

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

Lee, C. and Chang, C. (2008). "Tourism development and economic growth: a closer look at panels". *Tourism Management*, 29, 180-192.

McGuckin, R.H. (1995). “Establishment microdata for economic research and policy analysis: looking beyond the aggregates”. *Journal of Business Economics & Statistics*, 13(1), 121-126.

Pollak, R.A. and Wales, T.J. (1981). “Demographic variables in demand analysis”. *Econometrica*, 49(6), 1533-1551.

Chapter 1.- Determinants of Tourists' Length of Stay: A Hurdle Count Data Approach

David Boto-García

botodavid@uniovi.es

Department of Economics. University of Oviedo

Avenida del Cristo s/n, Oviedo (Asturias). Faculty of Business and Economics, Oviedo (Spain).

Abstract:

This article analyses tourists' length of stay at a particular destination using a Hurdle Count Data model that allows us to i) identify the determinants of the decision to be a same-day visitor or a tourist, and ii) explain the length of stay of those who stay overnight. Apart from sociodemographic characteristics, our interest relies on the effect of distance, mode of transport and destination-specific pull factors, such as tranquillity, natural environment or climate. Another relevant feature this paper addresses is how advertising, recommendation and previous experience influence both the probability of an overnight stay and the length of the stay. The results indicate that the determinants of the decision to stay overnight and how long to stay are not the same. Furthermore, positive previous experience and having seen advertising of the destination positively affect the decision to stay overnight and the number of days to stay.

Keywords: *length of stay, tourist's decision-making, conditional demand, Hurdle Negative Binomial model, same-day visitor*

JEL codes: C35, D12, D81

Acknowledgements:

I am grateful to Joao Santos-Silva, Helmut Farbmacher, Eduardo del Valle, Luis Valdés, members of the Department of Economics at the University of Oviedo, and my PhD advisors José Baños and Antonio Álvarez for helpful comments and suggestions.
I would also like to thank the *Tourist Information System of Asturias* for providing me with the database and helping me with doubts regarding the sampling procedure and the questionnaire.

*Note: a reduced version of this paper has been published in *Journal of Travel Research*.

Boto-García, D., Baños-Pino, J.F, and Alvarez, A. (2019). "Determinants of Tourists' Length of Stay: a Hurdle Count Data Approach". *Journal of Travel Research*, 58(6), 977-994.

Empirical Research Article

Journal of Travel Research
2019, Vol. 58(6) 977–994

DOI: 10.1177/0047287518793041
journals.sagepub.com/home/jtr
SAGE

Determinants of Tourists' Length of Stay: A Hurdle Count Data Approach

David Boto-García[1], José Francisco Baños-Pino[1], and Antonio Álvarez[1]

Abstract

This article analyses tourists' length of stay in a particular destination using a Hurdle Count Data model that allows us to first identify the determinants of the decision to be a same-day visitor or a tourist, and then to explain the length of stay for tourists. Apart from sociodemographic characteristics, we are interested in the effects of distance, mode of transport, and some relevant destination attributes of the destination such as tranquillity, natural environment, or climate. Another feature this article addresses is how advertising, recommendations, and previous experience at the destination affect both the probability of an overnight stay and the length of the stay. The results indicate that the determinants of the decision to stay overnight and how long to stay are not the same. Besides, a positive previous experience and having seen advertising of the destination positively affect the overnight stay decision and the number of days.

Keywords

length of stay, tourist's decision making, conditional demand, Hurdle Negative Binomial model

Introduction

Length of stay at a tourist destination is one of the most relevant issues in the tourist decision-making process (Decrop and Snelders 2004). The economic impact of tourism basically depends on the number of days the tourist stays at the destination. In this sense, many studies have found evidence of a strong correlation between length of stay and total expenditure (Leones, Colby, and Crandall 1998; Laesser and Crouch 2006). Because of this, uncovering the determinants of length of stay is critical for the proper design of marketing policies oriented to increase the revenues generated by tourism.

There are several studies in the literature that have examined the effects of sociodemographic characteristics, such as age, income, and nationality, on length of stay (e.g., Barros and Machado 2010). Additionally, other scholars analyze the relationship between the number of days the visitor spends at a destination and the mode of transport used, the type of accommodation selected, or the purpose of the trip (e.g., Alén et al. 2014). Another issue that deserves attention is the visitor's choice between being a same-day visitor or a tourist. In this sense, the determinants of the visitor's decision to stay overnight or not in a destination have been less studied in the literature, with the exception of Rodríguez, Martínez-Roget, and González-Murias (2018). Given that tourism products are in essence experiential, tourists normally face a high level of uncertainty when deciding whether or not to stay overnight at a particular destination and how long to stay. Therefore, their knowledge about the characteristics of the destination will be a critical factor and justifies the interest of our analysis.

The aim of this research is twofold. Firstly, we analyze the determinants of the decision to stay overnight, differentiating between same-day visitors and tourists. Second, we model the length of the stay, focusing on the role that tourists' knowledge about the destination and the attributes they most value play in tourist decision making. Specifically, we examine how a positive previous experience at the destination, looking for the natural environment, or for tranquility and recommendations from friends or relatives (*word of mouth* effect) affect the length of the stay. Another issue of interest is how stay duration is connected with distance to origin and the chosen mode of transport.

This article employs a pooled cross-sectional data set of tourists visiting Asturias, a region located in northern Spain, during the period 2010-2016. For analyzing the effects of different sources of information on the length of stay, we estimate a hurdle count data model (Mullahy 1986). This methodology allows us to both identify the factors that determine the decision to stay overnight and how long to stay for those who spend at least one night in Asturias. From a methodological point of view, we consider two competing specifications for modeling the positive outcomes, namely, a Zero-Truncated Negative Binomial P (ZTNBP) and a Zero-Truncated Poisson-lognormal (ZTPN) model, and compare

[1]Department of Economics, University of Oviedo, Oviedo, Spain

Corresponding Author:
José Francisco Baños-Pino, Department of Economics, University of Oviedo, Campus del Cristo, 33006, Oviedo, Spain.
Email: jbanos@uniovi.es

1. INTRODUCTION

Length of stay in a tourist destination is one of the most relevant issues in tourist decision-making process (Decrop and Snelders, 2004). The economic impact of tourism basically depends on tourists' length of stay. In this sense, several studies find evidence of a strong correlation between length of stay and total expenditure (Laesser and Crouch, 2006; Thrane and Farstad, 2011; Aguiló et al., 2017). Because of this, uncovering the determinants of length of stay is relevant to the appropriate design of marketing policies oriented to promote longer stays.

There are several studies that examine the effects of sociodemographic characteristics, such as age, income, and nationality, on length of stay (e.g. Barros and Machado 2010). Other scholars analyze the relationship between the number of stays and the mode of transport, the type of accommodation selected or the purpose of the trip (e.g. Alén et al., 2014). Another issue that deserves attention is the visitor's choice between being a same-day visitor or a tourist. In this sense, the determinants of the decision to stay overnight have been less studied in the literature, with the exception of Rodríguez et al. (2018). Since tourism products are experience goods, individuals usually face a high level of uncertainty when deciding whether to stay overnight and how long to stay. Therefore, their knowledge about the characteristics of the destination might play an important role and justifies the interest of our analysis.

The aim of this research is twofold. First, we analyze the determinants of the decision to stay overnight by modelling the likelihood of being a tourist versus a same-day visitor. Second, we model the length of the stay, focusing on the role that information about the destination plays in tourist decision-making. Specifically, we examine how having seen advertising, positive previous experience at the destination and recommendation from friends or relatives (*word of mouth*) affect the length of the stay. We also study how duration relates to distance to origin and the chosen mode of transport.

This paper employs a pooled cross-sectional dataset of tourists visiting Asturias, a region in northern Spain, during the period 2010-2016. For analyzing the effects of different sources of information on the length of the stay, we estimate a hurdle count data model (Mullahy 1986; Hellström and Nordström, 2008). This methodology allows us to identify the factors that determine the decision to stay overnight on the one hand, and how long to stay for those who spend at least one night on the other. From a methodological perspective, we consider two competing specifications for modelling the positive outcomes, namely, a Zero-Truncated Negative Binomial P model (ZTNBP) and a Zero-Truncated Poisson-lognormal (ZTPN) model. We compare which of them fits the data best. To the best of our knowledge, this is the first empirical study that employs a hurdle count data model in tourism research and compares two alternative specifications for the unobserved heterogeneity for the positive outcomes.

Our results show that recommendation from friends or relatives increases the likelihood of an overnight stay, but it has no effect on how long to stay. Having seen any type of advertising about the destination and positive previous experience are positively associated with the decision to stay overnight and with the number of stays. Furthermore, the climate and the natural and tranquil environment are found to be the main destination attributes that increase the length of the stay. Booking the trip through travel agencies and lodging at hotels leads to the longest stays. Foreign people tend to stay longer than Spanish tourists, whereas education is not significant for explaining the number of stays.

The paper is structured as follows. After this introductory section, in Section 2 we review the related literature. In Section 3 we develop the theoretical model. Section 4 describes the database and the variables employed. In Section 5 we present a methodological discussion together with the empirical model. Section 6 reports the results and discusses their implications. Finally, Section 7 outlines the main conclusions.

2. LITERATURE REVIEW

The economic relevance of tourism has sparked an increasing interest in analyzing its determinants. With reference to length of stay (hereafter LOS), some scholars have conducted descriptive analysis of the differences in LOS based on tourists' socio-demographic and/or trip-related characteristics, including Oppermann (1995, 1997) and Lew and McKercher (2002), among others. These studies show how LOS changes with nationality, age, labour status, repeat visit behavior, the stage in the life cycle and distance between the place of origin and the destination, among others. However, their descriptive nature hinders formal inference on the linkages between individual socio-demographic profiles and trip experiences and length of stay. Nonetheless, in the last decade, many researchers have employed regression models to explain LOS, which allows the researcher to study the effect of a covariate on the dependent variable, *ceteris paribus*. We now proceed to discuss the main findings on the effects of tourist's sociodemographic features, trip-related characteristics and supply-based factors on LOS.

Sociodemographic characteristics

The empirical evidence about the effect of gender on LOS is mixed. Barros and Machado (2010) and Machado (2010) show that males stay longer whereas Gomes de Menezes (2008), Gomes de Menezes and Moniz (2011), Oliveira Santos et al. (2015) and Rodríguez et al. (2018) find evidence of the opposite. Moreover, other scholars do not find significant differences (Martínez-García and Raya, 2008; Gomes de Menezes et al., 2008; Barros et al., 2010; Machado, 2010; Wang et al., 2012; Brida et al., 2013; Alén et al., 2014). As for the effect of age, several studies document that LOS is positively associated with tourist's age. Examples include Alegre and Pou (2006), Hellstrom (2006), Martínez-García and Raya (2008), Barros et al. (2008), Gomes de Menezes et al. (2008), Machado (2010), Barros and Machado (2010), Thrane and Farstad (2012), Ferrer-Rosell et al. (2014) and Alén et al. (2014). Regarding labor status, the evidence is also inconclusive. Alegre and Pou (2006) show that highly qualified workers display lower LOS whereas Martínez-García and Raya (2008) note that self-employed and low-level employees tend to stay shorter. Likewise, there is no consensus on the effect of education. While Barros and Machado (2010), Barros et al. (2010), Machado (2010) and Ferrer-Rosell et al. (2014) indicate that educational level and LOS are positively related, Gokovali et al. (2007), Gomes de Menezes et al. (2008), Martínez-García and Raya (2008) and Rodríguez et al. (2018) provide evidence of the contrary. Gomes de Menezes and Moniz (2011) and Brida et al. (2013) do not find significant effects whereas Oliveira-Santos et al. (2015) argues that the relationship between the level of education and LOS exhibits a complex pattern. With regard to nationality, the literature agrees that foreign tourists tend to stay longer (Thrane and Farstad, 2012; Oliveira-Santos et al., 2015). Regarding income, tourism is a normal good so that higher income is associated with longer stays. However, Rodríguez et al. (2018) do not find significant differences by income.

Supply-based factors

Another relevant group of variables are supply-based factors such as destination-attributes and prices. According to Gokovali et al. (2007) and Gomes de Menezes et al. (2008), tourists who attach high importance to natural environment, landscape and beautiful surroundings stay for longer. In this sense, climate is one of the attributes that encourage tourists to stay for more extended periods (Barros et al., 2008; Barros et al., 2010; Alén et al., 2014). Additionally, some studies include tourists' expenditures per day as a proxy of the price per stay (Alegre et al., 2011, Thrane, 2012, Gokovali et al., 2007, Alegre and Pou, 2006). As expected, they obtain a negative relationship with length of stay.

Knowledge of destination

When deciding how long to stay, tourists (and especially first-timers) face a risk of making a bad decision since the specific characteristics of a destination are unknown until the individual arrives there (i.e. intangibility). This 'experience' nature of tourism induces travelers to carry out extensive information search strategies (Roehl and Fesenmaier 1992). Consequently, some authors have included informational-type variables for explaining LOS. One of the most important ones is advertising, which reduces search costs and provides useful information to potential consumers. Woodside and Dubelaar (2002) note that advertising helps to gain positive perceptions about the destination. Brochures and advertising seem to positively affect LOS (Barros et al., 2008; Rodríguez et al., 2018), being nowadays considered the most influential information source for prospective visitors (Woodside and King, 2001; Kim et al. 2005; Park and Nicolau 2015). However, the problem lies on the credibility that potential tourists give to the information retrieved from advertising. For this reason, some scholars point out that individuals rely more on recommendations from friends and relatives, since they perceive them as trustworthy (Gitelson and Kerstetter, 1994; Fodness and Murray, 1997; Bieger and Laesser 2004). In this sense, the "word-of-mouth" effect is well-documented in the tourism industry (Yeoh et al., 2013; Luo and Zhong 2015). Nevertheless, Govokali et al. (2007) do not find a significant relationship with LOS.

Less information search is needed when the individual has been at the destination before and has first-hand information. In such case, the tourist has more confidence in the decisions made and the perceived risk is lower (Kerstetter and Cho 2004). The effect of previous experience has been widely analyzed without a clear conclusion. On the one hand, some studies have shown that first-time visitors stay longer (Thrane, 2012; Nicolau et al., 2018). On the other hand, Gomes de Menezes and Moniz (2011) and Machado (2010) provide evidence suggesting that repeaters stay more days, while Alegre et al.

(2011), Oliveira-Santos et al. (2015) and Rodríguez et al. (2018) find the opposite. When researchers take into account not only whether the tourist has been at the destination before but also how many times, a clearer picture emerges, with the number of previous visits being positively associated with LOS (Barros et al., 2008; Barros and Machado, 2010; Thrane and Farstad 2012). This can be explained by the fact that, as the number of visits to the same destination increases, tourists acquire expertise, which minimizes the uncertainty and leads to more efficient choices regarding accommodation or what to visit.

Trip-related characteristics

The distance between tourist's origin and destination is another key factor in tourism demand (Bell and Leeworthy 1990). As Nicolau et al. (2016) state, "the literature shows little consensus about the effects of distance on length of stay". On the one hand, Taylor and Knudson (1976) argue that distance reduces utility as it entails physical, temporal and financial effort. Moreover, self-drivers or train riders may prefer to stop at various places along the way (e.g. Zillinger 2007). On the other hand, since travel costs are fixed, longer stays allow tourists to spread the costs over a longer period. In line with this, a positive relationship between distance and LOS is found by Gronau (1970) and Paul and Rimmawi (1992). Nonetheless, Gomes de Menezes et al. (2008) and Alén et al. (2014) report that distance does not affect LOS. When the mode of transport is considered, it seems that those who travel by public modes of transport stay for longer (Rodríguez et al., 2018).

Concerning trip purpose, some researchers find that tourists visiting friends or relatives stay for the longest periods (Gomes de Menezes and Moniz, 2011; Oliveira-Santos et al., 2015, Alén et al., 2014). By contrast, others like Rodríguez et al. (2018) argue that those who travel for business purposes are who stay for the largest number of days, whereas non-significant differences by travel purpose are found by Martínez-García and Raya (2008). Party size exerts a negative influence in LOS (Alegre and Pou, 2006; Alén et al., 2014; Oliveira-Santos et al., 2015), with tourists travelling with friends staying fewer days than those who travel with a partner (Gomes de Menezes et al. 2008; Ferrer-Rosell et al., 2014). Concerning accommodation, those staying at hotels stay at the destination for the shortest periods, with the longest stays being associated with those lodged at private accommodations (Martínez-García and Raya, 2008; Alén et al., 2014; Oliveira-Santos et al., 2015). Focusing on hotel tourists, Alegre and Pou (2006) indicate that those lodged at higher-quality hotels stay longer than their lower-quality counterparts, whereas Ferrer-Rosell et al. (2014) find the reversal.

Booking a holiday package is associated with longer stays according to Ferrer-Rosell et al. (2014), but it is not found to be significant in Alegre and Pou (2006) and Martínez-García and Raya (2008). Tourists visiting more than one destination stay for shorter periods at each one than those who spend their whole trip period at a single destination (Gomes de Menezes et al. 2008). Similarly, some researchers find significant differences in LOS depending on the geographical area where they stay (Oliveira-Santos et al. 2015). As for seasonal differences, LOS is longer during the summer season (Thrane, 2012; Oliveira-Santos et al., 2015; Martínez-García and Raya, 2008; Grigolon et al., 2014; Ferrer-Rosell et al., 2014).

Different econometric strategies have been used in the literature: OLS regression (Mak et al. 1977; Thrane and Farstad, 2012), the Heckman model (Rodríguez et al. 2018), duration models (Gokovali et al. 2007; Martinez-Garcia and Raya 2008; Gomes de Menezes et al. 2008; Barros et al. 2008; Barros et al. 2010; Barros and Machado 2010; Machado 2010; Gomes de Menezes and Moniz 2011; Wang et al., 2012; Oliveira-Santos et al. 2015), Logit (Alegre and Pou 2006), Ordered Logit (Ferrer-Rosell et al. 2014), Multinomial Logit (Grigolon et al. 2014), Nested Logit (Nicolau and Más 2009), Latent Class (Alegre et al. 2011) and Count Data models (Brida et al. 2013; Alén et al. 2014; Prebensen et al. 2015; Nicolau et al. 2018). We believe the latter is the most suitable for modelling LOS. We discuss and justify it in Section 5. Whereas previous studies that use count data modeling for studying LOS specified a zero-truncated count data model for the positive stays, we extend it by including a previous hurdle that models the probability of being a tourist relative to being a same-day visitor.

A summary of some recent studies about tourist's LOS is presented in Table 1.1. This table provides a description of the geographic area and the population under analysis, the methodology used and the main findings. Since the different studies presented there refer to different countries, periods and type of tourists, conclusions should be drawn with caution. Nonetheless, a stylized fact is that LOS can be explained by time and economic constraints (daily prices of accommodation, travel costs, income, etc.), travel characteristics (trip purpose, mode of transportation, who you travel with, party size, etc.), sociodemographic characteristics (age, gender, labor status, etc.) and the level of information about the place.

Author(s)	Population, destination and study period	Research question	Methodology	Main results
Alegre and Pou (2006)	British and German tourists in the Balearic Islands (Spain) during the high season (1993-2003)	The effect of sociodemographic and economic constraints on LOS	A Logit model where the dependent variable takes 0 if the tourist stayed for 7 days or less and 1 if she stayed for more than a week	Older people, travelling with a couple, mid-to-high accommodation, the number of yearly trips and having visited the destination before are the main factors that increase the probability of staying for more than 1 week.
Martínez-Roget and Rodríguez-González (2006)	Annual data (1996-2001) for 49 rural establishments in Galicia (Spain)	How important are promotional and quality factors for coming to a rural area?	Dynamic panel data with Fixed and Random Effects	Rural tourism demand depends on economic determinants such as accommodation prices, travel costs and tourists' income. The latter exhibits the highest elasticity. The reputation and peculiarities of each establishment also matters.
Govokali et al. (2007)	Tourists who travelled to Bodrum (Turkey) by plane in the summer of 2005	Determinants of tourists' LOS	Duration models (exponential, Weibull's and Gompertz's parametric approaches)	The probability of staying longer increases with income, previous experience and party size, but decreases with late accommodation booking, daily expenditures and high education.
Martínez-García and Raya (2008)	Low-cost travelers visiting Catalonia (Spain) in 2005	Effect of personal and trip-related characteristics on LOS	Duration models (accelerated failure time model)	The type of accommodation, travelling in the high season and the level of education are quantitatively the most important factors for explaining LOS.
Barros et al. (2008)	Portuguese tourists travelling to South America on charter flights	How do brochures and the degree of advanced booking affect LOS?	Duration models.	The time spam a tourist stays at a destination is positively related to having booked in advance, having seen advertising, previous visits and travel frequency.
Gomes de Menezes et al. (2008)	Tourists departing from the Azores (Portugal) in the summer of 2003	The effect of tourist's reported valuation of environmental initiatives to improve the overall quality of the destination on LOS	Duration model (Cox Proportional Hazard model).	Repeat visitors and those who choose Azores due to its weather and remoteness stay for longer periods. Tourists who live farther away (Nordic or German people) stay for shorter.
Barros and Machado (2010)	Foreign tourists departing from Funchal Airport (Madeira Island) in 2008	Determinants of LOS controlling for possible sample selection bias	Survival sample selection model	Age, gender, education and hotel quality increase LOS but expenditure reduces it. Besides, Germans stay longer than British, Dutch and French tourists.
Barros et al. (2010)	Golfers who visit the Algarve (Portugal) in the spring of 2004	Determinants of golfers' LOS	Duration model (Weibull)	Golfer's LOS depends on nationality, education, age, the type of hotel where the individual stays, climate and the hospitality experience.
Gomes de Menezes and Moniz (2011)	Tourists departing from the Azores (Portugal) in the summer of 2003	How do trip experiences affect LOS?	Duration models	Educational level is not significant for explaining LOS. Besides, repeat visitors, taking charter flights and those who visit friends or relatives exhibit longer stays.
Alegre et al. (2011)	International visitors to Norway during the summer 2007	Differences in LOS based on nationality	OLS and duration models	Tourists from neighboring countries to Norway stay for shorter than those from elsewhere in Europe.

Thrane and Farstad (2012)	Visitors to the Archaeological Museum of Bolzano (Italy) from June to August 2010	Determinants of cultural visitors' LOS	Count data (Zero Truncated Negative Binomial model)	Visitors under 30 are who stay for the shortest period. Hosting the Ötzi museum is the most valuable attribute for visiting the city. Bad weather and travel costs negatively influence LOS.
Brida et al. (2013)	Spanish residents over 55 in 2012	Determinants of LOS for the case of senior tourists	Count data (Zero Truncated Negative Binomial model)	The variables that are associated with longer stays are age, visiting friends or relatives, the pursuit of good climate conditions, accommodation in a holiday apartment, travelling alone and the IMSERSO type of holiday.
Alén et al. (2014)	Dutch tourists in the period 2002-2009	Determinants of going on holidays for short, medium or long periods	Dynamic Mixed Multinomial Logit model for panel data	There is a great need for vacation when medium or long trips were not taken in the current year. Tourists exhibit a high seasonality, being the summer season the preferred period for travelling.
Grigolon et al. (2014)	Visitors to Brazilian destinations between 2004-2010	Determinants of LOS considering multiple destinations	Shared heterogeneity duration model (log-normal gamma model)	Income does not have a significant effect on LOS. Asians and Oceanians are the ones with the longest stays. Those who visit more than one destination stay for shorter periods. Interestingly, party size exerts a non-linear negative effect on LOS.
Oliveira-Santos et al. (2015)	Norwegian students at summer vacation destinations in 2014	Differences in LOS between those who decide the return date beforehand and those who take the decision along the way	OLS regression, a Weibull survival model and a Zero-Truncated Negative Binomial regression	'Open-returners' have the longest LOS, suggesting that 'pre-fixed' returners face more economic and time constraints.
Thrane (2016)	Visitors to an Atlantic Coast destination of the United States (2010-2012)	How distance and previous visits affect LOS	Count data (Zero Truncated Negative Binomial Model)	As distance increases, LOS increases too to compensate for the effort made in the journey. First-time visitation has a significant positive effect on LOS, maybe due to the willingness to widely explore a new destination.
Nicolau et al. (2018)	Visitors to Santiago de Compostela (Spain) during 2005-2012	Tourists and same-day visitors' sociodemographic profiles	Heckman model and separate Probit and Zero-truncated OLS regression	Young and retired people who visit Santiago for leisure purposes display a higher probability of being a same-day visitor, whereas labor-related visitors are the ones with the longest stays.
Rodríguez et al. (2018)	International visitors to Norway during the summer 2007	Differences in LOS based on nationality	OLS and duration models	Tourists from neighboring countries to Norway stay for shorter than those from elsewhere in Europe.

Table 1.1.- Summary of related literature.

3. THEORETICAL FRAMEWORK

Lancaster's theory of value (Lancaster, 1966) indicates that consumer's utility is derived by the consumption of the attributes of the goods. According to this theory, it is the characteristics of the goods (physical entities) what produce utility. Consumers compare the utility of different destination alternatives and choose the one that maximizes utility subject to time and economic constraints (McFadden, 1974).

In our study context, conditional on having chosen the utility-maximizing destination, a tourist gets utility from *being* at the destination for a certain period of time (Rugg, 1973). That is, the specific attributes that make a tourist to choose a destination A over a destination B need time to be transformed into utility. Accordingly, given a chosen destination, individuals demand time (stay at the destination).

The choice of destination and how long to stay constitute a classic case of discrete-continuous choice in consumer demand (Dubin and McFadden, 1984; Hanemann, 1984). Empirical studies on the joint modelling of choice of destination and length of stay include Van Nostrand et al. (2013), Bhat et al. (2016) and Gosens and Rouwendal (2018). However, given the nature of our data, here we take the choice of destination as given and focus on the modelling of the demand for time at the destination. Therefore, our analysis is conditional on having decided to travel to a particular location[3].

The standard tourism demand model is based on a multistage budgeting process in which the economic budget is separately allocated into leisure goods and other consumption goods. Following Dubin and McFadden (1984), tourist's LOS is the result of a utility maximization process subject to budget and time constraints so that:

$$Max\ U\ (q, Z, t_{trans}, LOS, \eta, \omega)$$

subject to

$$\begin{gathered} p'q\ +\ p_{trans} + p_{tou}LOS \leq Y \\ t_{trans} + LOS \leq T \\ q, Z,\ t_{trans}, p, p_{trans}, p_{tou} \geq 0 \end{gathered} \tag{1.1}$$

where q is a vector of consumer goods (excluding tourist services); Z refers to the characteristics of the trip (accommodation, mode of transport, etc.); t is the total length of the trip, disaggregated into the necessary travel time for reaching to the destination (t_{trans}) and the length of the stay there (LOS); η represents a set of sociodemographic characteristics that determine tourist preferences (taste shifters), and ω is a random error term for non-observable factors (McFadden, 1981). Similarly, p is the price vector of goods other than tourism, p_{trans} and p_{tou} are the daily prices of transport and accommodation, respectively, and Y denotes available income for travelling.

[3] In any case, studies that jointly model the choice of destination and the time allocation proceed in two stages. The time allocated to leisure activities is modelled as a second step (Gosens and Rouwendal, 2018).

Under the assumption that the tourist has previously chosen destination j based on utility maximization, the conditional demand function for LOS given the characteristics of the trip can be expressed as:

$$LOS_j = f\ (p, p_{j-tou}, Z_j, Y - pq - p_{j-trans}, T - t_{j-trans}, \eta, \omega) \qquad (1.2)$$

The previous expression conditions LOS on the selected trip characteristics. This conditional demand function (Pollak, 1969, 1971) allows us to estimate LOS taking pre-fixed values of the selected destination and trip characteristics so that length of stay explicitly depends on these arguments[4].

Under the assumption that the utility function of the individual is weakly separable into tourism and non-tourism goods[5], the conditional demand function for LOS can be written as:

$$LOS_j = f\ (p_{j-tou}, Z_j, Y^* - p_{j-trans}, T - t_{j-trans}, \eta, \omega) \qquad (1.3)$$

Since the demand for time at destination conditions on the trip features, LOS is assumed to be decided *after* the selection of destination, the mode of transport and the accommodation. This is a reasonable assumption[6].

Following this approach and in line with the empirical literature[7], LOS can be thus modelled as a function of daily prices ($Price$), income ($Income$), trip characteristics ($Trip$), destination-specific amenities ($Attrib$) and sociodemographic features (Soc), plus a random error term for non-observable characteristics (ω).

Our empirical model for the LOS conditional demand function can be written as:

$$LOS = f\ (Price, Trip, Attrib, Income, Soc) + \omega \qquad (1.4)$$

4. DATABASE

Our analysis of tourist's length of stay employs a pooled cross-sectional database of individuals visiting Asturias in the period 2010-2016. The *Tourist Information System of Asturias* conducts a detailed survey throughout the whole year to a representative sample of all visitors over 18 to Asturias, a northern Spanish region with 10,604 square kilometers. Data were collected through personal interviews using a mixture of i) quota random sampling procedure based on type of visitor, type of accommodation, geographical area, day of the week and month[8], and ii) pure random sampling. The

[4] See earlier characterization by Alegre and Pou (2006).

[5] This implies that "goods can be partitioned into different groups so that preferences within groups can be described independently of the quantities of other goods" (Deaton and Muellbauer, 1980, p.122).

[6] Nevertheless, the true order in which the tourist takes her decisions is unknown.

[7] See Oliveira-Santos et al. (2015) for a review.

[8] As opposed to random sampling, quota sampling allows the sample to be representative of the total population under study, overcoming the possible selection bias that may arise. We refer to Santos-Silva (1997) for a discussion on this.

sample size was determined according to a 95% confidence level with a 5% error[9]. The questionnaires were completed both in the street and in collective establishments all over the Asturian geography at different locations. They were available in Spanish, German, English and French[10]. The survey gathers microdata about tourist sociodemographic characteristics, travel motivation, places visited, total number of nights spent, mode of transport, place of origin, expenditure and type of accommodation, among others. During the study period, a total of 33,461 surveys were collected.

Visitors who stayed in Asturias for more than 30 days were excluded, since they should not be considered as tourists (Hellström and Nordström, 2008). Similarly, local tourists (those who live in the region) were not considered in our analysis. After further excluding some individuals with missing values for certain variables, our final sample consists of 19,111 individuals[11]. An appealing feature of this dataset is that it gathers information about both same-day visitors and *true* tourists[12], so it allows us to identify the factors that affect the decision to stay overnight.

Asturias is a region characterized by its natural surroundings, its beautiful landscapes and mild weather. It has experienced a notable increase in the number of visitors during the last decade, from six million in 2006 to more than seven in 2016. The tourist sector is currently one of the most important ones for this region, representing 10% of its Gross Domestic Product and 12% of its total employment. Average LOS continuously fell between 2010-2014, decreasing from 4.6 average stays in 2010 to 4.2 in 2014. However, in 2015 and 2016 it increased to 4.6 and 4.5, respectively. Only 8.5% of the total visitors are same-day visitors. Around 35% have seen some type of advertising about Asturias. About 39% declare they have come for the first time and 35% consider 'novelty' as the main reason for coming. The principal trip purpose is holiday/leisure (86%) and they mainly travel by car (82%) and in a couple (51%). Most visit Asturias between May-August (49%) and have organized the trip themselves (90%). The preferred area is the central one (45.8%), which includes the three main cities (Oviedo-Gijón-Avilés) of the region. Distance to origin is, on average, 675 kilometers. However, only 8.1% come from abroad. The average expenditure per person and day is €72. Finally, hotels are the preferred type of accommodation (56.4%).

We consider the following groups of variables as explanatory of LOS:

- *Sociodemographic characteristics (Soc):* gender (*male*); age, both in levels and in a quadratic form, denoted by *age* and *agesq*; labor status, distinguishing among *employed*, *self-employed* (base category), *student*, *housewife*, *unemployed* and *retired*; education level, considering *primary* (base category), *secondary* and high education (*higheduc*); and whether the respondent does not live in Spain, labelled *foreign*. Unfortunately, we lack data on income in our dataset. We are aware of its critical importance from an economic point of view.

[9] Annex 1 in the Supplementary Material provides further details on the survey design.

[10] The English version of the questionnaire is shown in Annex 2 in the Supplementary Material.

[11] The number of surveyed individuals for each of the seven years considered is a representative sample of the total visitors per year (2,581; 2,479; 2,635; 2,167; 2,832; 3,438; and 2,979 respectively).

[12] The UNWTO classifies a visitor as a 'tourist' if the trip includes an overnight stay at the destination and as a 'same-day visitor' otherwise.

Since we have information about age, education level and labor status, and according to the Mincer earnings function (Mincer, 1974), income can be proxied with these three variables.

- *Supply-based factors (Attrib):* in the survey tourists are asked about the main reason for travelling to Asturias. They can choose among the following alternatives: natural environment (*natural*), novelty seeking (*novelty*)[13], tranquility (*tranquility*), climate conditions (*climate*), gastronomy (*gastronomy*) and heritage (base category). All of them are defined as binary indicators. As for the daily prices, we consider the daily price paid for accommodation per person (denoted as *accom_price*) in euros. Lodging expenditures per day are a good proxy of the minimum cost of each day spent[14] (Mak et al., 1977; Silberman, 1985; Gokovali et al., 2007; Alegre et al., 2011).
- *Knowledge of destination (Knowledge):* individuals are asked whether they have seen any type of advertising about Asturias and whether they have previously visited the region. If so, they also report the total number of visits. Furthermore, they indicate their main reason for choosing Asturias, with two of the possibilities being recommendation from friends or relatives and a past positive experience. We define the following variables: *first* (if it is the first time the individual visits Asturias), *num_vis* (the number of visits made during the year, to control for the frequency of visits), *experience* (if the individual states that a positive previous visit is the main reason for returning), *advert* (which takes the value 1 if the tourist has seen any type of advertisement, regardless of whether it was via the internet, a brochure or a TV spot) and *recommend* (if she declares she has visited Asturias due to recommendation).
- *Trip-related characteristics (Trip):* distance to origin, measured as the total number of kilometers from the tourist's residence to Oviedo (the centroid), both in levels and in a squared form to allow for non-linearities (*distance* and *distancesq*, respectively); mode of transport, distinguishing among car (base category), *bus*, *train*, or *plane*; the purpose of the trip, considering *leisure*, labor-related (*labor*), visiting relatives (*family*) or other options, such as sport events, doctor visits, religious peregrination or making purchases (reference category); party size (number of members in the travel group, denoted by *party_size*); trip companions, considering *alone*, as a couple (*couple*) or in a group, being the latter the base category; type of accommodation, distinguishing among *hotel*, *hostel*, *rural house*, private accommodation (*private*) or campsite, being the latter the base category; how the trip was organized, considering the following options: the individual did it herself (base category), through a travel agency (*travel_agency*) or the company where she works/a club to whom she belongs organized it for her (*club_firm*); and two binary indicators for whether the

[13] It refers to the "inclination of consumers to shift from a choice they made on the most recent occasion" (Ratner et al., 1999). It also includes boredom alleviation, surprise, thrill and adventure.

[14] Using daily accommodation expenditure as a measure of the daily price is subject to criticism, since different expenditures may arise from different qualities of the goods and services consumed (Oliveira-Santos et al., 2015). However, since we also control for several features, the daily price of lodging per person represents the price of the tourism good (each overnight stay) *conditional* on tourist's preferences and characteristics.

individual only visits Asturias in this trip (*only_ast*) and for whether the tourist conducts active tourism activities (*act_tour*)[15].

- We also consider temporal factors (*Temp*) that may influence the decision regarding how long to stay. Specifically, we include year fixed effects (2010 is the base category) and two dummy variables for the period within the year the trip takes place[16]. Finally, we also control for the regional area (Area) where the tourist stays, distinguishing among the following: west area (*west*), central area (base category), capital city area (*capcity*), east inner (*east_inner*) and east coast (*east_coast*).

Annex 3 presents descriptive statistics of all the variables used in the analysis, their acronym and definition.

5. METHODOLOGY

5.1. Methodological discussion

There are several econometric models to explain LOS as a function of variables. The choice of one or another depends on the assumptions about the dependent variable. Researchers that consider length of stay to be continuous have mainly employed duration models[17]. Duration models estimate the length of time elapsed in a given state before transition to another state as a function of variables (Cameron and Trivedi, 2005). Applied to our context, the relevant feature is not the unconditional duration of the trip but the probability of ending the stay in period *t*, conditional on having stayed until that moment (Kiefer, 1988, p.651). That is, the hazard function. In duration models, the estimation results can be interpreted in two ways: as the effect of a covariate on the conditional probability that the trip ends or as the effect of a variable on the expected value of the length of the stay. Consequently, if a variable positively affects the conditional probability of the travel ending in the next period, then it negatively affects the expected value of the length of the stay.

In our view, duration models are not appropriate for modeling tourists' LOS. As Thrane (2012) criticizes, tourists' LOS cannot be understood as a process by which in each period tourists face a real 'risk' of leaving the destination. Most tourists organize their trip in advance, so accommodation and transport are booked for specific dates. This means that if a tourist has decided to stay for a week, her probability of leaving it before the day seven would be close to zero, being the one for the seventh day almost one. Therefore, if the hazard function is 'fixed', it makes little sense to apply duration models. Indeed, these models are a useful econometric tool for longitudinal data, but its application to cross-sectional data seems to be unsuitable (Thrane, 2012).

[15] This refers to a style or philosophy of leisure travel that combines adventure, nature and cultural tourism, with a particular emphasis on low-impact and sustainable tourism and the use of local guides. It involves a wide range of activities such as trekking, horse routes, climbing and cycling, among others.

[16] We block the year into three groups of four months each: January-April (base category), May-August (*t2*) and September-December (*t3*).

[17] Earlier studies have alternatively used Ordinary Least Squares (Mak et al., 1977; Paul and Rimmawi, 1992).

Because of this reason, the dependent variable (total number of stays) is assumed here as discrete and non-negative so that LOS $\in N = \{0,1,2,\dots\}$. Its modeling as a function of a set of regressors should be better done using count data models (Hellerstein and Mendelsohn, 1993), widely used in the literature (Alegre et al., 2011; Salmasi et al., 2012; Brida et al. 2013; Alén et al., 2014; Nicolau et al., 2018)[18]. Whereas previous studies that used this approach for studying LOS specified a zero-truncated count data model for explaining the positive stays, we extend it by including a previous hurdle that models the probability of being a tourist relative to a same-day visitor.

5.2. A Hurdle count data model

One of the basic assumptions of the standard count data models is that both zeros and positive outcomes come from the same data generating process. However, it makes sense that length of stay is a two-part decision process: first, visitors decide whether to stay overnight at the destination; second, they decide how long to stay. Therefore, there are two types of visitors: those who spend at least one night at the destination (tourists) and those who do not (same-day visitors).

Mullahy (1986) suggests that the effect of the explanatory variables on the probability of participation and on the number of positive counts should not be restricted to be equal. To do so, it seems necessary to first separate participants from non-participants through a binary model, and then in a second step model the number of positive stays using a count data model. There are two alternatives for this purpose: zero-inflated models[19] and hurdle models[20]. The main difference between them is that a zero-inflated model assumes two processes as sources of zeros and combines a count distribution with a discrete point mass as a mixture. By contrast, a hurdle model separately handles zero observations and positive counts, where then a truncated-at-zero count distribution is used for the non-zero state. When choosing among them, it is convenient to think about the data generating processes. In our case, the hurdle model seems to be more convenient since the nature of *true* tourists (those who spend the night at the destination) and that of same-day visitors is different. In our view, individuals who visit a destination decide either to stay overnight or not. As such, zeros can be regarded as *genuine* zeros. Therefore, contrary to other situations like cultural participation (e.g. Ateca-Amestoy and Prieto-Rodríguez, 2013) in which we need to differentiate between 'non-attendants' and 'goers', in this context there might not be scope for *potential participants*. What is more, Min and Agresti (2005) listed some advantages of the hurdle model compared with the zero-inflated based on simulation[21].

[18] Authors like Alegre and Pou (2006) have transformed LOS into a binary variable with two categories (more and less than a week). This has the disadvantage that the *cutoff* is arbitrary.

[19] See for instance Lambert (1992) on industrial processes and Greene (1994) on credit defaults.

[20] This model is called the 'hurdle' model because for observing a positive value of the dependent variable it is necessary to cross the first hurdle (participation equation). See Mullahy (1986), Yen and Adamowicz (1994) and Rose et al. (2006) on separate modeling of participation and usage.

[21] It works well in both zero-inflation and zero-deflation situations and it is easier to fit since it separately handles the count and the zero parts.

The hurdle model is specified as follows:

a) *Participation equation*: we define a latent participation variable (d_i^*) that is explained by a set of explanatory variables Z_i in the following way:

$$d_i^* = Z_i'\gamma + u_i \tag{1.5}$$

where u_i is a random error term that follows a logistic distribution, which results in the Logit model. The observation mechanism assigns $d_i = 1$ if $d_i^* > 0$ and $d_i = 0$ otherwise. The probability of being a tourist is:

$$Prob(d_i = 1|Z_i) = P(d_i^* > 0) = \frac{e^{Z_i'\gamma}}{1 + e^{Z_i'\gamma}} \tag{1.6}$$

b) *Intensity equation:* the positive values of the dependent variable come from a zero-truncated count data model[22].

The hurdle model is estimated by Maximum Likelihood. Assuming that the error terms of the binary and the truncated model are uncorrelated, the log likelihood function is the sum of the log likelihoods of the two parts. Therefore, maximizing the hurdle log likelihood is equivalent of maximizing both log likelihood functions separately[23].

Alternatively, we have considered the possibility of correlated errors to control for potential endogenous selection. However, this model relies on strong distributional assumptions about the structure of the correlation (Papadopoulos and Santos-Silva, 2012). Although the correlation parameter is theoretically identified, in practice it cannot be accurately estimated (Smith and Moffatt, 1999)[24]. In addition, recent evidence by Drukker (2017) proves that the two-part model is robust to endogenous selection[25].

This two-part model has been widely applied in the economic literature, especially in health (e.g. Sarma and Simpson, 2006) and environmental economics (e.g. Bilgic and Florkowski, 2007). However, it has not been employed in tourism to date.

Initially, we assume that our dependent variable "total number of nights spent at the destination (LOS)" follows a Poisson distribution whose conditional mean and variance are given by:

$$E(LOS_i|X_i) = Var\,(LOS_i|X_i) = e^{X_i\beta} = \lambda_i, \tag{1.7}$$

[22] Truncated count models are the discrete counterparts of truncated and censored models for continuous variables. Truncation at zero is the most common form. See Gurmu (1991) and Gurmu and Trivedi (1992).
[23] The estimation can be also jointly conducted by Non-Linear Least Squares, but the estimates are generally less precise (Papadopoulos and Santos-Silva, 2012).
[24] As Winkelmann (2004) argues, the interpretation of the latent demand in a hurdle model with correlated errors is not straightforward.
[25] This author demonstrates that the effect of a given covariate in a two-part model can be consistently estimated without the necessity of imposing the conditional mean independence property. Like the identification of the treatment effect on the treated (ATET), the two-part model handles endogenous selection because the parameters are recoverable from the data and the unobservable parameter is multiplied by zero.

where X_i is a vector of covariates that explains the length of the stay and *i* = 1,...,*N*, indexes the *N* observations in the sample. We explicitly assume that there is a constant term in the model.

One of the main limitations of the Poisson regression model is that it imposes the conditional mean and variance to be equal (equidispersion property). This assumption is quite restrictive and commonly violated in applied work, generating the *overdispersion* problem (i.e. the conditional variance exceeds the conditional mean). Under these circumstances, inefficiencies arise, and inference based on standard errors is not valid (Wang and Famoye, 1997)[26]. Since the data generally exhibits *overdispersion*, researchers usually look for alternatives to the Poisson model.

Table 1.2 shows the descriptive statistics of our dependent variable (LOS).

Variable	N	Mean	SD	Min	Max
LOS	19,111	4.313	3.833	0	30

Table 1.2. - Descriptive statistics of the dependent variable (LOS)

Since the mean of the length of stay (LOS) is 4.31 whereas its variance is 3.83^2= 14.66, it seems there is an *overdispersion* problem, which prevents us from using the Poisson regression model (this is tested formally later).

Cameron and Trivedi (2009) indicate that the *overdispersion* problem arises due to the presence of unobserved heterogeneity, suggesting the need for a new specification in which the error term adequately represents unobservable or omitted variables. The econometric literature has proposed several alternatives to accommodate unobservable heterogeneity. The most common way to deal with *overdispersion* is to introduce multiplicative randomness (v) in the Poisson regression model so that:

$$LOS_i \sim Poisson\ (LOS_i|\lambda_i\ v) \tag{1.8}$$

We now present two alternative models depending on the assumption about the distribution of the unobserved heterogeneity term (v).

5.2.1. The Negative binomial model: a Poisson-gamma mixture.

Suppose we specify v such that $E(v) = 1$ and $Var(v) = \sigma^2$. Therefore, the first two moments of the dependent variable are given by:

$$E(LOS_i|X_i) = \ e^{X_i\beta+v} = \lambda_i h_i \tag{1.9}$$

$$Var\ (LOS_i|X_i) = \ \lambda_i\ (1 + \lambda_i\ \sigma^2) \tag{1.10}$$

[26] In the presence of the *overdispersion* problem, the Poisson model tends to *under*predict the frequency of zeros (Cameron and Trivedi, 2009) and t-statistics are grossly inflated. Moreover, in settings that involve truncation or censoring, *overdispersion* leads to inconsistency.

where $h_i = exp(v)$ is assumed to have a one parameter gamma distribution, $G(\theta, \theta)$ with mean 1 and variance $1/\theta = \alpha$ (see Greene, 2005), so that its density function is:

$$f(h_i) = \frac{\theta^\theta \exp(-\theta h_i) h_i^{\theta-1}}{\Gamma(\theta)}, \qquad h_i \geq 0, \qquad \theta > 0 \tag{1.11}$$

In the particular case that $v \sim Gamma\ (1, \alpha)$, we obtain the Negative Binomial (NB) model (also known as Poisson-gamma mixture model)[27], which has been shown to be more flexible and suitable for empirical research (Gurmu and Trivedi, 1992; Winkelmann and Zimmermann, 1995)[28].

After integrating h_i out of the joint distribution, the probability mass distribution of the NB is:

$$Prob\ (Y = y_i | X_i) = \frac{\Gamma\ (\theta + y_i) r_i^\theta (1 - r_i)^{y_i}}{\Gamma(1 + y_i) \Gamma(\theta)}, \tag{1.12}$$

where $\Gamma(.)$ denotes the gamma integral that specializes to a factorial for an integer argument, $y_i = 0,1, ...,$ $\theta > 0$ and $r_i = \theta/(\theta + \lambda_i)$.

The introduction of latent heterogeneity induces *overdispersion* while preserving the conditional mean as $E(v) = 1$:

$$E(LOS_i | X_i) = \lambda_i \tag{1.13}$$

$$Var(LOS_i | X_i) = \lambda_i (1 + (1/\theta)\, \lambda_i) = \lambda_i (1 + \alpha \lambda_i) \tag{1.14}$$

where $1/\theta = \alpha = Var\ (h_i)$

The Negative Binomial distribution can be seen as a more general case that collapses to the Poisson model if $\alpha = 0$. An *overdispersion* test consists on testing the null of $\alpha = 0$ (equidispersion) against the alternative $\alpha \neq 0$[29].

Cameron and Trivedi (1986) suggested a reparametrization of the conditional variance so that it depends on a parameter P:

$$Var\ (LOS_i | X_i) = \lambda_i\ (1 + \alpha \lambda_i^{P-1}) \tag{1.15}$$

When P takes the values 1 and 2 we obtain the well-known NB1 and NB2 models (Gurmu and Trivedi, 1996)[30]. The former specifies a linear variance function whereas the latter considers a quadratic variance function. Both variants of the model are easily estimated

---

[27] The regression model is developed in detail in Hausman et al. (1984), Cameron and Trivedi (1986, 1998, 2005), Winkelmann (2008) and Greene (2008). We refer the reader to these authors for further details.
[28] The NB model is the most common specification for count data (Hilbe, 2007).
[29] This test is made using a modified likelihood ratio test to be valid asymptotically. Specifically, the distribution under the null is not χ_1^2 but a 50-50 mixture of χ_1^2 and χ_0^2 (Gutierrez et al., 2001).
[30] We adopt the same terminology as Cameron and Trivedi (1998). The NB1 model is also known as the constant dispersion model while NB2 is called the mean dispersion model. See Cameron and Trivedi (2013, p.80-89) for further details.

by Maximum Likelihood. The main drawback of these models is that the form of the conditional variance is exogenously imposed by the researcher.

Cameron and Trivedi (1998, p.73) also note that other exponents apart from 1 and 2 in the conditional variance would be possible. By replacing θ with $\theta\lambda_i^{2-P}$ in the probability mass function, we obtain the NBP model, whose probability mass function is then given by:

$$Prob\ (Y = y_i|X_i) = \frac{\Gamma\left(\theta\,\lambda_i^{2-P}+y_i\right)s_i^{\theta\,\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta\,\lambda_i^{2-P})}, \tag{1.16}$$

$$y_i = 0,1,\dots; \qquad s_i = \frac{\lambda_i}{\lambda_i+\theta\,\lambda_i^{2-P}} \tag{1.17}$$

Greene (2008) suggests that since the NBP model estimates the parameter P endogenously, this model is likely to be the most suitable among the negative binomial family. Unlike the Poisson maximum likelihood estimator, NB2 is not consistent if the variance specification is incorrect. Although a quadratic conditional variance (NB2) often works well in empirical works and is the most used, it may be badly specified in case the true P is higher than 2. Cameron and Trivedi (1986, 1998) argue that “NB2 is favored by econometricians whereas NB1 is extensively used by statisticians”, but they do not state a preference for one or another.

Because of these reasons, in our empirical model we estimate the general NBP model for explaining LOS in the intensity equation, conditional on being a *true* tourist (participant). In case the estimated value of P is 1 or 2, the NBP model reduces to the classical NB1 and NB2. As LOS is a positive variable, it is necessary to truncate the distribution of the dependent variable. Therefore, we model the intensity equation in terms of a Zero Truncated Negative Binomial P Model (hereafter ZTNBP).

$$LOS^*|X_i \sim ZTNBP \tag{1.18}$$

Its truncated probability mass function is given by dividing the probability function by $Prob\ (y_i = 0|X_i)$:

$$Prob\ (Y = y_i|\ y_i\ > 0, X_i) = \frac{\mathrm{Prob}\ (\mathrm{Y}=\mathrm{y}_i|\mathrm{X}_i)}{\mathrm{Prob}\ (\mathrm{y}_i > 0|\mathrm{X}_i)} = \frac{\frac{\Gamma\left(\theta\,\lambda_i^{2-P}+y_i\right)s_i^{\theta\,\lambda_i^{2-P}}(1-s_i)^{y_i}}{\Gamma(1+y_i)\Gamma(\theta\,\lambda_i^{2-P})}}{1-(1+\alpha\lambda\mathrm{i})^{-\alpha^{-1}}} \tag{1.19}$$

being $y_i = 0,1,\dots; s_i = \frac{\lambda_i}{\lambda_i+\theta\,\lambda_i^{2-P}}$, and $Prob\ (y_i > 0|X_i) = 1-(1+\alpha\lambda_i)^{-\alpha^{-1}}$

The estimation of the ZTNBP model is done by Maximum Likelihood (Greene, 2008). Nonetheless, as the log-likelihood function to maximize is not globally concave and there is no certainty of a unique maximum, the estimates of the truncated NB2 model are normally used as starting points. An application of this general ZTNBP model can be found in Farbmacher (2013).

The reason why we truncate the distribution after having introduced the latent heterogeneity (v) as gamma distributed is not innocuous. To have a closed-form solution of the truncated model based on the NB distribution, it is required to perform the mixing first. This implies that the assumption of independence between v and X only necessarily holds *before* the integration and for the whole population (see Santos-Silva (2003) and Farbmacher (2013) for a discussion on this). Nonetheless, in a hurdle model, the population of interest is the positive counts. Therefore, for consistent estimators we require the covariates and the unobservable heterogeneity to be orthogonal in the truncated population. The problem with the negative binomial model is that independence in the actual population can rule out independence in the truncated one. Consequently, another alternative is to first truncate the distribution and then introduce the mixing over it. This is what we develop below.

5.2.2. The Poisson lognormal mixture model

Instead of assuming that the multiplicative randomness (v) follows a gamma distribution, another option is to suppose it is normally distributed with zero mean and σ standard deviation (Hellström and Nordström, 2008). Under this assumption, the mixing can be done over the zero-truncated Poisson (i.e. unobserved heterogeneity is introduced after truncation).

The Zero-Truncated Poisson Log-Normal model (ZTPN), conditioning on both X_i and v, is given by:

$$Prob\ (Y = y_i | y_i > 0, X_i, v) = \frac{\exp(-h_i\lambda_i)(-h_i\lambda_i)^{y_i}}{\{1-\exp(-h_i\lambda_i)\}\, y_i!},$$
$$h_i\lambda_i = exp\ (X_i\beta + \sigma v),$$
$$v \sim N(0,1) \qquad (1.20)$$

The density of y_i conditioning only on X_i is then:

$$Prob\ (Y = y_i |\ y_i > 0, X_i)\ = \int_{-\infty}^{\infty} \mathrm{Prob}\ (\mathrm{Y} = \mathrm{y}_i | y_i > 0, \mathrm{X}_i, v)\ \phi(v) dv \qquad (1.21)$$

As it happens in the NB2 model, the conditional variance in the log normal model is quadratic in the conditional mean so it accounts for *overdispersion* in the same way as the NB2 model does (Greene, 2009). The integral in the log likelihood function does not exist in closed form, so the estimation needs to be conducted by Gauss-Hermite quadrature using the BHHH estimator after having reparametrized the log likelihood following Butler and Moffitt (1982).

Greene (2009) argues that the Poisson-log normal model seems to be a more natural specification than the Poisson-gamma mixture. The reason is that if v captures unobserved heterogeneity across the sample, then the normality of v can be established by central limit theorems (Winkelmann, 2008). In line with this, some authors also point out that the normal distribution would be a preferable alternative for the unobserved heterogeneity instead of the traditional gamma (Riphahn et al., 2003; Winkelman, 2004). In any case, which of them is the most suitable model needs to be tested empirically.

6. RESULTS

Before discussing the parameter estimates, we need to first choose which of the two alternatives for the intensity equation fits the data best. The Vuong test (Vuong, 1989) is the most used for statistically discriminate between non-nested models. However, ZTNBP and ZTPN are overlapping models since they collapse to Zero Truncated Poisson in case $\alpha = 0$ and $\sigma = 0$ respectively. Furthermore, the standard use of the Vuong test has been heavily criticized. On the one hand, this test is based on the mean and the standard deviation of a statistic constructed as the deviations in the individual contributions to the likelihood function, assuming it is normally distributed. However, in some applications this statistic is non-normal (Wilson, 2015). On the other hand, the statistic possesses a standard normal distribution only if the variance of the log-likelihood ratio between the two models is different from zero. This does not hold when both models are observationally equivalent, and their pseudo-true densities are identical (Schennach and Wilhelm, 2016). Because of these reasons, we employ instead the HPC test proposed by Santos-Silva, Tenreyro and Windmeier (Santos-Silva et al., 2015). These authors develop a testing procedure based on Davidson and MacKinnon's seminal work (Davidson and MacKinnon, 1981), which basically discriminates between model A and model B by checking whether the conditional expectation of the dependent variable under the alternative hypothesis outperforms the conditional mean under the null hypothesis[31]. If so, we reject the null as the alternative improves the prediction of the outcome.

Table 1.3 presents the results of the HPC test. As Santos-Silva et al. (2015) suggest, we reverse the roles of the null and the alternative so that model choice does not depend on which one you compare against the other. The test indicates that the ZTNBP model fits our data best and it is thus the chosen one.

Model comparison	t-Statistic (p-value)	Selected model
ZTNBP vs ZTPN	-3.187 (0.99)	ZTNBP
ZTPN vs ZTNBP	4.603 (0.00)	ZTNBP

Table 1.3.- Santos-Silva, Tenreyro and Windmeier HPC test for model choice

Table 1.4 presents the estimation results of the proposed hurdle count data model. The first column shows the estimates of the Logit model for the participation decision, whereas the second and the third report the estimates of the Zero Truncated Negative Binomial P model (ZTNBP) and the Zero Truncated Poisson Log Normal model (ZTPN), respectively[32]. The variables *accom_price* and the ones for the type of accommodation (*hotel, hostel, rural* and *private*) are only considered in the intensity equation, since they

[31] The HPC test differs from the P and C tests developed by Davidson and MacKinnon (1981) in the sense that it accounts for heteroskedasticity in the auxiliary regression. Each observation is weighted in the moment condition as a percentage difference between the two conditional means.

[32] We used *ztnbp* and *ztpnm* modules in Stata 14.0, outlined in Farbmacher (2011). The estimates are robust to the number of quadrature points (Q). The reported results used Q=30, but we have also estimated it using Q=20 and Q=40. The parameters remain unchanged (available upon request).

are only observed for those who stay overnight. The α and σ parameters, which account for *overdispersion*, are statistically significant at the 1 percent level. This provides evidence of the necessity of accounting for unobserved effects when modelling LOS in comparison to the basic Poisson model (Cameron and Trivedi, 1998).

The estimated parameter P in the ZTNBP model is 3.65 and it is statistically significant. This figure is far from the imposed 1 and 2 that correspond to the ZTNB1 and ZTNB2 alternatives. This supports the necessity of allowing the model to be flexible in the structure of the conditional variance.

Starting with the sociodemographic characteristics, gender is not significant in either the participation or the intensity equations. This is in line with the literature (Martínez-García and Raya 2008; Brida et al. 2013). In the same vein, age is not significant for explaining the overnight stay decision at conventional levels (only at the 10 percent level). It is, however, positively related with the number of days spent, although at a decreasing rate according to the negative sign of the squared term. This is consistent with Fleischer and Pizam (2002), who find a concave relationship between age and length of stay. Regarding labor status, we set *self-employed* as the reference category. The estimates indicate that these individuals display the largest probability of staying overnight. However, retired people (*retired*), students (*student*), unemployed people (*unemployed*) and housewives (*housewife*) stay longer than self-employed individuals. This falls in line with previous findings that report that full-time working people stay fewer nights (e.g. Hellstrom and Nordström, 2008). As for the education level, compared to primary education (reference category), visitors with secondary and university studies (*secondary* and *higheduc* respectively) have a higher probability of staying overnight in Asturias. Nonetheless, the education level is not significant for explaining the intensity of the stay. It is important to highlight here that these last three variables (age, educational level and labor status) may also account for income differences among individuals. With reference to the place of residence, we differentiate between people who live in Spain and those who do not with the dummy variable *foreign*. Foreign individuals do not show a statistically different probability of an overnight stay relative to Spaniards. However, conditional on having decided to stay, they stay for longer.

Regarding the motivations to visit Asturias, we find that tranquility (*tranquility*), the natural environment (*natural*), and its oceanic weather (*climate*) positively influence both the length of the stay and the probability of staying overnight (relative to reporting heritage). The latter is in line with Nicolau and Más (2009) and Barros et al. (2010), who document that climate as a pull factor is associated with longer stays. However, those who report that their main reason for travelling was either novelty seeking (*novelty*) or Asturian gastronomy (*gastronomy*) have a higher likelihood of staying overnight but do not stay for longer.

Contrary to our expectations, the daily price of accommodation per person (*accom_price*) is not found to be significant in the intensity equation. One possible explanation is that the model controls for the type of accommodation in the regression, which might implicitly reflect price differences.

We now move to the effect of tourist's knowledge about the destination. The estimates for first-time visitors (*first*) are positive and significant in both equations, implying that those who have never been to Asturias stay for longer than repeat visitors. This positive relation can be explained in terms of the willingness to widely get acquainted with the destination (Nicolau et al. 2018). A similar explanation is that repeat visitors have already explored the destination in previous visits, so they are less incentivized to have extended stays. This is consistent with previous research that has documented important differences in travel patterns between first-time and repeat visitors (Li et al., 2008). The number of visits during the year (*num_vis*) is negatively related to the likelihood of staying overnight, whereas it is not significant in the intensity equation. Conversely, those who declare that a positive previous experience at the destination is the main reason for coming back (*experience*) exhibit a higher probability of an overnight stay and longer stays. For some tourists, if their previous experience was satisfactory, a good risk-reduction method is to return to the same destination and stay for longer periods.

Those who state they have seen some type of advertisement (*advert*) show a higher probability of staying overnight and stay for longer, in line with previous evidence (e.g. Barros et al., 2008). This is not surprising, since the experience nature of tourism induces people to build indirect experience from advertising contents such as texts, images or videos (Park and Nicolau, 2015). Indeed, tourism advertising is one of the main external information sources, as it both consciously and unconsciously affects consumer decision-making (Woodside and King 2001). In the same way, the recommendation of the destination from friends or relatives (*recommend*) also increases the likelihood of staying overnight. Surprisingly, *recommend* is not significant for explaining the number of stays. Although this is contrary to our expectations, this result matches those by Gokovali et al. (2007).

Regarding distance to origin, this variable is significant in the participation and the intensity equations, revealing a positive relationship between distance and length of stay (although at a decreasing rate). This is in line with Gronau (1970) and Nicolau et al. (2018). Concerning the chosen mode of transport, the longer it takes to the tourist to reach the destination, the less time she can then allocate to staying there. As our analysis is *conditional* on trip characteristics, tourists travelling by plane arrive sooner and, consequently, may stay for longer. However, faster modes of transport are more expensive so, given the budget constraint, the individual would have less money to spend at the destination and may stay for shorter. Setting *car* as the reference category, this double reasoning may justify why none of the means of transport is significant in the intensity equation. The trade-off between monetary and time savings might cancel out the differences across the modes of transport. Nonetheless, travelling by train or by plane positively affects the likelihood of an overnight stay. Conversely, travelling by bus reduces the probability of an overnight stay. This may account for the fact that most same-day visitors come to the destination by bus.

Dependent variable: LOS	Participation	Intensity	
Independent variables	Logit	ZTNBP	ZTPN
male	-0.0875	-0.0064	-0.0061
	(0.059)	(0.011)	(0.010)
age	0.0297*	0.0180***	0.0207***
	(0.017)	(0.003)	(0.003)
agesq	-0.0003	-0.0001***	-0.0001***
	(0.000)	(4.57e-05)	(4.07e-05)
housewife	-0.5183***	0.0776**	0.0906***
	(0.154)	(0.033)	(0.030)
retired	-0.3100*	0.0789**	0.0891***
	(0.179)	(0.039)	(0.034)
employed	-0.2051**	-0.0171	-0.0198
	(0.086)	(0.015)	(0.014)
student	-0.4648***	0.0609**	0.0753***
	(0.144)	(0.029)	(0.027)
unemployed	-0.6361***	0.1339***	0.0898**
	(0.191)	(0.046)	(0.043)
secondary	0.2311**	0.0329	0.0269
	(0.106)	(0.025)	(0.021)
higheduc	0.3983***	0.0108	0.0141
	(0.106)	(0.024)	(0.021)
foreign	-0.1692	0.2099***	0.1481***
	(0.131)	(0.031)	(0.024)
natural	1.4149***	0.0547**	0.0473*
	(0.109)	(0.027)	(0.025)
novelty	1.3746***	-0.0050	-0.0120
	(0.095)	(0.025)	(0.022)
tranquility	1.2438***	0.0972*	0.1109*
	(0.313)	(0.058)	(0.058)

Dependent variable: LOS	Participation	Intensity	
Independent variables	Logit	ZTNBP	ZTPN
climate	1.4340***	0.1828***	0.2151***
	(0.418)	(0.054)	(0.051)
gastronomy	0.6610***	-0.0204	-0.0676
	(0.215)	(0.073)	(0.061)
accom_price		-0.0003	0.0001
		(0.000)	(0.000)
first	0.3501***	0.0779***	0.0938***
	(0.079)	(0.014)	(0.013)
num_vis	-0.0266***	0.0005	-0.0034
	(0.003)	(0.003)	(0.003)
experience	1.3585***	0.0533**	0.0493**
	(0.089)	(0.026)	(0.023)
advert	0.3482***	0.0297**	0.0472***
	(0.067)	(0.012)	(0.011)
recommend	1.0665***	-0.0044	-0.0270
	(0.116)	(0.028)	(0.026)
distance	0.0008***	0.0002***	0.0002***
	(0.000)	(3.72e-05)	(3.36e-05)
distancesq	-7.74e-08**	-2.71e-08***	-2.83e-08***
	(3.29e-08)	(3.91e-09)	(3.78e-09)
bus	-1.0888***	-0.0572	-0.0256
	(0.243)	(0.040)	(0.037)
train	1.2759***	0.0412	0.0540
	(0.412)	(0.039)	(0.033)
plane	0.5714***	-0.0053	0.0374
	(0.196)	(0.027)	(0.023)
leisure	-0.8672***	0.1767***	0.2241***
	(0.182)	(0.045)	(0.040)

Dependent variable: LOS	Participation	Intensity	
Independent variables	Logit	ZTNBP	ZTPN
labor	-0.3875*	0.2510***	0.1448***
	(0.218)	(0.070)	(0.053)
family	-0.1984	0.1603***	0.2197***
	(0.210)	(0.050)	(0.043)
party_size	-0.0002	0.0002	-0.0004
	(0.005)	(0.001)	(0.000)
alone	0.6313***	-0.0055	-0.0500
	(0.174)	(0.037)	(0.030)
couple	0.2639***	-0.0365***	-0.0508***
	(0.061)	(0.012)	(0.011)
hotel		-0.3322***	-0.3264***
		(0.027)	(0.023)
hostel		-0.0858**	-0.0539*
		(0.033)	(0.029)
rural		-0.1527***	-0.1381***
		(0.028)	(0.024)
private		0.2716***	0.2239***
		(0.032)	(0.026)
travel_agency	0.8898***	0.1213***	0.1459***
	(0.270)	(0.026)	(0.024)
club_firm	0.6747***	-0.1543**	-0.1135***
	(0.198)	(0.062)	(0.043)
only_ast		0.0907***	0.1294***
		(0.016)	(0.015)
act_tour	1.5424***	0.1481***	0.1873***
	(0.201)	(0.019)	(0.016)
y11	0.6240***	-0.1212***	-0.1012***
	(0.120)	(0.020)	(0.018)
y12	0.1460	-0.1282***	-0.1109***
	(0.107)	(0.021)	(0.019)
y13	0.1176	-0.1328***	-0.1257***
	(0.108)	(0.023)	(0.020)
y14	0.2451**	-0.0848***	-0.0706***
	(0.109)	(0.020)	(0.018)
y15	-0.1078	-0.0290	-0.0330*
	(0.102)	(0.021)	(0.019)
y16	0.1456	-0.0534**	-0.0698***
	(0.112)	(0.024)	(0.019)
t2	0.3751***	0.4000***	0.3887***
	(0.073)	(0.014)	(0.013)
t3	0.3456***	0.1377***	0.1050***
	(0.076)	(0.017)	(0.014)
west	-1.6733***	0.0816***	0.1244***
	(0.083)	(0.016)	(0.015)
capcity	0.0021	-0.0027	0.0025
	(0.183)	(0.027)	(0.025)
east_inner	-1.3624***	0.0591***	0.0533***
	(0.085)	(0.017)	(0.015)
east_coast	-0.9123***	0.1236***	0.1428***
	(0.090)	(0.015)	(0.014)
constant	1.0192**	0.4412***	0.2323***
	(0.443)	(0.102)	(0.089)

α		0.0107***	
		(0.001)	
P		3.6564***	
		(0.118)	
σ			0.4130***
			(0.004)
Observations	19,111	17,478	

Table 1.4. – Parameter estimates of the hurdle model
(Robust standard errors in parentheses)
*** p<0.01, ** p<0.05, * p<0.1

Visiting Asturias for leisure and entertainment (*leisure*) or due to job or study-related issues (*labor*) reduce the probability of an overnight stay. Although this might initially seem counterintuitive, the omitted category here (*other*) includes, among others, sport events or doctor visits, which normally require the visitor to spend at least one night at the destination. Nonetheless, for the case of the number of stays, leisure purposes lengthen the stay, in line with the literature (Gomes de Menezes et al. 2008). Visiting friends or relatives (*family*) and labor-related reasons are also associated with longer stays.

Party size and its composition also matter for explaining LOS, as tourism consumption is usually a social activity in which activities are mainly group-based (Thornton et al. 1997). The fact that only 6.3% of the tourists traveled alone indicates that they generally travel in a couple or in a group. Compared to travelling in a group (*group*), those who come to Asturias alone (*alone*) or in a couple (*couple)* display a higher probability of spending at least a night there. Regarding the length of the stay, there are no statistical differences between travelling alone or in a group, whereas couples are associated with longer stays. The size of the party size is not statistically significant either in the participation or in the intensity equation.

As for the accommodation type, staying at a private dwelling leads to the longest stays. Focusing on formal market-based accommodations, tourists lodged at a rural house (*rural*), a *hostel* or a *hotel* stay for shorter than those who stay on *campsites* (reference category). Being the tourist himself the omitted category, we find that organizing the trip through a travel agency (*travel_agency*) increases both the probability of spending the night at the destination as well as LOS. Booking the trip through a club or a firm (*club_firm*) is also positively related to the decision to stay overnight but reduces the number of expected stays.

As expected, those who only visit Asturias in the current trip (*only_ast*) stay for longer. Taking part in active tourism (*act_tour*) activities exerts a positive impact on both the decision to stay overnight and on the number of stays. Since these tourists engage in outdoor activities that require time to be performed, it makes sense that this market segment exhibits longer stays.

Finally, the regression also controls for temporal and geographical variables. The year dummies mainly reflect the effect of the business cycle, while the geographical ones gather differences in preferences over the territory. Everything else being equal, stays were shorter from 2011 onwards. This may be associated with the fact that during the economic crisis people faced important economic constraints (Smeral and Song 2015). Even if they did not, uncertainties and fears about the near future and labor instability might have urged them to spend more on necessities and less on luxuries to save money (Gunter and Smeral 2016). As for seasonal effects, tourists exhibit both a higher likelihood of an overnight stay and longer stays in the May-August (*t2*) and September-December (*t3*) periods. Regarding geographical preferences, the probability of staying in the surroundings of the capital city is higher than in the east or the west. However, the opposite pattern is observed for the number of stays.

To address potential collinearity concerns, we computed the Variance Inflation Factor (VIF) after running an OLS regression on LOS. All the values are below 10 (min=1.07; max=6.7; mean=2.05), which is usually considered the threshold value to detect multicollinearity problems.

To examine the magnitude of the effects, Table 1.5 reports the relative average marginal effects (AME) on the probability of an overnight stay in the first column, and on the conditional expected number of nights in the second one[33]. They indicate the percentage change in the probability of an overnight stay and the percentage change in the number of stays if there is a marginal change in a given variable, respectively. Specifically, the marginal effect for the participation equation is given by:

$$\sum_{i=1}^{N} \frac{1}{N} \frac{\partial\, P(LOS_i > 0)}{\partial\, Z_{ik}} = \sum_{i=1}^{N} \frac{1}{N}\ \beta_k \frac{e^{Xi\beta}}{(1+ e^{Xi\beta})} \tag{1.22}$$

Following Farbmacher (2013), the relative marginal effects for a ZTNBP model are given by:

$$\sum_{i=1}^{N} \frac{1}{N} \frac{\frac{\partial\, E(LOS_i | LOS_i > 0,\ X_i)}{\partial\, X_{ik}}}{E(LOS_i | LOS_i > 0,\ X_i)} = \beta_k - \frac{r_i m_i \left(Q\beta_k \ln\left(\frac{\left(m_i + e^{Xi\beta}\right)}{m_i}\right) + \frac{m_i Q\beta_k + \beta_k\, e^{Xi\beta}}{m_i + e^{Xi\beta}} - Q\beta_k\right)}{1 - r_i} \tag{1.23}$$

where $Q = (2 - P)$; $m_i = \frac{1}{\alpha}\lambda_i{}^{2-P} = e^{(2-P)X_i\beta - \ln\alpha}$; and $r_i = \left(\frac{m_i}{\lambda_i + m_i}\right)^{m_i}$

The AME are computed for the informative-type variables (*first*, *num_vis*, *experience*, *advert*, *recommend*) and the destination attributes (*natural*, *novelty*, *tranquility*, *climate*, *gastronomy*).

[33] In the participation equation each coefficient indicates the log of the odds ratio (i.e. the ratio of the probability of observing a positive value of LOS divided by the probability of non-participation). In the intensity equation, each coefficient could be interpreted as a semi-elasticity (i.e. relative change in LOS for a unitary change in an explanatory variable) in a standard Poisson or NB model. However, due to the truncation at zero, the derivation of the semi-elasticities requires some extra computation.

Variable	Logit model	ZTNBP
first	2.292***	7.098***
num_vis	-0.180***	0.047
experience	7.724***	4.621**
advert	2.289***	2.706**
recommend	5.492***	-0.406
natural	7.032***	4.984**
novelty	8.636***	-0.462
tranquility	5.600***	8.857*
climate	6.058***	16.657***
gastronomy	3.609**	-1.858

Table 1.5. - Average marginal effects on the participation (1) and intensity (2) equations (in percentage). *** p<0.01, ** p<0.05, * p<0.1

First-time visitors (*first*) show a 2.3% higher probability of staying overnight and are associated with a 7.1% longer stays than repeat visitors. A unitary change in the number of visits during the year (*num_vis*) reduces the probability of becoming a tourist by almost 0.2%. Previous experience (*experience*) and having seen advertising (*advert*) increases the expected number of stays by 4.6 and 2.7%, respectively. Moreover, those who had previously visited the destination and those who declare that advertising has persuaded them to visit it display 7.7 and 2.3% higher probability of an overnight stay. Finally, recommendation from friends or relatives (*recommend*) only influences the probability of being a tourist (5.5% higher), but it does not affect the expected stay.

Regarding the destination attributes, novelty seeking (*novelty*) stands as the most relevant factor, increasing the likelihood of an overnight stay by 8.6%. The natural environment (*natural*) is another relevant attribute that encourages visitors to stay overnight. With reference to the expected stay, the Asturian climate (*climate*) and its tranquility (*tranquility*) are associated with 16.6 and 8.8% longer stays, respectively.

Based on the average relative marginal effects reported above, we can conclude that previous experience (*experience*) is the information source that is associated with the highest probability of an overnight stay and that has the largest impact on LOS, *ceteris paribus*. This implies that there is no better source of information than having a positive experience in the past. To a lesser extent, advertising also positively contributes to lengthen the stay.

Overall, we find that the different explanatory variables considered do not have the same effect on the probability of staying overnight (participation equation) and on the number of days spent (intensity equation). This highlights the relevance of distinguishing between tourists and same-day visitors, which is one of the novel aspects of this paper. Furthermore, our regression framework controls for a wide range of sources of observable heterogeneity among tourists, allowing us to properly isolate the effects of the different information sources about the destination on LOS.

7. CONCLUSIONS

Using a hurdle count data model, this study has examined the determinants of both the decision to stay overnight at a destination and the length of the stay. The determinants of tourists' LOS have been widely analyzed in the literature but, to the best of our knowledge, no studies have considered the different nature of same-day visitors and tourists using a hurdle count data approach. Moreover, in this research we have devoted particular attention to the effect of information about the destination on LOS, once having controlled for a large spectrum of sources of observable heterogeneity. Specifically, we have studied the effect of being a first-time visitor, having seen any type of advertising, visiting because of a recommendation or due to having had a pleasant experience on the length of the stay.

From a methodological perspective, we have proposed two alternative specifications for the intensity equation in the hurdle model. On the one hand, a Zero-Truncated Negative Binomial P model, which seems to be the best alternative among the Negative Binomial family because, at the same time it handles the *overdispersion* problem, it also endogenously estimates the structure of the conditional variance through the parameter P. On the other hand, we have estimated a Zero-Truncated Poisson Log-Normal model, which specifies a normal distribution for the unobserved heterogeneity instead of the gamma distribution assumed for the ZTNBP model. The HPC test clearly indicates the better adequacy of the ZTNBP model, at least in our data.

As our case study, we have used a pooled cross-sectional database of visitors (including same-day visitors and tourists) to Asturias (Spain) for the period 2010-2016. Our empirical model is based on a conditional demand function for time given individual characteristics. Our results show that first-time visitors, those who had seen any type of advertising about the destination, those who declare that having had a positive experience at the destination or that someone recommended it to them as the main reason for visiting have a higher probability of staying overnight. The same pattern holds for the expected number of stays, except for the fact that a recommendation is not significant for explaining LOS. Quantitatively, we find that first time visitation leads to the highest increase in LOS. Tourists seem to rely relatively more on their personal past experiences than on recommendations from friends or relatives.

Apart from the effects of the information sources, we have also considered other relevant variables in our empirical model. To summarize, we find that the sociodemographic profile appears to be less relevant for LOS in comparison to previous studies once we control for a large number of trip-related characteristics. In this sense, the Asturian climate and the aim of performing active tourism emerge as two relevant factors that increase the length of the stay. Regarding the chosen mode of transport, travelling by train or by plane positively increase the likelihood of an overnight stay in comparison to travelling by car. With reference to the accommodation, staying at a private apartment is associated with the longest stays, followed by campsites. Our results also show that LOS is not significantly related to travel party size.

Regarding policy implications, the identification of the drivers of LOS seems to be a relevant issue since revenues from tourism are directly related to LOS. The results provided in this study about the linkages between a wide set of factors and tourists' length of stay can thus help local policy makers i) to improve the promotion campaigns and ii) to develop proper strategies to adapt the tourism products to the desires of tourists, focusing on those who stay for longer periods. In this sense, the estimates of the relative marginal effects reveal that those who had seen advertising campaigns display a high likelihood of an overnight stay and longer-period stays. Consequently, policy makers should try to broadcast promotional campaigns at wider audiences. Moreover, a pleasant past experience seems to be another relevant information source. Hospitality entrepreneurs thus need to provide tourists with an enjoyable stay to encourage them to come back in the near future. Additionally, the natural environment of the destination, its climate and its tranquility are also three characteristics of Asturias that pull visitors to stay for longer. Policy makers should reinforce these appealing features, highlighting the "green tourism" brand.

The limitations of this study include the fact that our analysis of LOS is conditional on having decided to visit Asturias, a decision that we cannot model with the data we have. We also lack information on income, which does not allow us to examine how the length of the stay relates to it. Furthermore, our sample does not allow the extrapolation of the results to other regions except for those sharing similar characteristics with Asturias. Nonetheless, the proposed methodology can be applied to any destination.

REFERENCES

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

Alegre, J., Mateo, S., and Pou, L.L. (2011). "A latent class approach to tourists' length of stay". *Tourism Management*, 32(3), 555–563.

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

Alén, E., Nicolau, J.L, Losada, N. and Domínguez, T. (2014). "Determinant factors of senior tourist's length of stay". *Annals of Tourism Research*, 49, 19-32.

Ateca-Amestoy, V. and Prieto-Rodríguez, J. (2013). "Forecasting accuracy of behavioral models for participation in the arts". *European Journal of Operational Research*, 229(1), 124-131.

Barros, C.P., Butler, R. and Correia, A. (2010). "The length of stay of golf tourism: a survival analysis". *Tourism Management*, 31, 13-21.

Barros, C.P., Correia, A. and Crouch, G. (2008). "Determinants of the length of stay in Latin American tourism destinations". *Tourism Analysis*, 13, 329-340.

Barros, C.P. and Machado, L.P. (2010). "The length of stay in tourism". *Annals of Tourism Research*, 37(3), 692-706

Bell, F.W. and Leeworthy, V.R. (1990). "Recreational demand by tourists for saltwater beach days". *Journal of Environmental Economics and Management*, 18, 189-205.

Bhat, C.R., Astroza, S. and Bhat, A.C. (2016). "On allowing a general form for unobserved heterogeneity in the multiple discrete-continuous probit model: formulation and application to tourism travel". *Transportation Research Part B*, 86, 223-249.

Bieger, T. and Laesser, C. (2004). "Information sources for travel decisions: toward a source process model". *Journal of Travel Research,* 42, 357–371.

Bilgic, A. and Florkowkski, W.J. (2007). "Application of a hurdle negative binomial count data model to demand for bass fishing in the southeastern United States". *Journal of Environmental Management*, 83, 478-490.

Brida, J.G., Meleddu, M., and Pulina, M. (2013). "Factors influencing length of stay of cultural tourists". *Tourism Economics*, 19, 1273-1292.

Butler, J. and Moffitt, R. (1982). "A computationally efficient quadrature procedure for the one factor multinomial probit model". *Econometrica*, 50, 761-764.

Cameron, A.C. and Trivedi, P.K. (1986). "Econometric models based on count data. Comparisons and applications of some estimators and tests". *Journal of Applied Econometrics*, 1(1), 29–53.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression analysis of count data*. Nueva York: Cambridge University Press.

Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics. Methods and applications*. NY: Cambridge University Press.

Cameron, A. C., and Trivedi, P.K. (2009). *Microeconometrics using Stata.* Stata Press, College Station, Texas, USA.

Cameron, A.C. and Trivedi, P.K. (2013). *Regression analysis of count data*, 2nd edition, Econometric Society Monograph, Cambridge University Press.

Davidson, R. and MacKinnon, J.G. (1981). "Several tests for model specification in the presence of alternative hypotheses". *Econometrica*, 49, 781-793.

Deaton, A.S., and Muellbauer, J. (1980). *Economics and consumer behaviour*. Cambridge: Cambridge University Press.

Decrop, A. and Snelders, D. (2004). "Planning the summer vacation. An adaptable process". *Annals of Tourism Research*, 31(4), 1008–1030.

Drukker, D.M. (2017)."Two-part models are robust to endogenous selection". *Economics Letters*, 152, 71-72.

Dubin, J.A., and McFadden, D. (1984). "An econometric analysis of residential electric appliance holdings and consumption". *Econometrica*, 52(2), 345–362.

Farbmacher, H. (2011). "Estimation of hurdle models for overdispersed count data". *The Stata Journal*, 11(1), 82-94.

Farbmacher, H. (2013). "Extensions of hurdle models for overdispersed count data". *Health Economics*, 22, 1398-1404.

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

Fleischer, A. and Pizam, A. (2002). "Tourism constraints among Israeli seniors". *Annals of Tourism Research*, 29(1), 106–123.

Fodness, D. and Murray, B. (1997). “Tourist information search”. *Annals of Tourism Research*, 24, 503–523.
Gitelson, R. and Kerstetter, D. (1994). "The influence of friends and relatives in travel decision-making". *Journal of Travel and Tourism Marketing*, 3(3), 59-68.
Gokovali, U., Bahar, O. and Kozak, M. (2007). “Determinants of length of stay: a practical use of survival analysis”. *Tourism Management*, 28(3), 736–746.
Gomes de Menezes, A. and Moniz, A. (2011). “Determinants of length of stay. A parametric survival analysis”. *Tourism Analysis*, 16, 509-524.
Gomes de Menezes, A., Moniz, A. and Cabral, J. (2008). “The determinants of length of stay of tourists in the Azores”. *Tourism Economics*, 14(1), 205-222.
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. (1994). “Sample selection in credit-scoring models”. *Japan and the World Economy,* 10, 299-316.
Greene, W. (2005). “Functional form and heterogeneity in models for count data”. *Foundations and Trends in Econometrics*, 1 (2), 113-218.
Greene, W. (2008). “Functional forms for the negative binomial model for count data”. *Economic Letters*, 99, 585-590.
Greene, W. (2009). "Models for count data with endogenous participation". *Empirical Economics*, 36, 133-173.
Grigolon, A.B., Borgers, A.W.J., Kemperman, A.D.A.M. and Timmermans, H.J.P. (2014). "Vacation length choice: a dynamic mixed multinomial logit model". *Tourism Management*, 41, 158-167.
Gronau, R. (1970). “The effect of travelling time on the demand for passenger transportation”. *Journal of Political Economy*, 78(3), 377‐394.
Gunter, U. and Smeral, E. (2016). “The decline of tourism income elasticities in a global context”. *Tourism Economics,* 22 (3), 466–83.
Gurmu, S. (1991). “Test for detecting overdispersion in the positive Poisson regression model”. *Journal of Business and Economics Statistics*, 9, 215–22.
Gurmu, S. and Trivedi, P.K. (1992). “Overdispersion tests for truncated Poisson regression models”. *Journal of Econometrics*, 54(1), 347–370.
Gurmu, S. and Trivedi, P.K. (1996). “Excess zeros in count models for recreational trips”. *Journal of Business and Economic Statistics*, 14(4), 469–477.
Gutiérrez R.G., Carter S.L. and Drukker D.M. (2001). “On boundary-value likelihood-ratio tests”. *Stata Technical Bulletin*, 60, 15–18.
Hanemann, W.M. (1984). “Discrete/continuous models of consumer demand”. *Econometrica*, 52(3), 541–561.
Hausman, J.A., Hall, B.H., and Griliches, Z. (1984). “Econometric models for count data with an application to the patents – R and D relationship”. *Econometrica*, 52, 909–938.
Hellerstein, D. and Mendelsohn, R. (1993). “A theoretical foundation for count data models”. *American Journal of Agricultural Economics*, 75(3), 604-611.
Hellström, J. (2006). “A bivariate count data model for household tourism demand”. *Journal of Applied Econometrics*, 21, 213–226.
Hellström, J. and Nordström, J. (2008). "A count data model with endogenous household specific censoring: the number of nights to stay". *Empirical Economics*, 35, 179-192.
Hilbe, J. M. (2007). *Negative binomial regression*, New York: Cambridge.
Kerstetter, D. and Cho, M.H. (2004). “Prior knowledge, credibility and information search”. *Annals of Tourism Research*, 31, 961–985.
Kiefer, N. (1988). “Economic duration data and hazard functions”. *Journal of Economic Literature*, 26, 646–667.
Kim D.Y., Hwang Y.H. and Fesenmaier D.R. (2005). “Modeling tourism advertising effectiveness”. *Journal of Travel Research*, 44(1), 42-49.
Laesser, C. and Crouch, G.I. (2006). “Segmenting markets by travel expenditure patterns: the case of international visitors to Australia”. *Journal of Travel Research*, 44(4), 397–406.
Lambert, D. (1992). “Zero-Inflated Poisson regression, with an application to defects in manufacturing”. *Technometrics*, 34(1), 1-14.
Lancaster, K.J. (1966). “A new approach to consumer theory”. *Journal of Political Economy*, 74, 132–157.
Lew, A.A. and McKercher, B. (2002). “Trip destinations, gateways and itineraries: the example of hong kong”. *Tourism Management*, 23(6), 609-21.
Li, X., Chia-Kuen, C., Hyounggon, K. and Petrick, J.F. (2008). “A systematic comparison of first time and repeat visitors via a two-phase online survey”. *Tourism Management*, 29, 278-293.

Luo, Q. and Zhong, D. (2015). “Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social network sites”. *Tourism Management*, 46, 274-282.
Machado, L.P. (2010). “Does destination image influence the length of stay in a tourism destination?” *Tourism Economics*, 16 (2), 443-56.
Mak, J., Moncur, J. and Yonamine, D. (1977). “Determinants of visitor expenditures and visitor length of stay: a cross-section analysis of US visitors to Hawaii”. *Journal of Travel Research*, 15, 5–8.
Martínez-García, E. and Raya, J.M. (2008). “Length of stay for low-cost tourism”. *Tourism Management*, 29(6), 1064–1075.
Martínez-Roget, F. and Rodríguez-González, X.A. (2006). “Rural tourism demand in Galicia, Spain”. *Tourism Economics*, 12(1), 21-31.
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. (1981). *Econometric models of probabilistic choice*. In C. Manski, & D. McFadden (Eds.), Structural analysis of discrete data (pp. 198–272). Cambridge: MIT Press.
Min, Y. and Agresti, A. (2005). “Random effect models for repeated measures of zero-inflated count data”. *Statistical Modelling*, 5, 1-19.
Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: National Bureau of Economic Research.
Mullahy, J. (1986). “Specification and testing of some modified count data models”. *Journal of Econometrics*, 33, 341–365.
Nicolau, J.L. and Más, F.J. (2009). “Simultaneous analysis of whether and how long to go on holidays”. *Service Industries Journal*, 29 (8), 1077-92.
Nicolau, J.L., Zach, F. and Tussyadiah, I. (2018). "Effects of distance and first-time visitation on tourists’ length of stay". *Journal of Hospitality and Tourism Research*, 42(7), 1023-1038.
Oliveira-Santos, G.E., Ramos, V. and Rey-Maquieira, J. (2015). “Length of stay at multiple destinations of tourism trips in Brazil”. *Journal of Travel Research*, 54(6), 788-800
Oppermann, M. (1995). “Travel life cycle”. *Annals of Tourism Research*, 22(3), 535–552.
Oppermann, M. (1997). “First-time and repeat visitors to New Zealand”. *Tourism Management*, 18, 177–181.
Papadopoulos, G. and Santos-Silva, J.M.C (2012). "Identification issues in some double-index models for non-negative data". *Economic Letters*, 117, 365-367.
Park, S. and Nicolau, J.L. (2015). "Differentiated effect of advertising: joint vs separate consumption". *Tourism Management*, 47, 107-114.
Paul, B.K. and Rimmawi, H.S. (1992). “Tourism in Saudi Arabia: Asir National Park”. *Annals of Tourism Research*, 19, 501–515.
Pollak, R.A. (1969). “Conditional demand functions and consumption theory”. *The Quarterly Journal of Economics*, 83(1), 60–78.
Pollak, R.A. (1971). “Conditional demand functions and the implications of separable utility”. *Southern Economic Journal*, 37, 423–433.
Prebensen, N.K., Altin, M. and Uysal, M. (2015). “Length of stay: a case of northern Norway”. *Scandinavian Journal of Hospitality and Tourism*, 15, 28-47.
Ratner, R.K., Kahn, B. and Kahneman, D. (1999). “Choosing less-preferred experiences for the sake of variety”. *Journal of Consumer Research*, 26(1), 1-15.
Riphahn, R., Wambach, A. and Million, A. (2003). “Incentive effects in the demand for health care: a bivariate panel count data estimation”. *Journal of Applied Econometrics*, 18(4), 387-405.
Roehl, W. and Fesenmaier, D.R. (1992). “Risk perceptions and pleasure travel: an exploratory analysis”. *Journal of Travel Research*, 30, 17-26.
Rose, C., Martin, S., Wannemuehler, K. and Plikaytis, B. (2006). “On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data”. *Journal of Biopharmaceutical Statistics*, 16, 463–481.
Rugg, D. (1973). "The choice of journey destination: a theoretical and empirical analysis". *The Review of Economics and Statistics*, 55(1), 64-72.
Salmasi, L., Celidoni, M. and Procidano, I. (2012). “Length of stay: price and income semi-elasticities at different destinations in Italy”. *International Journal of Tourism Research*, 14, 515–530.
Santos-Silva J.M.C. (1997). “Unobservables in count data models for on-site samples”. *Economics Letters*, 54, 217–220.
Santos-Silva, J.M.C. (2003). "A note on the estimation of mixture models under endogenous sampling". *Econometrics Journal*, 6, 46-52.

Santos-Silva, J.M.C., Tenreyro, S. and Windmeier, F. (2015). "Testing competing models for non-negative data with many zeros". *Journal of Econometric Methods*, 4, 29-46.
Sarma, S. and Simpson, W. (2006). “A microeconometric analysis of Canadian health care utilization”. *Health Economics*, 15, 219-239.
Schennach, S. and Wilhelm, D. (2016). “A simple parametric selection test”. *Journal of the American Statistical Association*, 112(520), 1663-1674.
Silberman, J. (1985). “A demand function for length of stay: The evidence from Virginia Beach”. *Journal of Travel Research*, 23(4), 16-23.
Smeral, E. and Song, H. (2015). “Varying elasticities and forecasting performance”. *International Journal of Tourism Research*, 17(2), 140-150.
Smith, M.D. and Moffatt, P.G. (1999). “Fisher’s information on the correlation coefficient in bivariate logistic models”. *Australian and New Zealand Journal of Statistics*, 41, 315–330.
Taylor, C.E. and Knudson, D.M. (1976). “Area preferences of Midwestern campers”. *Journal of Leisure Research*, 7, 39-48.
Thornton, P., Shaw, G. and Williams, A.M. (1997). "Tourist group holiday decision-making and behavior: the influence of children". *Tourism Management*, 18(5), 287-297.
Thrane, C. (2012). “Analyzing tourists’ length of stay at destinations with survival models: a constructive critique based on a case study”. *Tourism Management*, 33, 126-132.
Thrane, C. (2016). “Students’ summer tourism: determinants of length of stay (LOS)”. *Tourism Management*, 54, 178-184.
Thrane, C. and Farstad, E. (2011). "Domestic tourism expenditures: the non-linear effects of length of stay and travel party size". *Tourism Management*, 32, 46-52.
Thrane, C. and Farstad, E. (2012). “Tourists' length of stay: the case of international summer visitors to Norway”. *Tourism Economics*, 18(5), 1069-1082.
Van Nostrand, C., Sivaraman, V. and Pinjari, A.R. (2013). "Analysis of long-distance vacation travel demand in the United States: a multiple discrete-continuous choice framework". *Transportation*, 40, 151-171.
Vuong, Q.H. (1989). “Likelihood ratio tests for model selection and non-nested hypotheses”. *Econometrica*, 57(2), 307–333.
Wang, W. and Famoye, F. (1997). “Modeling household fertility decisions with generalized Poisson regression”. *Journal of Population Economics,* 10(3), 273-283.
Wang, E., Little, B.B. and DelHomme-Little, B. A. (2012). “Factors contributing to tourists’ length of stay in Dalian Northeastern China—a survival model analysis”. *Tourism Management Perspectives*, 4, 67-72.
Wilson, P. (2015). "The misuse of Vuong test for non-nested models to test for zero-inflation". *Economic Letters*, 127, 51-53.
Winkelmann, R. (2004). “Health reform and the number of doctor visits – an econometric analysis”. *Journal of Applied Econometrics*, 19(4), 455-472.
Winkelmann, R. (2008). *Econometric analysis of count data*. 5th ed. Berlin, Springer-Verlag.
Winkelmann, R. and Zimmermann, K.F. (1995). “Recent developments in count data modelling: theory and applications”. *Journal of Economic Surveys*, 9(1), 1-24.
Woodside, A.G. and Dubelaar, C. (2002). “A general theory of tourism consumption systems: A conceptual framework and an empirical exploration”. *Journal of Travel Research*, 41(2), 120-132.
Woodside, A. G. and King, R. (2001). “An updated model of travel and tourism purchase consumption systems”. *Journal of Travel and Tourism Marketing*, 10(1), 3–26.
Yen, S.T. and Adamowicz, W.L. (1994). “Participation, trip frequency, and site choice: a Multinomial-Poisson hurdle model of recreation demand”. *Canadian Journal of Agricultural Economics,* 42(1), 65-76.
Yeoh, E., Othman, K. and Ahmad, H. (2013). “Understanding medical tourists: word-of-mouth and viral marketing as potent marketing tools*”. Tourism Management*, 34, 196–201.
Zillinger, M. (2007). “Tourist routes: a time-geographical approach on German car-tourists in Sweden”. *Tourism Geographies*, 9(1), 64-83.

SUPPLEMENTARY MATERIAL

Determinants of Tourists' Length of Stay: a Hurdle Count Data Approach

ANNEX 1.- Survey design

- **Target Population:** visitors to the Principality de Asturias over 18.
- **Data gathering:** personal interviews to visitors to Asturias from January to December.
- **Languages:** Spanish; English; French; German.
- **Sample size:**

Year	Total	Collective establishments	On public road
2010	4,926	1,389	3,537
2011	5,076	1,574	3,502
2012	5,150	1,301	3,849
2013	4,268	701	3,567
2014	4,510	686	3,824
2015	4,831	1,525	3,306
2016	4,700	1,368	3,332

Table A1.- Sample size per year

- **Sampling procedure:**
  - **In collective establishments:** quota random sampling depending on the time period; type of accommodation; day of the week and geographical area.
  - **On public roads:** stratification by period; day of the week and geographical area. The interviewer randomly asked visitors at tourist places of interest.
- **Confidence interval:** 95%

Year		2010			2011			2012		
		N*	n	Sample error	N*	n	Sample error	N*	n	Sample error
Collective establishments		1,844,193	3,264	± 1.71 %	1,841,593	3,228	± 1.72 %	1,710,793	3,307	± 1.70 %
Geographical area	**West**	194,770	338	± 5.33 %	160,054	318	± 5.49 %	165,143	303	± 5.62 %
	Centre	1,158,590	1,969	± 2.21 %	1,112,301	1,805	± 2.30 %	1,015,768	1,744	± 2.34 %
	East	490,832	957	± 3.16 %	569,238	1,105	± 2.95%	529,882	1,260	± 2.76 %
Period	**Q1**	237,645	466	± 4.54%	217,463	445	± 4.64 %	212,714	388	± 4.97 %
	Q2	434,801	723	± 3.64 %	485,468	825	± 3.41 %	432,841	675	± 3.77 %
	Q3	849,009	1,467	± 2.56 %	849,445	1,272	± 2.75 %	788,671	1,453	± 2.57 %
	Q4	322,737	608	± 3.97 %	289,217	686	± 3.74 %	276,567	791	± 3.48 %
Purpose of the trip	**Holidays**	1,317,481	2,881	± 1.82 %	1,304,960	2,764	± 1.86 %	1,283,589	3,019	± 1.78 %
	Work	423,441	172	± 7.47 %	409,841	207	± 6.81 %	359,913	132	± 8.53 %
Private accommodation		2,325,245	552	± 4.17%	2,354,388	534	± 4.24 %	2,187,673	483	± 4.46 %
Same-day visitors		1,779,524	1,110	± 2.94 %	1,801,827	1,314	± 2.70 %	1,674,240	1,360	± 2.66%

Year		2013			2014			2015		
		N*	n	Sample error	N*	n	Sample error	N*	n	Sample error
Collective establishments		1,769,931	2,323	± 2.03 %	1,935,484	2,474	± 1.97 %	2,129,171	2,787	± 1.86 %
Geographical area	**West**	218,250	262	± 6.05 %	271,591	300	± 5.65 %	205,674	302	± 5.64 %
	Centre	1,023,827	1,215	± 2.81 %	1,103,793	1,330	± 2.69 %	1,327,295	1,642	± 2.42 %
	East	527,854	846	± 3.37 %	560,099	844	± 3.37 %	596,202	843	± 3.37 %
Period	**T1**	328,354	539	± 4.22 %	356,514	417	± 4.80 %	408,926	390	± 4.96 %
	T2	947,499	928	± 3.22 %	1,029,636	1,268	± 2.75 %	1,117,441	1,573	± 2.47 %
	T3	494,078	856	± 3.35 %	549,333	789	± 3.49 %	602,804	824	± 3.41 %
Purpose of the trip	**Holidays**	1,284,676	2,087	± 2.14 %	1,432,524	2,307	± 2.04 %	1,496,170	2,347	± 2.02 %
	Work	384,464	89	± 10.39 %	430,790	54	± 13.34 %	502,322	228	± 6.49 %
Private accommodation		2,222,882	525	± 4.28 %	2,438,468	429	± 4.73 %	2,655,869	620	± 3.94 %
Same-day visitors		1,701,185	1,420	± 2.60 %	1,866,174	1,607	± 2.44 %	2,032,553	1,424	± 2.6%

Year		2016		
		N*	n	Sample error
Collective stablishments		2,257,173	2,558	± 1.94 %
Geographical area	**West**	251,140	299	± 5.66 %
	Centre	1,321,130	1,523	± 2.51 %
	East	684,903	736	± 3.61 %
Period	**T1**	435,376	513	± 4.32 %
	T2	1,197,460	1,278	± 2.74 %
	T3	624,337	767	± 3.54 %
Purpose of the trip	**Holidays**	1,644,894	2,093	± 2.14 %
	Work	491,052	286	± 5.79 %
Private accommodation		2,794,675	731	± 3.62 %
Same-day visitors		2,138,782	1,411	± 2.61 %

Table A2.- Sample size per year, disaggregated by category

N* measures the total number of tourists in each category predicted by the National Statistics Institute (INE);
n refers to the number of sampled individuals in each category

ANNEX 2.- Questionnaire

1.– IS IT THE FIRST TIME YOU VISIT ASTURIAS?

1. No
2. Yes
3. Do you live in Asturias

Make the corresponding questions ☑

2.– HOW MANY TIMES HAD YOU BEEN IN ASTURIAS BEFORE?

1. Once
2. Twice
3. Three–five times
4. More than five times

3.– HOW OFTEN DO YOU VISIT ASTURIAS?

1. No 2. Yes ⇒ 4.– HOW MANY TIMES APPROXIMATELY? ____

5.– FOR REASONS OF _______________ ☐

☑ 6.– WHAT IS THE REASON FOR YOUR VISIT:

1. Holidays, spare time
2. Business
3. Conference, fairs _________________
4. Visit to family or friends
5. Studies
6. Health and medical treatments
7. Religious / pilgrimages
8. Sport competitions
9. Shopping (where) _________________

☑ 7.– WHO HAS ORGANIZED THIS VISIT TO ASTURIAS?

1. Yourself
2. Your company
3. Thorough a travel agency but only specific products (no tourist package)
4. A travel agency: tourist package ⇒ COST ________________ by person
5. An association, sports club ⇒ COST ________________ by person

☑ 8.– WHO DO YOU TRAVEL WITH?

1. Alone
2. In couple
3. With your family
4. With friends
5. With an organized group
6. With workmates
7. Others _________

9.– WHICH IS THE MAIN REASON WHY YOU CHOSE ASTURIAS AS YOUR DESTINATION?

1. Discovering new places
2. Recommendation (family, friends)
3. Experience/ previous visits
4. Country's natural environment / Landscape
5. Cultural heritage
6. No overcrowding
7. Gastronomy
8. Geographical proximity
9. Asturian origin
10. Mild weather
11. Hunting / fishing activities
12. Others _________________

☑ 10.– HAVE YOU SEEN ANY ADVERTISING ABOUT TOURISM IN ASTURIAS?

1. No 2. Yes

☑ 11.– WHAT KIND OF ADVERTISING? PLEASE, GRADE IT FROM 0 TO 10.

1. Mass media (point out)_________________________________ [......]
2. Tourist leaflets (handed in tourism offices) [......]
3. Fairs and presentations (point out)_________________________ [......]
4. Internet (general information) [......]

☑ 12.– HAS ADVERTISING INFLUENCED YOUR DECISION TO MAKE THIS TRIP? 1. No 2. Yes

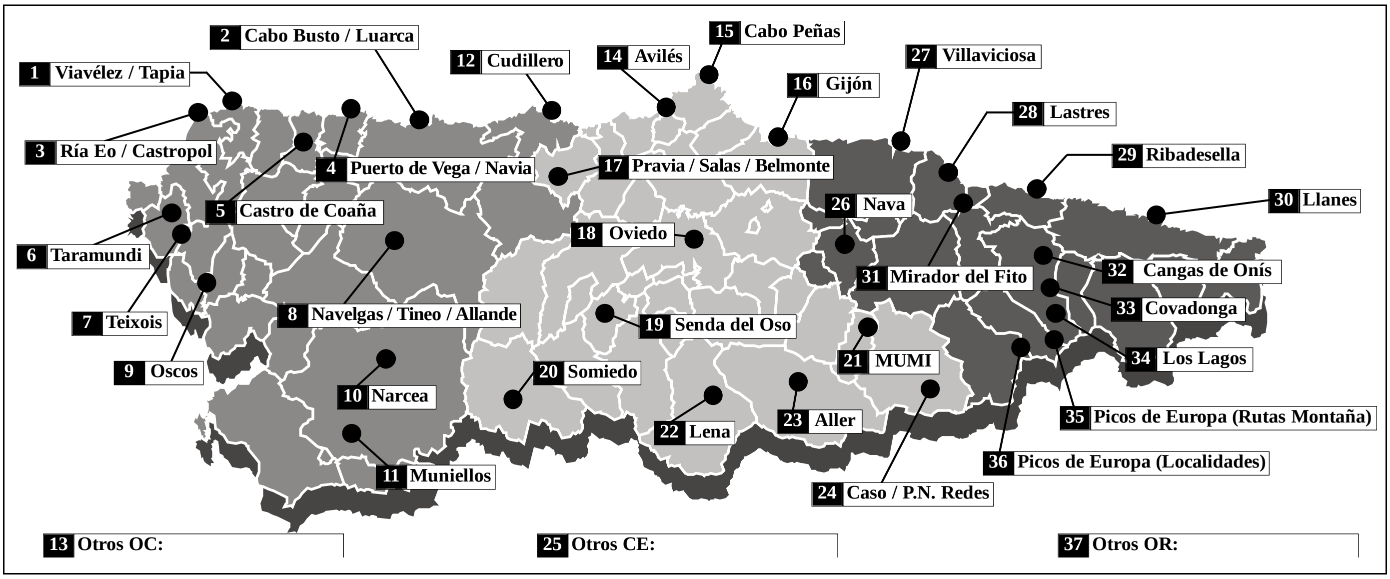
☑ 13.– HAVE YOU SEARCHED INFORMATION ABOUT ASTURIAS ON THE INTERNET?

1. No 2. Yes

☑ 14.– POINT OUT AND VALUE, FROM 0 TO 10.

1. Webs in general	[......]
2. Official Tourism Web of Asturias	[......]
3. Tourism Blogs	[......]
4. Social Networks	[......]
5. Others (point out) _______________________________________	[......]

☑ 15.– WHICH OF THE FOLLOWING PLACES OF ASTURIAS HAVE YOU VISITED, OR ARE YOU PLANNING TO VISIT?



16.– WHAT MEANS OF TRANSPORT HAVE YOU USED TO COME TO ASTURIAS?

1. Own car / motorcycle
2. Bicycle / Trekking
3. Hired car
4. Hired bus
5. Bus
6. Train
7. Airplane
8. Yacht / Ship / Cruise ships

17.– IN CASE YOU GOT TO ASTURIAS BY ROAD, CAN YOU TELL ME WHICH ROAD HAVE YOU GOT?

....

☑ 18.– HOW ARE YOU TRAVELLING AROUND ASTURIAS?

1. Own car /motorcycle
2. You do not travel
3. Friend's car
4. Bus
5. Hired car
6. Train
7. Bicycle / Trekking
8. Yacht / Ship
9. Tax

☑ 19.– ARE YOU GOING TO SPEND SOME NIGHTS IN ASTURIAS?

1. No 2. Yes

☑ 20.– HOW MANY? ______________________

☑ 21.– WILL YOUR STAY IN ASTURIAS INCLUDE A WEEKEND? 1. No 2. Yes

☑ 22.– IN THIS TRIP, ARE YOU GOING TO LODGE JUST IN ASTURIAS OR ALSO IN SOME OTHER REGION?

1. Just Asturias   2. Other regions   ⇒ ☑ 23.– WHICH REGIONS?

1. Cantabria   2. Galicia   3. Castilla y León   4. País Vasco   5. Others_________

☑ 24.– WHAT KIND OF ACCOMMODATION ARE YOU GOING TO STAY AT?

24.– TYPE OF LODGING	25.– WHAT WILL BE YOUR DAILY EXPENDITURE?	26.– HOW MANY PEOPLE?	27.– MEAL PLAN				28.– QUALITY / PRICE	
	€	People	AO1	BB2	HB3	FB4		

☑ 29.– IN CASE OF LODGING IN CAMP SITE: 1. Mobile home   ⇒ COST MONTH ___________€ / caravan

2. Bungalow   3. Tent   4. Motorhome

☑ 30.– THE APPLIED RATE HAS BEEN:

1. Official rate   3. Internet offer voucher   5. Travel agency rate
2. It tariffs discount voucher   4. Company rate   6. Others ________________________

FOR COLLECTIVE TOURIST ACCOMMODATION AND RENTED ACCOMMODATION

☑ 31.– WHERE DID YOU FIND THE INFORMATION ABOUT PLACES TO STAY?

1. Press or magazine
2. Travel agency
3. Real estate agency
4. Offices of tourist information
5. Internet
6. Tourist leaflets / Fairs
7. Specialized books and guides
8. Relatives or friends suggestions
9. Others _______________________

☑ 32.– HOW HAVE YOU BOOKED YOUR ACCOMMODATION? (Read all possible answers: only ONE valid answer)

1. By telephone directly with the accommodation
2. On arriving, (without reservation)
3. Through traditional travel agency
4. Through computerized systems
5. Through the Internet directly with the accommodation
6. Your company / organization has booked
7. Friends and family
8. Others _________________

☑ 33.– HOW LONG IN ADVANCE DID YOU BOOK YOUR ACCOMMODATION?____________________ (Point out the number of months or days)

☑ 34.– WHICH IS THE FINAL REASON TO CHOOSE THIS LODGING? WHAT ASPECTS DID YOU CONSIDER MOST IMPORTANT?

1. ________________ ☐ 2. ________________ ☐ 3. ________________ ☐

ONLY FOR SECOND RESIDENCE AND / OR RENTED ACCOMMODATION

☑ 35.– WHERE IS IT? ______________________

☑ 36.– IN WHAT NEIGHBOURHOOD / AREA? ______________________

☑ 37.– WHAT KIND OF ACCOMMODATION IS? 1. Chalet 2. Flat 3. Apartment 4. Rural building 5.Others__________

☑ 38.– ONLY COLLECTIVE TOURIST ACCOMMODATION. VALUE THE FOLLOWING ASPECTS OF THE ESTABLISHMENT IN WHICH IS YOU STAYED.

ASPECTS	Value from 0 to 10
1. Situation and environment	
2. Comfort of their facilities	
3. Building, architecture, design, atmosphere	
4. Service and attention of the personnel	
5. Added services (sports activities, leisure...)	
6. Services of restoration of the establishment	
7. Cleaning of the establishment	

☑ 39.– APPROXIMATELY, HOW MUCH ARE YOU GOING TO SPEND, DAILY PER PERSON (not considering accommodation)

HOSTELRY	DISTRIBUTION	EXPENDITURE
BREAKFASTS / LUNCHES / DINNERS	**In the accommodation**	€ / person / day
	In restaurants, sidrerias	€ / person / day
OTHER EXPENSES	**In bars / cafeterias / pubs / discos**	€ / person / day

☑ 40.– WHICH IS THE APPROXIMATE TOTAL EXPENDITURE YOU ARE GOING TO UNDERTAKE DURING YOUR TRIP? THIS SPEND IS ONLY FOR YOU OR FOR ALL THE TRAVELERS?

CONCEPT	DISTRIBUTION	EXPENSE	HOW MANY PEOPLE?
ACTIVITIES	**Cultural: tickets to museums, cinemas…**	€	People
	Sports activities (adventure travel)	€	People
TRANSPORT	**Petrol, bus… IN THE REGION**	€	People
PURCHASES	**Craftmanship, Souvenirs, fashion, trade…**	€	People
	Food and drinks in groceries, supermarkets…	€	People

☑ 41.– HAVE YOU VISITED ANY MUSEUM? 1. No 2. Yes

☑ 42.– WHICH ONE? 1.______________________________ Cod: ☐ 2.______________________________ Cod: ☐

☑ 43.– HAVE YOU VISITED ANY MONUMENT? 1. No 2. Yes

☑ 44.– WHICH ONE? 1.______________________________ Cod: ☐ 2.______________________________ Cod: ☐

☑ 45.– WHAT ACTIVITIES HAVE YOU MADE DURING YOUR STAY?

1. GO TO THE BEACH
2. SHOPPING
3. ACTIVITIES OF ACTIVE TOURISM
4. SHORT MOUNTAIN ROUTES/TREKKING
5. MOUNTAIN (HIKING)
6. VISIT TOWNS OR SIGHTSEEING SPOTS
7. GO OUT AT NIGHT / BARS / DISCO
8. REMAIN IN THE LODGING
9. GO TO THE MOUNTAIN
10. CYCLE TOURISM
11. THE WAY OF ST. JAMES
12. OTHER ACTIVITIES: ___________________________

☑ 46.– PLEASE, INDICATE IF YOU HAVE HIRED OR IF YOU INTEND TO HIRE SOME ACTIVITIES DURING YOUR TRIP ROUND ASTURIAS. IN THAT CASE, VALUE .

					Value from 0 to 10
1.	Marine activities (surfing, sailing, diving…)	1. No	2. Yes	3. If it would hire	
2.	Canoeing	1. No	2. Yes	3. If it would hire	
3.	Bungee jumping, rafting	1. No	2. Yes	3. If it would hire	
4.	Guided routes: trekking, hiking	1. No	2. Yes	3. If it would hire	
5.	Horse riding	1. No	2. Yes	3. If it would hire	
6.	Trips in all road, quad, motorbike	1. No	2. Yes	3. If it would hire	
7.	Golf	1. No	2. Yes	3. If it would hire	
8.	Rent of Bicycles (Mountain Bike)	1. No	2. Yes	3. If it would hire	
9.	Ski / Snow sports	1. No	2. Yes	3. If it would hire	
10.	Other ________________________________	1. No	2. Yes	3. If it would hire	

☑ 47.– VALUE FROM 0 TO 10 THE FOLLOWING ASPECTS ABOUT ASTURIAS.

ASPECTS	VALUE FROM 0 TO 10
BARS, COFFEE SHOPS	
RESTAURANTS / SIDRERIAS (CIDER HOUSES)	
OFFICES OF TOURIST INFORMATION	
HIGHWAYS / SIGNALLING	
TREATMENT RECEIVED / HOSPITALITY	
CONSERVATION OF THE ENVIRONMENT AND THE CULTURAL HERITAGE	
GASTRONOMY	
PRICES	

☑ 48.– IN YOUR OPINION WHICH OF THE PLACES VISITED IN ASTURIAS DURING THE TRIP ARE THE MOST INTERESTING?

1. ____________ 2. ____________ 3. ____________

49.– TELL US SOMETHING THAT YOU HAVE REALLY MISSED DURING YOUR STAY OR SOMETHING THAT MUST BE IMPROVED.

____

50.– WHAT DO YOU LIKE THE MOST OF ASTURIAS? WHAT HAS CAUGHT YOUR ATTENTION?

____

☑ 51.– WHERE DO YOU LIVE?

City: | Cod: | Province: | Cod: | CC.AA.: | Cod:

☑ 52.– IN WHICH COUNTRY?

Country: | Cod: | City: | How did you travel to Spain? (Transport and place of arrival)

☑ 53.– WOULD YOU LIKE TO INDICATE YOUR JOB? ____________________ Cod:

☑ 54. – ED.LEVEL: _________________ Cod:

☑ 55.– COULD YOU TELL US YOUR AGE? ______________ YEARS OLD

☑ 56.– SEX: 1. Masculine 2. Feminine

☑ 57.– FOR POSSIBLE FOLLOW UPS, COULD YOU TELL US YOUR NAME AND YOUR TELEPHONE NUMBER TO CONTACT YOU?

Contact person: Telephone:

THANK YOU VERY MUCH

ANNEX 3.- Descriptive statistics

Type of variable	Variables	Mean	SD	Min	Max	Definition
TYPE OF TRAVELER	*tourist*	0.914	.279	0	1	The individual sleeps in the destination at least 1 night
TYPE OF TRAVELER	*same-day*	0.085	.279	0	1	The individual does not spend the night at the destination
DEPENDENT VARIABLE	*LOS*	4.313	3.830	0	30	Number of nights spent at the destination
SOC	*male*	0.543	0.498	0	1	Man
SOC	*age*	40.315	12.110	18	91	Age
SOC	*housewife*	0.036	0.186	0	1	Housewife/ househusband
SOC	*retired*	0.056	0.231	0	1	Retired
SOC	*employed*	0.652	0.476	0	1	Employed
SOC	*student*	0.076	0.265	0	1	Student
SOC	*unempl*	0.018	0.136	0	1	Unemployed
SOC	*selfempl*	0.155	0.362	0	1	Self-employed
SOC	*primary*	0.072	0.259	0	1	Primary studies
SOC	*secondary*	0.309	0.462	0	1	Secondary studies
SOC	*higheduc*	0.617	0.485	0	1	Higher education
SOC	*foreign*	0.081	0.274	0	1	The individual lives in another country
ATTRIB	*natural*	0.113	0.317	0	1	The individual visits Asturias due to its natural environment
ATTRIB	*novelty*	0.353	0.477	0	1	The individual visits Asturias due to novelty seeking
ATTRIB	*tranquility*	0.007	0.087	0	1	The individual visits Asturias looking for tranquility
ATTRIB	*climate*	0.007	0.088	0	1	The individual visits Asturias due to its climate
ATTRIB	*heritage*	0.007	0.083	0	1	The individual visits Asturias due to its heritage
ATTRIB	*gastronomy*	0.010	0.103	0	1	The individual visits Asturias due to its gastronomy
PRICE	*accom_price*	28.602	22.536	0	575	Daily expenditure per person (€) on accommodation
KNOWLEDGE	*recommend*	0.092	.289	0	1	The individual visits Asturias due to recommendation
KNOWLEDGE	*experience*	0.215	0.411	0	1	The individual visits Asturias due to positive previous experience
KNOWLEDGE	*advert*	0.347	0.476	0	1	The individual has seen advertising.
KNOWLEDGE	*first*	0.387	0.487	0	1	First time the individual visits Asturias
KNOWLEDGE	*num_vis*	0.928	6.295	0	100	Number of visits during the year.
TRIP	*distance*	675.12	1,171.70	0	17,713	Distance between origin and Oviedo (in km)
TRIP	*car*	0.824	0.380	0	1	Car
TRIP	*bus*	0.026	0.161	0	1	Bus
TRIP	*train*	0.031	0.173	0	1	Train
TRIP	*plane*	0.075	0.264	0	1	Plane
TRIP	*alone*	0.063	0.244	0	1	The individual travels alone
TRIP	*couple*	0.517	0.499	0	1	The individual travels with a partner (as a couple)
TRIP	*group*	0.418	0.493	0	1	The individual travels with his/her family, friends or workmates (in a group)
TRIP	*party_size*	3.704	7.110	1	250	Party size (number of members in the trip)
TRIP	*leisure*	0.830	0.375	0	1	The individual comes for leisure or on holidays.
TRIP	*labor*	0.063	0.242	0	1	The individual comes because of studies of job-related reasons.
TRIP	*family*	0.075	0.263	0	1	The individual comes for visiting relatives.

TRIP	*other*	.0315	.1748	0	1	The individual comes for a doctor visit, making purchases, a religious peregrination or a sport competition.
TRIP	*hotel*	0.564	0.495	0	1	The individual stays at a hotel
TRIP	*hostel*	0.038	0.192	0	1	The individual stays at a hostel
TRIP	*rural*	0.112	0.316	0	1	The individual stays at a rural house
TRIP	*campsite*	0.054	0.226	0	1	The individual stays at a campsite
TRIP	*private*	0.144	0.351	0	1	The individual stays at a private accommodation
TRIP	*himself*	0.907	0.289	0	1	The individual organized the trip himself
TRIP	*travel_agency*	0.044	0.206	0	1	The trip was organized by a travel agency
TRIP	*club_firm*	0.047	0.213	0	1	The trip was organized by a club or the company the individual works for.
TRIP	*act_tour*	0.074	0.262	0	1	The individual performs active tourism activities
TRIP	*only_ast*	0.832	0.372	0	1	The individual only visits Asturias.
AREA	*west*	0.161	0.367	0	1	West area
AREA	*capcity*	0.458	0.498	0	1	Capital city area (Oviedo-Gijon-Avilés)
AREA	*central*	0.046	0.210	0	1	The rest of the central area
AREA	*east_coast*	0.172	0.377	0	1	East coast
AREA	*east_inner*	0.161	0.367	0	1	East inner area
TEMP	*t1*	0.194	0.395	0	1	January-February-March-April
TEMP	*t2*	0.492	0.499	0	1	May-June-July-August
TEMP	*t3*	0.304	0.460	0	1	September-October-November-December
TEMP	*y10*	0.135	0.341	0	1	Year 2010
TEMP	*y11*	0.129	0.335	0	1	Year 2011
TEMP	*y12*	0.137	0.344	0	1	Year 2012
TEMP	*y13*	0.113	0.317	0	1	Year 2013
TEMP	*y14*	0.148	0.355	0	1	Year 2014
TEMP	*y15*	0.179	0.384	0	1	Year 2015
TEMP	*y16*	0.155	0.342	0	1	Year 2016
OBSERVATIONS		19,111				

Table A3.- Descriptive statistics, variable definition, and acronym

Chapter 2.- Modelling Heterogeneous Preferences for Nature-based Recreational Trips

David Boto-García

botodavid@uniovi.es

Department of Economics. University of Oviedo

Avenida del Cristo s/n, Oviedo (Asturias). Faculty of Business and Economics, Oviedo (Spain).

Abstract:

This paper studies individual preferences for nature-based recreational trips. We estimate a Random Parameter Multinomial Logit with Error Components that controls for i) unobserved preference heterogeneity for the attributes, and ii) correlation in unobserved destination features. We examine the influence of a set of mean shifters in the marginal utilities for destination attributes, and how individuals are willing to trade distance for warmer (cooler) locations in the form of Marginal Rates of Substitution. We make use of a rich dataset of trips for nature and sport purposes from a sample of Spanish residents. We find evidence of large heterogeneity in preferences for temperature differentials and distance, with trip purposes acting as moderators. Our results also show that nature-based tourists appreciate regions with national parks, a high number of kilometres for skiing and tourism sightseeing spots. Conversely, high rainfalls and prices deter recreational site choice probabilities. Own and cross elasticities for prices and relative temperatures are also derived and interpreted.

Keywords: *nature-based tourism; recreational demand; Random Parameter Logit; temperature; distance; marginal rates of substitution*

JEL codes: C35, Q26, R21, Z30

Acknowledgements:

I am grateful to my PhD advisors José Baños and Antonio Álvarez for helpful comments and suggestions.

This paper also benefited from suggestions by Juan Luís Eugenio, Jaume Rosselló, Juan Prieto and Carlos Arias. I also want to thank seminar participants at the Universities of Oviedo and Las Palmas de Gran Canarias for their valuable comments.

Financial support from the Spanish Ministry of Education, Culture and Sport (FPU 16/00031) is gratefully acknowledged.

1. INTRODUCTION

Tourism can be understood as a form of trade in services that involves the temporal displacement of consumers across regions, whose comparative advantage is determined by natural endowments (Biagi and Pulina, 2009). It is nowadays a fast-growing industry and a crucial driver of economic growth for certain areas (Paci and Marrocu, 2014). In this sense, there is robust evidence that earnings from tourism activities drive a substantial share of domestic real income in the long run (Balaguer and Cantavella-Jordá, 2002; Lee and Chang, 2008; Faber and Gaubert, 2019), both in developed and developing countries (Sequeira and Nunes, 2008).

The tourism sector is particularly relevant in Spain, a country for which it has become an important economic force. According to UNWTO (2018), Spain is the second most visited country in the world behind France. In 2017, Spain received more than 81 million international arrivals. In 2018 this figure increased up to 82.6 million. The tourism industry accounts for 11.7% of GDP and constitutes 12.8% of total employment according to the last available estimates by the Tourism Satellite Account (INE, 2019).

Despite the relevance of international tourism, the Spanish domestic travel market has increased its importance in the last decade. In 2018, it represented over two thirds of the aggregate demand[34]. More than 100 million domestic trips were made for leisure purposes, which represents an aggregate expenditure of more than 28 billion euros. However, there are relatively few studies aimed at examining the factors that pull people to travel to one region or another. In this sense, the economic modelling of domestic trips has been overlooked in favour of analysing the international market. We seek to fill this gap in the literature.

Within the different trip purposes, nature-based tourism is gaining increasing popularity and attention in the literature (Gosens and Rouwendal, 2018; Kim et al., 2019)[35]. The share of tourists that declare nature and/or sport as their main trip purpose has significantly increased during the last decade. It is estimated to represent 15% of total tourism in the world (UNWTO, 2018). In Spain, the number of tourists who travel for nature-based reasons has increased by 32% during the 2009-2016 period, reaching 3.6 million tourists in 2016 (INE, 2017). About 81% of the total nature-based travellers are residents. Moreover, total spending from nature-based tourists is estimated to be about 9,000 million euros, which constitutes 11% of overall spending (SGAPC, 2017).

In this paper, we examine how regional attributes affect individual destination choices for nature-based tourism within Spain. We specifically focus our attention on the role of distance and temperatures relative to the place of residence. Previous studies have shown that it is not only climate conditions at the destination which matter for tourists'

[34] This figure is obtained as the ration of the number of people that travelled domestically in 2018 (any purpose, 197 million) and the total number of tourists that visited Spain in 2018 (sum of domestic and international trips, 279 million).

[35] Nature-based tourism refers to broad array of outdoor recreational activities such as picnic, trekking, visiting natural areas, guided routes, or the observation of the fauna in their natural environment. It is sometimes also referred to as rural tourism, ecotourism, or wilderness/adventure tourism. This type of tourism also includes sport activities performed in the natural environment, such as scuba diving, canoeing, skiing or climbing.

choices but climate differences between the place of origin and potential destinations (Bigano et al., 2006; Rosselló-Nadal et al., 2011). Intuitively, individuals who are used to colder (warmer) climates may be looking to travel to warmer (colder) regions. Although there is a large body of literature on this issue for coastal destinations (e.g. Bujosa and Rosselló, 2013), the preferences for warmer or cooler destinations have been less studied in the context of nature-based tourism. We consider the ratio between temperature at each possible destination and that at the origin to study 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 recreational choice, since tourists exhibit heterogeneous preferences regarding travelling to nearby or distant destinations (e.g. Nicolau, 2008; 2010a). We explore whether the type of activities to engage in at the destination moderate or intensify the disutility of distance. This relates to previous evidence that links leisure trips to the satisfaction of needs (e.g. Arentze and Timmermans, 2009; Dekker et al., 2014).

We use microdata for monthly domestic trips by Spanish residents between February 2015 and September 2017. We focus on leisure trips for nature-based and sport purposes to the 17 Spanish Autonomous Communities. We combine this dataset with: i) monthly regional data on tourism prices, temperatures, rainfalls and ski track kilometres available for practising winter sports; ii) weighted bilateral distances between the origin and potential destinations that explicitly take into account the sampling probability of residence location within the region, and iii) time-invariant regional-specific characteristics such as the number of tourism spots, the number of national and natural parks, the size of protected natural areas and the presence of coast.

We provide a microeconometric foundation of tourists' destination choice by combining Lancaster's product-characteristics approach (Lancaster, 1966) with Berry & Pakes' discrete choice modelling, with both observed and unobserved preference heterogeneity for the attributes (Berry and Pakes, 2007). In this way, this paper offers some additional guidance to the empirical modelling of tourist destination choice.

Our econometric model is a Random Parameter Multinomial Logit with Error Components that controls for unobserved heterogeneity in preferences for the attributes and for the alternatives (Greene and Hensher, 2007)[36]. We allow the random parameters to be a function of a mean parameter and a set several individual-specific characteristics such as age, income, party size and trip purpose, among others, plus a random term varying across individuals. In this way, while introducing stochastic variability in the marginal utilities for the attributes, we explore the factors that shift these marginal utilities. 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 between regions arising from common unobserved characteristics.

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 these two dimensions. Based on the model estimates, we derive

[36] This model is also known as the Mixed Logit Model with Error Components.

the conditional means of the individual-specific marginal utilities for these two attributes and compute the marginal rates of substitution at the individual level. In doing so, we address the problem of potential attribute non-attendance following the approach proposed by Hess and Hensher (2010). Therefore, we assess how much distance individuals are willing to travel for warmer (cooler) temperatures, *ceteris paribus*.

We contribute to the literature by examining how individual characteristics and travel motivations are related to marginal utilities. 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. From this perspective, our work is similar to that of Swait et al. (2020), who examined how goal pursuit mitigates the distance decoy effect. Our study differs from theirs in that we use revealed preference data and cover different periods within the year. The latter allows us to study potential seasonal effects in preferences.

Our results show that, on average, tourists attach a negative utility to distance and temperature differentials. However, marginal utilities are quite heterogeneous. Trekking and mountaineering as the main trip purpose, travelling in weekends and party size increase the dissuasive effect of distance, whereas income and the seeking of adventure and aquatic sports moderate it. In addition, recreationists seem to prefer cooler destinations in the summer season, especially those coming from regions with above-mean temperatures. Interestingly, tourists are willing to cover 159 kilometres to gain a marginal increase in temperature relative to that at the origin. Moreover, tourists attach positive utility to the availability of kilometres for skiing, national and natural parks, the size of protected natural areas and tourism spots.

The paper is structured as follows. In Section 2 we review the related literature. Then, Section 3 presents the theoretical framework. Section 4 describes the database and the variables employed. The following Section outlines the econometric modelling and the empirical strategy. The estimation results are discussed in Section 6. Finally, Section 7 concludes with the main remarks and some policy implications.

2. LITERATURE REVIEW

The vacation decision-making process has received substantial attention in the tourism literature. Two formal characterizations can be found in Seddigui and Theocharous (2002) and Hyde and Laesser (2009). Scholars agree that destination choice is a complex process that involves various stages. Eugenio-Martín (2003) conceptualizes tourism demand as a five-stage process in which early-stage decisions condition subsequent ones. A similar hierarchy is also postulated by Nicolau and Mas (2005; 2008). The two initial decisions tourists must make are whether to go on holidays and (if so) where to go. Although we focus on the choice of destination conditional on participation, personal motivations along with constraints are key determinants of the participation decision[37].

There is a large body of literature that investigates tourists' destination choice, both from theoretical (e.g. Papatheodorou, 2001) and from empirical perspectives (e.g. Grigolon et al., 2013)[38]. In general terms, the determinants of vacation choice can be categorized into two main types: personal motivations (also referred as 'push factors') and destination attributes ('pull factors'). That is, people travel because they are pushed by internal motivations and pulled by the attractiveness of potential destinations. Nonetheless, personal and budget constraints must also be considered. We now proceed to discuss each of them.

2.1. Personal motivations and situational inhibitors

Many studies have examined the vacation decision making process from a psychological perspective[39]. An extensive review of this can be found in Qiu et al. (2018). Researchers have conceptualized tourist behaviour and destination choice as a process in which tourists initially construct a choice set with the different possible destinations according to their motivations. However, since individuals cannot freely choose where to go because they are restricted by several factors (income, time availability, etc.), the final choice is made from the so-called feasible choice set (Crompton and Ankomah, 1993; Lam and Hsu, 2006; Decrop, 2010). Consistent with this, we first introduce trip motivations and we then turn to travel constraints.

2.1.1. Personal motivations

Travel motivations play a key role in tourist destination choice since the choice of a certain holiday destination implies a desire for some benefit[40]. In this regard, travel motivations are strongly related to tourists' psychographic characteristics (e.g. Plog, 2002). The so-called 'push motivations' refer to individuals' wish to escape from routine, rest and relaxation, adventure, social interaction, excitement or family togetherness (Crompton, 1979; Pearce and Lee, 2005). A formal characterization can be found in Arentze and Timmermans (2009), who develop a theoretical framework about how households engage in different leisure activities to satisfy certain needs. At the empirical

[37] Eugenio-Martín and Campos-Soria (2010) show that the probability of going on holidays domestically for European households is positively associated with income, education, living in coastal areas, the presence of children in the household and the size of the place of residence. Alegre et al. (2010) conduct a microeconometric analysis of tourism participation decision among Spanish residents. They find that the probability of being able to afford a holiday decreases when households are older than 55 and when there are minors living in the household. By contrast, it increases with income (especially median income households), education level and a good health status.

[38] A review of the tourism decision-making literature can be found in Smallman and Moore (2010).

[39] Some examples are Van Raaij and Francken's vacation sequence (Van Raaij and Francken, 1984), the 'structure of vacation destination choice sets' (Crompton 1992; Um and Crompton, 1992; Ankomah et al., 1996), the 'vacation tourist behaviour model' (Moutinho, 1987) and the 'destination choice model' (Mansfeld, 1992).

[40] Travel motivations have been defined as 'psychological/biological needs and wants, including integral forces that arouse, direct and integrate a person's behaviour and activity' (Yoon and Uysal, 2005, p.46). According to McCabe (2000), their characterization requires the combination of both behaviourist and cognitivist approaches.

level, the measurement of motivations and needs imposes a challenge when using revealed preference data since they are latent constructs that are normally unobserved.

However, some studies using stated preferences have provided some insights into the role of the satisfaction of needs on leisure trip choices. By means of a discrete choice experiment, Dekker et al. (2014) find that anticipated needs-satisfaction for walking in nature is highly correlated with the need for physical exercise, relaxation and being outdoors. Strikingly, needs-satisfaction for walking in nature is lower among males and decreases with age. Similarly, Swait et al. (2020) examine how goal pursuit determines site choice probabilities. Their results show that those destinations that are perceived more suitable for relaxation, spending time with family or contact with nature are significantly more likely to be chosen.

The ability of each destination to meet tourists' goals depends on how the destination is perceived (i.e. destination image). This abstract perception is collectively constructed and is subject to social biases (Chen et al., 2013). For example, press media coverage (Castelltort and Mäder, 2010) and framing of marketing campaigns (Zhang et al., 2018) have been shown to affect destination image formation. Nonetheless, once stablished, destination image tends to remain stable over time, even under the shock of bad events (Gkritzali et al., 2018). Furthermore, destination image perceptions depend on individuals' psychographic variables such as self-congruity (Beerli et al., 2007) or animosity (Stepchenkova et al., 2019).

Similarly, the literature has paid attention to the effect of past visitation on future choices. On the one hand, tourists who seek novelty and discovering new places will have a lower probability of returning to an already visited destination (e.g. McKercher and Guillet, 2011). On the other hand, tourists might prefer to travel to the same destination on a routine basis. This is because as tourists become more familiar with the destination, they might feel more confident about finding the services and products they want. In this vein, García et al. (2015) show that the probability of returning to a destination increases with the number of previous visits. Furthermore, a high degree of satisfaction is another argued reason for coming back.

Despite all this, some evidence points to the existence of great disparity between travel intentions and actual behaviour (Kah et al., 2016). There is some consensus that tourists plan their leisure trips in advance. Consistent with construal level theory (Trope and Liberman, 2010), longer temporal distances are associated with more abstract and ideal destination choices. However, as the trip period approaches, tourists' preferences change from desirability to feasibility. By means of four choice experiments, Li et al. (2019) find that tourists have a strong preference for feasible destinations as the time to decision approaches.

2.1.2. Travel constraints

As introduced before, when choosing a vacation destination, tourists are subject to some constraints. The most important one is income. Since tourism is a normal good with positive income elasticity of demand (e.g. Davis and Mangan, 1992; Alegre et al., 2010), people with high incomes have a higher attainable choice set from which to choose. Put another way, apart from its effect on participation, income further restricts the number of destinations among to go[41].

Time constraints also affect recreational choices. Most people travel during holidays and the amount of time available for the journey might affect the decision about where to go (Dellaert et al., 1998). In this sense, McConnell (1999) develops a time-allocation model for labour supply and recreational demand and shows that recreational choices depend on the household labour market situation.

The size and the composition of the travel party is another relevant factor in destination choice (e.g. Basala and Klenosky, 2001). Apart from its effect through the budget constraint, empirical evidence shows that the attractiveness of each destination varies depending on who you travel with. For instance, Kaoru (1995) provides evidence on how site choice changes depending on trip companions. Campo-Martínez et al. (2010) document that loyalty to Mallorca differs depending on the travel party composition. Furthermore, in the context of mountain-biking site choice, Morey and Krizberg (2012) find that whether one has a companion and her relative ability are as important as the costs or the physical characteristics of the places for deciding where to go.

2.2. Destination attributes

Tourism products cannot be relocated. Hence, the intrinsic physical environment and characteristics of the different destinations are what make tourists to travel to one place or another. The so-called 'pull factors' refer to destinations' attractiveness in terms of recreation facilities, cultural attractions, natural scenery, beaches, etc. In this subsection, we review the literature on the effect of distance, climate and prices on tourists' destination choice. In addition, we also discuss some other destination-specific attributes that have been shown to affect recreational choice. We refer here to studies that model either individual or aggregate demand, since the modelling of flows and arrivals is the aggregation of individual preferences.

Distance

Given the spatial dimension of vacation site choice, the distance between the region where the tourist lives and the possible destinations is one of the most important factors for explaining destination choice. Conditional on the origin, distance to each alternative destination can be thus understood as a destination attribute. Although it has received great attention in the literature, evidence about tourists' preference for nearby or farther destinations is inconclusive.

On the one hand, travelling to faraway destinations entails higher monetary and time costs (Taylor and Knudson, 1976; Chandra et al., 2014). Distant locations are associated with higher travel times, which impose costs both in terms of the opportunity cost of time and reducing the time spent at the destination. To alleviate this, tourists normally opt for fast modes of transport such as the plane or the high-speed rail (Thrane, 2015).

[41] Nonetheless, Bernini et al. (2017) show that whereas international tourism can be understood as a luxury good, domestic tourism is less sensitive to income (basic need).

Therefore, given time and budget constraints, people will choose nearby destinations, *ceteris paribus*, with higher probability. Under this viewpoint, distance is a dissuasive factor. In this sense, distance decay has been revealed as a tourism geography law by which tourism demand decreases as distance increases (Lise and Tol, 2002; McKercher and Lew, 2003; McKercher et al., 2008; Kah et al., 2016; McKercher, 2018; Gosens and Rouwendal, 2018; Mckercher and Mak, 2019)[42].

On the other hand, tourists who are looking to discover new places and environments might be willing to travel far away if the additional utility that distant destinations provide them is greater than the costs they entail. Nicolau and Más (2006) were among the first to provide empirical evidence that motivations like the search for a better climate or discovering new places make people willing to cover longer distances to satisfy those needs. In such situations, distance can be perceived as an appealing attribute.

These two opposite effects constitute the so-called antinomy of distance (see Cao et al., 2020). A growing body of research has started to study the factors behind it.

Nicolau (2008) examines how tourists' sensitivity to distance can be explained by sociodemographic characteristics. He shows that the disutility of distance decreases with income, organizing the trip through intermediaries, the size of the city of residence and travelling in faster modes of transport. Conversely, the higher the number of children under 16 in the household, the higher the preference for nearby destinations. Van Nostrand et al. (2013) analyse long-distance vacation travel patterns in USA. They find that faraway destinations are less attractive and that those who travel to distant destinations tend to stay there for longer to make the effort worthwhile. Interestingly, low-income households spend more time travelling than their high-income counterparts, possibly because they opt for slower modes of transport. Wong et al. (2016) show that long-haul tourists are mainly older, better educated and high-income couples travelling without children.

In a similar fashion, Nicolau (2010a) explores how the effect of distance on destination choice depends on tourist's motivations (whether tourist looks for variety or display inertial behaviour)[43]. His results show that *variety-seeking behaviour*[44] decreases the disutility of distance whereas *inertial behaviour* increases it. Put it simply, tourists who seek variety from one trip to another are more willing to cover longer distances to enjoy new experiences. Conversely, for those who prefer to travel to the same type of destination, distance acts as an important deterrent factor. By means of a discrete choice experiment, De Valck et al. (2017) also arrive to similar conclusions. These authors show that, in the context of recreational demand, dog walkers and joggers are highly sensitive to distance, while cyclists are less affected. More recently, Swait et al. (2020) show that recreationists are willing to travel further distances if in exchange they can achieve their goals. Therefore, goal fulfilment makes tourists to become less distance-sensitive, being

[42] Notwithstanding this, the development of transport infrastructures has reduced travel costs over time, thereby alleviating the dissuasive effect of distance and increasing the share of tourism flows between distant destinations (Albalate and Fageda, 2016; Pagliara et al., 2017).
[43] Inertial behaviour is conceptualized as routinized consumption by which the tourist travels to the same destination.
[44] This refers to a psychological need for continuously discovering new things and places.

the mitigation effect higher as goal importance increases. Overall, these studies point to the existence of a high degree of heterogeneity with reference to distance that depends on trip purposes and motivations.

Some scholars have put forward the idea that the effect of distance on destination choice is not constant over time. Yang et al. (2018) explore this issue by studying whether preferences for distant or nearby destinations change over time due to improvements in transportation systems and the corresponding savings in terms of costs and travel time. They find that people are willing to choose distant countries as *dream* destinations. However, past choices and intended destinations are negatively associated with distance. Interestingly, this negative relationship has decreased over time. The authors also report the existence of substantial heterogeneity in the sensitivity to distance. Additionally, Wong et al. (2017) indicate that the dissuasive effect of distance is relaxed during the growing phase of the business cycle (i.e. higher income increases the willingness to travel farther away). Similar findings are reported by Eugenio-Martin and Campos-Soria (2014) and Cafiso et al. (2016), who show that during the Great Depression tourists have turned to closer destinations to reduce their expenditure.

Climate

Climate is another important attribute when choosing where to travel (Eymann and Ronning, 1997; Moreno Sánchez, 2010)[45], particularly for trip purposes that involve outdoor activities. Kozak (2002) finds that 'enjoying good weather' is the most important factor among German and British tourists travelling to Mallorca and Turkey.

The effect of climate on tourism demand has been modelled both from micro and macro perspectives. Rosselló-Nadal (2014) provides a discussion about the different empirical strategies used in the literature. A stylized finding is the existence of an inverted U-shaped relationship between temperature and tourism demand (i.e. tourists prefer to travel to warmer destinations but, after reaching a comfort threshold, rises in temperature deter demand).

Using data on aggregate tourists' arrivals, Maddison (2001) models the effect of climate on the number of return visits from the United Kingdom considering the average maximum daytime temperature and rainfall at the destination. His results point to the existence of an optimal maximum daytime temperature around 30ºC. Conversely, and in line with expectations, rainfalls are negatively associated with visitation rates. Similarly, Lise and Tol (2002) examine destination choices made by Dutch tourists as a function of rainfalls and temperatures at the possible destinations. They show that optimal temperature is independent of the country of origin and is about 21ºC. Falk and Lin (2018) examine tourism demand to South Tyrolean mountains in winter, focusing on the role of temperatures. Their results indicate that a one percent increase in temperature decreases arrivals by 8 percent, whereas a decrease in temperature does not exert any effect. Becken (2012) studies the impact of weather on both intra- and inter-annual

[45] It is important to highlight the distinction between climate and weather. The term *climate* refers to the prevailing condition of the atmosphere drawn from long periods of observation. Contrariwise, the term *weather* refers to the state of the atmosphere at a particular time (Gómez-Martín, 2005). By the time the trip if booked, tourists cannot anticipate the specific weather at each possible destination, so they rely on expected conditions (climate).

seasonality in visitors to New Zealand. She considers minimum and maximum temperatures, rain and sunshine hours for explaining monthly overnight stays. Both minimum and maximum temperatures along with sunshine hours are significant predictors of seasonality, whereas rain is less important. According to her findings, temperature is the most important climate indicator. Also for a sample of international visitors to New Zealand, Becken and Wilson (2013) examine how weather conditions induce travellers to change their trip plans once at the destination. They report that those who were forced to change their plans (in terms of length of stay or leaving a place earlier than desired) due to unexpected weather conditions were less satisfied with the trip.

For the case of nature-based tourism, a stream of literature has analysed the relationship between climate conditions and demand, with mixed findings. Fisichelli et al. (2015) study the linkages between climate and park visitation in USA using historical data. They find that visitation rates increased with monthly average temperatures but strongly decreased when temperatures were above 25ºC. Hewer et al. (2017) analyse summer campers climatic preferences in Canada according to the type of activities performed. Their results indicate that older campers devote greater importance to weather whereas active ones are more tolerant to adverse weather conditions. Additionally, campers coming from nearby locations are more likely to leave the park early if adverse weather takes places. In a similar study, Hewer et al. (2016) identify temperature threshold points that discourage park visitation. Temperatures over 33ºC (under 10ºC) appear to be 'too hot' ('too cold') to visit natural parks. Smith et al. (2018) report that daily maximum temperatures that exceed 25ºC reduce monthly park visitation rates in the 3 of the 5 most important parks in southern USA, but for the other two the number of visitors continues rising over that threshold. Pongkijvorasin and Chotiyaputta (2013) report that increases in temperature and rainfall adversely impact the number of visitors to a National Park in Thailand. In line with this, Liu (2016) finds that rainfall is the factor that mostly deters tourists from visiting national parks in Taiwan, although increases in temperatures also negatively impact the number of visitors. For the case of cycling tourism, Helbich et al. (2014) show that temperature has a positive effect while wind speed and rainfall act as inhibitors.

Although traditionally climate has been considered a pull factor, some scholars have shown that climate at the place of origin also plays a role. This is because tourists might choose a destination as to gain a climatic advantage in comparison to where they live. Under this reasoning, unfavourable climate conditions at the place of origin are a 'push' factor. Ridderstaat et al. (2014) study tourism flows from the United States and Venezuela to Aruba focusing on the effect of rainfall, temperature, wind and cloud coverage. They show that climate at the origin and the destination matter for explaining the total number of visitors to Aruba. Li et al. (2017) analyse tourism flows from Hong Kong to China considering climate conditions at the origin, at the destination and the absolute difference between them. Interestingly, these authors find that home climate is more relevant for explaining flows than climate at the destination or the difference between the two.

Similarly, at the micro level some studies have documented how climate conditions at the place of origin affect destination choice. For example, Eugenio-Martín and Campos-Soria (2011) and Roselló-Nadal et al. (2011) show that higher temperatures at the origin exert a negative effect on outbound tourism and thus increase the likelihood of travelling domestically. This result is further confirmed by Eugenio-Martín and Campos-Soria (2010), who study the role of the climate in the region of origin on European households' probability of going on holidays abroad or domestically. Their estimates clearly show that those who live in places with better climate conditions travel domestically, whereas those from colder regions have a higher probability of travelling abroad. Chandra et al. (2014) also find that cross-border travel rises with average temperature at the origin.

Another stream of research has studied the effects of climate change (hereafter CC) on tourism flows. A general agreement between academics is that global warming may cause the weather conditions of lower altitude and latitude locations to become less attractive for tourists (e.g. Amelung et al., 2007). In this sense, skiing and sun and beach tourism are the most affected by global warming[46]. One of the purposes of this literature is to predict recreationists' adaptation to climate change. In the context of skiing, several studies find that spatial substitution would be the most frequent response, followed by temporal substitution and activity substitution (e.g. Rutty et al., 2015). Richardson and Loomis (2004) examine visitation rates to North American mountains under moderate warming and extreme heatwave scenarios. They find that visits will increase in the former case whereas they will notably decrease in the latter. Studies on the impacts of CC on sun and beach tourism also indicate that under certain CC scenarios with sea level rises or in which temperatures become uncomfortably hot, tourists would travel to different destinations or change the dates for travelling (Scott et al., 2012; Atzori et al., 2018; Lithgow et al., 2019)

Concerning the impact of CC on tourists' destination choice for the Spanish case, the literature has mainly focused on its effects for coastal destinations. Using a gravity model for domestic flows during the summer season, Priego et al. (2015) conduct a simulation study to see the consequences of a generalized rise in temperature on the redistribution of flows. According to their results, northern regions would notably increase their arrivals at the cost of southern regions (mainly Andalusia). At the individual level, Bujosa and Rosselló (2013) conduct a similar analysis to assess the impact of two CC scenarios on choice probabilities. Their results also show that colder Northern provinces would be the most benefited from a rise in temperature, whereas trips to south provinces would significantly decrease. In a similar vein, Bujosa et al. (2015) document that CC will produce a reallocation of tourists across Spanish coastal regions during the high season. Their estimates suggest that as Mediterranean and southern coastal regions will become too warm, they will become less attractive and lose part of their market share. These authors find that the probability of choosing a destination increases at a decreasing rate up to 39ºC, beyond which a rise in temperature reduces it.

Finally, it is important to highlight that a range of meteorological variables have been used for measuring climate conditions: temperature, humidity, rainfall, sunshine hours and wind, among others. Furthermore, tourism climatic indexes such as the early proposed by Mieczkowski (1985) have also been widely adopted (Amelung et al., 2007; Goh, 2012). Additionally, to avoid problems of multicollinearity, temperatures alone have

[46] Detailed reviews of tourism demand responses to climate change can be found in Gössling et al. (2012) and Steiger et al. (2019).

been by far the most used measure (Serquet and Rebetez, 2011; Rosselló-Nadal, 2014). Among those using temperature, different definitions have been considered: average temperature (e.g. Bigano et al., 2006), maximum temperature (e.g. Maddison, 2001), average temperature in the warmest month (e.g. Lise and Tol, 2002) and monthly maximum and minimum temperature (Rosselló-Nadal et al., 2011).

Prices

Prices constitute a third major determinant of tourist destination choice. According to microeconomic theory, prices must exhibit a negative relationship with demand. Therefore, given a budget constraint, the more expensive a destination is, the lower the probability of being selected, *ceteris paribus*. This expected negative relationship between prices and recreational demand has been documented in Morey et al. (1991), Morley (1994), Riera (2000a), Van Nostrand et al. (2013) and Bujosa et al. (2015), to cite some.

However, other studies have shown mixed findings on the sign of prices on site choice. Given the experience nature of tourism products, there is an inherent uncertainty about the quality and the characteristics of a destination, especially when the individual has not been there before. Under uncertainty, prices are perceived by consumers as signals of quality (Keane, 1997; Rao, 2005). Hence, consumers might infer inferior quality from a low-priced product. Depending on tourists' risk aversion, they might be willing to pay high prices to guarantee a certain standard of quality (Alegre and Juaneda, 2006). In such cases, the negative relationship between prices and demand could be *empirically* inverted. Nevertheless, this cannot be interpreted as evidence against the law of demand but simply that in those cases quality is not being properly controlled for.

In this regard, a growing body of research has studied how loss aversion and anchoring explain tourists' reaction to prices. This literature considers that the same prices could be perceived as cheap or expensive depending on past experiences, which act as benchmarks. Nicolau (2012) examines loss aversion by comparing tourists' choice of destination when destination prices are above and below reference points, measured as the price paid at the last purchase occasion. Consistent with prospect theory, he finds that individuals are price loss averse, since losses weight more than gains. A similar analysis is conducted by Nicolau (2011), who provide evidence that cultural tourists exhibit lower loss aversion, thereby being less reluctant to pay more than expected. Park and Nicolau (2018) examine the potential existence of irrational behaviour in tourism demand by means of a sticker shock model in which alternatives are evaluated based on the comparison between actual and reference prices. Their estimates show that, on average, price has a negative effect on destination choice, but 21% of tourists are willing to pay higher than expected prices. More recently, Tanford et al. (2019) experimentally explore price anchoring effects, showing that price judgements are asymmetric and willingness to pay is significantly higher under high anchors. What is more, the literature has even pointed to the possible existence of conspicuous consumption in tourism so that travelers might be willing to pay a premium price just to exhibit their wealth and meet their social need for esteem (e.g. Kim and Jang, 2013).

Some scholars have explored the linkages between heterogeneity in price sensitivity and subsequent expenditures. For a sample of Spanish domestic tourists, Nicolau (2009) finds that the relationship between price sensitivity to regional prices and expenditure is smile-shaped. In doing so, price sensitivity is estimated in terms of the effect of regional price indexes on regional choice probabilities. Nicolau and Masiero (2013) replicate a similar analysis using a choice experiment in Switzerland. They find the same pattern: price sensitivity exhibits a convex relationship with expenses.

Other studies have explored whether income and tourist motivations moderate the negative effect of price. Nicolau (2010b) examines price sensitivity for a sample of tourists travelling within Spain, and how it relates with income. His results show that income moderates the negative effect of prices on regional choice probabilities, but the moderating effect is non-linear. This means that from an income threshold point onwards, tourism demand becomes more elastic. Using the same data, Nicolau and Mas (2006) study how price sensitivity varies depending on tourists' motivations. Their results show that cultural tourists who look for discovering new places are willing to pay higher prices. However, those who declare 'search for climate' as their trip motivation are more deterred by high-priced destinations. Similarly, Masiero and Nicolau (2012) report that the dissuasive effect of prices is heterogeneous in the population and strongly related to motivations. The largest price sensitivities are found among those whose main motivation is 'free of charge', such as being physically active or experiencing landscape and nature. Conversely, those who seek new experiences, which are usually highly priced, show lower price sensitivity.

A key issue is how to measure destination prices. Since tourism consists of numerous components, it is unlikely that tourists know all costs beforehand. Researchers have been forced to use of different proxies. A review of the different measures used in the literature can be found in Dwyer and Forsyth (2011).

Morley (1994) argues that the consumer price index (henceforth CPI) for tourist services of a region is a good proxy of actual tourist prices, since CPI are found to be highly correlated with tourism expenditures over time[47]. Empirical studies that use CPI include Capacci et al. (2015) for Italian provinces and Yang et al. (2014) for domestic tourism in China. However, the use of price indexes exhibits important shortcomings. There is wide discussion in the economic literature about whether they properly measure the cost of living. Basically, the necessity of fixing quantities to control for price changes over time does not allow for changes in the basket composition due to substitution effects (Moulton, 1996; Hausman, 2003). Other well-known sources of biases in CPI are due to quality improvements and the appearance of new products (see Boskin et al. (1998) and Nordhaus (1998) for in-depth discussions on this). Furthermore, 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 (Massidda and Etzo, 2012; Marrocu and Paci, 2013).

Alternatively, Eugenio-Martín and Campos-Soria (2010) use a hotel price index for three and four-star hotels. However, this can only be understood as a proxy, since not all the tourists lodge at hotels and staying at a destination involves more costs than the

[47] His analogy between prices and expenditure imposes the price inelasticity of demand.

accommodation. Similarly, package tour prices (Papatheodorou, 2002) or the Big Mac Index (Dwyer et al., 2000) have also been used. Others like Eymann and Ronning (1997) and Nicolau (2011) have proposed to construct a specific cost index (also referred as quasi-hedonic prices) for each destination and individual. The procedure consists of regressing individual expenditures on length of stay and sociodemographic characteristics and then construct an estimated "cost" using the obtained parameters and average values of the stay of all the tourists that visit destination j. However, this is subject to criticism since the price construction considers quantities and therefore leads to endogeneity concerns. Moreover, evidence by Hill and Menser (2008) shows that the obtained quasi-hedonic prices are highly sensitive to the functional form used.

Eymann and Ronning (1992) argue that individuals do not only compare prices across possible destinations, but they also compare them with the existing ones at the place of residence, which act as a reference point. This issue is relevant for international tourism, where price differences are more pronounced. In this regard, to measure price competitiveness, many studies have used CPI differentials between the origin and potential destinations (adjusted for exchange rates in the case of international tourism), both for explaining aggregate flows (Morley, 1998; Rosselló et al., 2004; Nordström, 2005; Garín-Muñoz, 2006; Garín-Muñoz and Montero-Martín, 2007; Garín-Muñoz, 2009; Massidda and Etzo, 2012; Pattuelli et al., 2013) and individual choices (Nicolau, 2008; 2009; 2010b; Chandra et al., 2014; Bernini et al., 2017).

Other relevant attributes

In addition to distance, climate and prices, other attributes are usually considered. The literature agrees that nature-based tourists attach importance to the existence of national parks and its recreation facilities (Riera, 2000b; Neuvonen et al., 2010), the size of protected natural areas (Marrocu and Paci, 2013; Bernini et al., 2017) and the environmental diversity (Bujosa and Riera, 2009). Additionally, previous research shows that factors such as the environmental quality (Richardson and Loomis, 2004) and tranquillity (De Valck et al., 2017) also matter for those seeking nature-based recreation.

Another relevant attribute is the availability of ski tracks. Winter tourism is heavily dependent on snow cover and depth at mountain resorts (Falk, 2010). Indeed, for skiing tourists, it is not only the existence of ski runs but the number and variety of ski tracks available what matters (Konu et al., 2011).

A final set of factors that affects destination choice are risk-related issues. A large body of research has documented that security threats in the form of terrorism, crime, corruption and disease negatively impact choice probabilities (Santana-Gallego et al., 2019; Seabra et al., 2020). These factors thus act as impedance, making some destinations to be disregarded. The same applies to political instability. For the case of Catalonia, Perles-Ribes et al. (2019) report that arrivals and expenditure decreased in the final quarter of 2017 after some political events. Notwithstanding, for some people risky destinations can even be appealing. Lepp and Gibson (2008) report that tourists with high levels of sensation seeking (i.e. a personality trait associated with the need for novelty and stimulation) are more likely to travel to risky destinations. Moreover, there seems to be relevant gender differences in risk aversion, being males more willing to take high-risk travel activities (Kim and Seo, 2019).

An overview of the research question, the model and the results of some recent related empirical studies is presented in Table 2.1.

Reference	Research question	Model	Empirical results
Zimmer et al. (1995)	Seniors' choices regarding whether to travel or not and choice of destination.	MG-DFA[a]	Income, education and living in a city positively affect the probability of travelling to distant locations, whereas age and a bad health status make people to choose nearby regions.
Nicolau and Mas (2005)	The determinants of the decision to take a vacation, the choice between travelling domestically or abroad and the decision to take a single or a multi-destination trip.	RP-Logit[b]	Personal constraints, together with sociodemographic and psychographic characteristics are the main determinants of multistage travel organization.
Nicolau and Mas (2006)	The moderating role of motivations on the negative effect of price and distance exert on destination choice.	RP-Logit[b]	Interest in discovering new places, the search for climate and visiting family and friends mitigate the negative effect of distance.
Hong et al. (2006)	Explore the roles of categorization, affective image and constraints in forming destination choice sets and the final selection.	NLM[c]	Tourists' categorize alternative destinations into groups with similar characteristics and then decide among them based on affective image. In addition, 'active' and 'exhilarating' parks are the most likely to be chosen.
Lam and Hsu (2006)	How push and pull travel motives affect destination choice of Taiwanese travelers to Hong Kong.	SEM[d]	As the number of previous visits to Hong Kong increases, Taiwanese travelers' intention to revisit the destination also increases. Attitude towards the destination did not exert any effect on the choice of Hong Kong as destination.
Correia et al. (2007)	The determinants of Portuguese tourists' choices to travel to Latin America as opposed to Africa.	RP-Logit[b]	Gastronomy, nightlife and lodging are the main factors that pull Portuguese tourists to Latin America.
Pestana-Barros et al. (2008)	Tourists' choices between African and Latin American destinations.	RP-Logit[b]	Destination attributes and the socioeconomic profile of tourists are the main factors that affect the choice of destination when travelling abroad.
Nicolau (2008)	The role of sociodemographic characteristics in explaining tourists' sensitivity to distance.	RP-Logit[b]	Income, the size of the city of residence, organizing the trip through intermediaries and travelling by plane reduces the disutility of distance.
Nicolau and Mas (2008)	What is the order of choices when deciding where to travel?	RP-Logit[b]	The coastal nature of the destination precedes the decision about visiting a rural or an urban area. Tourists' choices exhibit a clear hierarchical pattern.
Lyons et al. (2009)	Destination-specific and individual-specific determinants of destination choice for holiday.	C-MNL[e]	Temperature, GDP and coastline positively affect choice probabilities, whereas age and distance have a negative effect on choice.
Hsu et al. (2009)	What are the factors that determine tourists' choice of destination?	AHP[f]	Visiting friends/relatives, safety, desire for scape and relaxation and destination image emerge as the most relevant factors for inbound tourists to Taiwan.
Nicolau (2010a)	The effect of distance on destination choice, focusing on tourists' motivations.	RP-Logit[b]	Variety-seeking behavior decreases the dissuasive effect of distance. Conversely, for tourists who exhibit inertial behavior, distance acts as a deterrent factor.

Reference	Research question	Model	Empirical results
Nicolau (2010b)	Tourists' price sensitivity when choosing where to go on holidays and the moderating role of income.	RP-Logit[b]	For some individuals, higher prices do not necessarily reduce utility. Income moderates the negative effect of prices but up to a threshold.
Eugenio-Martín and Campos-Soria (2010)	The effect of climate in the region of origin on the decision to travel domestically or abroad.	SUR-Biprobit[g]	Climate at the region of residence is a strong determinant of holiday destination choice. Individuals from regions with better climates have a higher likelihood of travelling domestically than abroad.
Nicolau (2011)	Differentiated loss aversion depending on cultural interest.	RP-Logit[b]	Cultural tourists appear to be less loss averse and to be willing to pay higher prices if they perceive that the destination worth it.
Eugenio-Martin and Campos-Soria (2011)	The role of income in the decision of travelling domestically, abroad, both of them or non-participation.	SUR-Biprobit[g]	Tourism demand is income elastic. Above an income threshold, a substitution pattern between travelling domestically or abroad takes place. The probability of the latter keeps growing while the former remains constant.
Nicolau (2012)	The existence of price loss aversion in destination choice.	RP-Logit[b] and cluster analysis	Tourists are loss averse when selecting a holiday destination. Price losses in comparison to the previous price paid are greater than gains. Moreover, there is large heterogeneity in price responsiveness.
Wu et al. (2011)	A better understanding of the heterogeneous interdependencies between destination and travel party choice.	LC-NLM[h]	Travel time, the attractiveness of the destination and the number of tourism spots were found to be important factors in destination choice, and gender, age and marital status on travel party choice.
Grigolon et al. (2013)	The determinants of the different facets arranged during the vacation planning process.	Binary Mixed-Logit Panel	The closer to the date of the trip, the higher is the probability of a facet being planned. The level of planning differs according to the life cycle stage, income and travel experience.
Bujosa and Rosselló (2013)	The impact of climate change on Spanish domestic coastal destination choice.	RP-Logit[b]	While Spanish northern colder regions would increase their arrivals under a rise in temperatures, regions located at the South would suffer a significant lost in tourist visitors, especially in the summer season.
Yang et al. (2013)	The role of spatial configuration of destinations on tourists' multi-destination choices.	NLM[c]	The longer the length of the stay at a given destination, the less likely for a tourist to continue on tour to another destination. Older tourists and those travelling with family and friends are the most likely to engage in multi-destination trips.
Van Nostrand et al. (2013)	American's holiday destination choices and time allocation patterns to the chosen trips.	MDCEV[i]	Higher travel time/travel costs reduce the attractiveness of a destination. Conversely, the length of the coastline and moderate temperatures are two of the dimensions that tourist most appreciate.
Eugenio-Martín and Campos-Soria (2014)	How the economic crisis has affected tourism expenditure.	SU-SOBP and S-SOBP[j]	Tourists' cutback decisions on tourism expenditures depend on climate conditions of the place of origin, GDP and GDP growth.

Reference	Research question	Model	Empirical results
Bujosa et al. (2015)	The effect of temperature and other destination-specific attributes on destination choice for the case of Spanish coastal tourism.	NLM[c]	The probability of choosing a destination increases at a decreasing rate with rises in temperature up to 39ºC. The climatic change might damage the market shares of the southern coastal regions in Spain.
García et al. (2015)	Tourists' preferences for the 'all-inclusive' travel mode.	MNL[k] and SLM[l]	The decision structure of those who prioritize destination is significantly different from that of those who prioritize the travel mode.
Mussalam and Tajeddini (2016)	Which destination attributes tourists value the most depending on whether they plan a short or a long holiday.	ANOVA and Cluster Analysis	Price, security, food quality and culture are the most important attributes for short stay holidays, whereas location, quality and variety of accommodation, natural resources, and availability of tourist information are among the most preferred ones for long stays.
Crouch et al. (2016)	How recent past tourism experience choices relates to future experience preferences.	LC-MNL[m]	There is a strong positive relationship between past experience choices and preferences for future trips. Tourists display stable preferences that can be inferred from recent past choices.
Bhat et al. (2016)	Joint modelling of recreational destination choices and the number of trips undertaken to each destination.	MDCP[n]	Youngers are more willing to be loyal to a destination if it offers a wide range of diversity. Young parents and seniors look for diversity by travelling to different regions.
Yang et al. (2018)	The effect of both geographic and cultural distance for past, dream and intended destination choices.	RP-Logit[b]	The negative effect of geographic distance on past and intended choices has diminished over time, whereas it is positively related to dream destinations.
Masiero and Qiu (2018)	How leisure tourists' past experiences affect current long-haul destination choices.	RP-Logit[b]	Tourists exhibit a reference-level bias by which destinations are judged relative to past experiences. They seem to be loss averse with reference to transport services and cultural, natural and entertainment attractions.
Park and Nicolau (2018)	Analysis of possible anomalies in tourist behavior in terms of price sensitivity.	RP-Logit[b]	For some tourists, prices act as a signal of quality so that, under risk aversion, people are willing to pay higher than expected prices to guarantee a certain level of quality.
Gosens and Rouwendal (2018)	Destination and travel mode choice for recreational trips together with the allocation of time to outdoor activities.	MDCEV[i]	Personal characteristics and accessibility both influence time allocation decisions. Higher temperatures increase time spent in outdoor recreation.

Table 2.1.- Summary of related literature

[a] Multiple Group Discriminant Function Analysis
[b] Random Parameter Logit Model (also referred as Mixed Logit Model or Random Parameter Multinomial Logit Model)
[c] Nested Logit Model
[d] Structural Equation Modelling
[e] Conditional Multinomial Logit Model
[f] Analytic Hierarchy Process
[g] Seemingly Unrelated Bivariate Probit Model
[h] Latent-Class Nested Logit Model
[i] Multiple Discrete-Continuous Extreme Value Model
[j] Simultaneous Semi-Ordered Bivariate Probit Model and Seemingly Unrelated Semi-Ordered Bivariate Probit Model
[k] Multinomial Logit Model
[l] Sequential Logit Model
[m] Latent Class Multinomial Logit Model
[n] Multiple Discrete-Continuous Probit Model

3. THEORETICAL FRAMEWORK

Our theoretical model for destination choice is based on the *Lancasterian* product characteristics approach (Lancaster, 1966). We build on the work of Seddighi and Theocharous (2002) to develop a conceptual framework for theoretically describing how consumers decide where to travel for a holiday depending on their personal characteristics and the attributes of the destinations.

Contrary to traditional demand theory that postulated that goods themselves produce utility, Lancaster (1966) introduced the notion that it is the characteristics of the goods from which utility is derived. This framework is especially suitable in the tourism context. Tourism can be understood as trade in services that is to a large extent determined by local natural and cultural features. Individuals obtain utility from being at a particular location for a period of time, during which they consume and enjoy the location-specific attributes of the destination, such as its climate or a relaxing environment.

Consider a given number of J alternative destinations characterized by a finite number of K objective and observable characteristics (X_{kj}). Each destination j can be thus regarded as a source of systematic utility (V_{ij}) in the form:

$$V_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} \quad (2.1)$$

where β_{ik} is the marginal utility of attribute k for individual i. Therefore, marginal utilities are allowed to vary in the population. This is because consumers have subjective perceptions (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 assumed to be the same for the entire population (b_k), and ii) an individual-specific component that varies across individuals and adds stochastic variation to the marginal utilities. This latter component can be further partitioned into two elements. The first one is a vector of observable individual characteristics Z_i, and the second one is a composite factor v_{ik} for random preference heterogeneity for each attribute. 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:

$$\beta_{ik} = g(b_k, Z_i, v_{ik}) \quad (2.2)$$

Since the characteristics of the goods are objectively given, consumers make choices among bundles of characteristics. Under his framework, preference rankings over goods are derived based on the characteristics they possess conditional on the individual-specific marginal utilities. A multidimensional preference map can be thus constructed according to the preferred bundle of characteristics (Lancaster, 1966). Since individuals are subject to a budget constraint, optimal choice does not only imply choosing the alternative with the bundle of characteristics that maximizes utility but also the one that minimizes costs. The systematic utility can be expanded to include the associated price for each destination j[48]:

$$V_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j \tag{2.3}$$

Destinations with a similar collection of characteristics can be regarded as close substitutes. The model would predict higher cross price elasticities for close substitutes. From a welfare perspective, higher differentiation makes it easier for individuals to find their desired bundle of characteristics.

Conditional on the marginal utilities, Lancaster's product characteristics approach is deterministic. A limitation of such assumption is that the model can predict zero market shares for some destinations. This makes the model quite sensitive to measurement error[49]. A common way to deal with this is to include an additive source of residual utility, generally in the form of an iid disturbance term (ε_{ij}). This idiosyncratic random component is independent across alternatives and individuals and measures the error in the calculation of the utility associated with each destination. Therefore, the utility of destination j for individual i can be expressed as:

$$U_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j + \varepsilon_{ij} = \sum_{k=1}^{K} g(b_k, Z_i, v_{ik}) X_{kj} - \gamma P_j + \varepsilon_{ij} \tag{2.4}$$

Lancaster's product characteristics approach is derived from the consumer's perspective. From that perspective, all the sources of utility are known by the individual. However, at the empirical level, not all the relevant attributes of the destinations are observed. Therefore, the utility is further expanded with the inclusion of an additive term ξ_j for unobserved (from the econometrician perspective) destination-specific characteristics (Murdock, 2006). As a result, the utility is given by:

$$U_{ij} = \sum_{k=1}^{K} \beta_{ik} X_{kj} - \gamma P_j + \xi_j + \varepsilon_{ij} = \sum_{k=1}^{K} g(b_k, Z_i, v_{ik}) X_{kj} - \gamma P_j + \xi_j + \varepsilon_{ij} \tag{2.5}$$

The term ξ_j is generically considered to be of dimension J, but it can also be defined to have a lower dimension. According to our previous discussion, close substitutes share similar observed features and might also share the same unobserved characteristics.

This is the *characteristics-based* demand model, widely applied in the analysis of differentiated products in industrial organization. This model has been shown to provide several advantages over the *taste for products* model, such as not imposing limits on the substitution patterns between alternative goods (destinations) or the advantage of setting a utility bound as the number of alternatives increase (see Berry and Pakes (2007). The model does not consider quantities because it is implicitly assumed that the consumer

---

[48] Without loss of generality, we assume the price indicator to be a weighted sum of the prices of accommodation, entertainment and catering (Papatheodorou, 2001). In principle, the effect of price on utility could be allowed to be heterogeneous in the population (i.e. $\gamma_i \neq \gamma_{i'}$ for $i \neq i'$).

[49] Athey and Imbers (2007) provide a clear illustration of this issue by showing that when the optimal bundle of characteristics based on the model prediction is not observed in the data, then all the remaining alternatives are equally likely. However, the analysist might expect alternatives with utilities closer to the optimal to be more likely to be chosen.

chooses a single unit of the good (only travels to one destination) at a time. Therefore, we model probabilistic demand.

Rugg (1973) was the first who applied the *Lancasterian* product characteristics approach to the choice of a journey destination, introducing monetary and time constraints. Morley (1992) extended Rugg's contribution by modelling holiday destination choice. He incorporated the decision to travel or not before the destination choice itself, an issue that had not been addressed before. In this sense, it is important to highlight that here the choice of destination for a holiday trip is necessarily conditional on having decided to travel (Eugenio-Martín 2003). We need to further impose that preferences for leisure activities are weakly separable (Deaton and Muellbauer, 1980) so that tourism demand can be expressed independently of non-tourism prices. As such, our model picks up at the second stage of budget decomposition, after income has been allocated for expenditure in tourism activities. Similarly, at the time of choosing a destination the consumer is assumed to have allocated a positive amount of time to travelling.

4. DATA

4.1. Database

Our database is drawn from the Spanish Domestic Tourist Survey (*ETR/FAMILITUR),* 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, accommodation dwelling, expenditure and sociodemographic characteristics, among others[50]. Participants are interviewed at their homes by telephone about trips that have taken place two months before.

Our study covers the period between February 2015 and September 2017. Therefore, we have information about trips undertaken over 32 months. However, our database does not have a panel structure but is a pool of monthly cross-sectional units. Since we are interested in modelling domestic destination choice, we only consider trips within Spain whose main purpose was leisure and holidays[51]. Hence, international trips are not included. Moreover, we restrict our sample to those who declare that nature and/or sport is their main travel purpose[52]. This trip purpose represents about 18 percent of leisure trips. After excluding some observations with missing values in the variables of interest, we have valid information for 6,661 tourists that take a nature-based domestic holiday to any of the 17 Spanish regions (NUTS 2). Trips to Ceuta and Melilla are excluded.

[50] A trip is defined as any journey away from the usual residence that implies, at least, an overnight stay and lasts for less than a year. Same-day trips are therefore not considered.
[51] In this way, visiting friends or relatives or job-related trips are not included in our database since in these cases the tourist does not choose where to go (i.e. the destination is exogenously given).
[52] Among them, those who declare they travel to a destination to attend a sport event (e.g. a soccer match) (325 individuals) are also excluded.

Summary statistics of a selection of sample characteristics are provided in Table 2.2. Although we only use part of these variables in the analysis, we present detailed descriptive statistics with the aim of properly characterizing the nature-based tourist profile in our data.

Variable	Description	Mean	SD	Min	Max
age	Age in years	43.23	12.45	15	85
female	Respondent is a 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
income1	Income: Less than 1,500 euros per month	0.26	0.44	0	1
income2	Income: Between 1,500-3,500 euros per month	0.57	0.49	0	1
income3	Income: More than 3,500 euros per month	0.15	0.36	0	1
LOS	Length of the stay at destination (days)	3.36	3.43	1	30
weekend	Travels for a weekend	0.51	0.49	0	1
q1	Travels in the first quarter	0.20	0.40	0	1
q2	Travels in the second quarter	0.25	0.43	0	1
q3	Travels in the third quarter	0.37	0.48	0	1
q4	Travels in the fourth quarter	0.16	0.36	0	1
*wint_sports**	Trip activity: winter sports practice (i.e. skiing, snowboarding)	0.07	0.25	0	1
*moun_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 2.2.- Descriptive statistics (N=6,661)
*Note: these activities are not mutually exclusive

Our sample of domestic nature-based tourists comprises slightly less females (46%) than males, with a mean age of 43 years old. Respondents are relatively highly educated (60%) and mostly participating in the labor market (74%). About half of the sample is married and the average number of travel companions is 2.2. Most respondents have middle income (56%) and the mean length of the stay at destination is 3.3 nights. Half of the sample travels during weekends (51%), being the third quarter of the year when more trips take place (37%). Regarding trip activities, mountaineering and trekking is the most declared trip activity (72%), followed closely by visiting natural areas (56%). Interestingly, a non-negligible share of respondents travels to perform adventure and risky sports (22%), whereas 16% opt for visiting rural areas and villages. Finally, only 7% of the sample practices winter sports. This low figure is explained due to the following. Since our data is a pool of monthly cross-sectional units and winter sports are mainly practised during the first and fourth quarters, this causes its share to be low.

As for the geographical composition of the sample, Figure 2.1 depicts the percentage of nature-based tourists according to the region of residence (NUTS 2). Darker colours imply larger shares. Madrid, Catalonia, Andalusia and the Basque Country are the regions with the highest shares of outbound tourism. Similarly, Figure 2.2 maps the percentage of inbound tourists by region. The specific figures for each destination, grouped at two levels of administrative aggregation (NUTS1 and NUTS2, respectively) are presented in Table 2.3. Here, we further decompose the total number of visitors to each region based on whether their main trip purpose is enjoying nature (*Nature*) or practicing any kind of sport (*Sport*). This way, *Nature* gathers mountaineering, trekking, visiting natural areas and villages (74 percent), whereas *Sport* collapses aquatic, winter and adventure sports (25 percent). This variable definition is provided in the survey. All the regions comprise at least 1 percent of the sample, which is the common threshold point for keeping destinations among the choice set (Bujosa and Riera, 2009).

Aragon is the region with the largest number of tourists (15.4%) for both nature- and sport-related purposes, followed by Catalonia (14.3%). Conversely, Murcia and La Rioja are the regions with the lowest number of visitors (1.1% and 1.6%, respectively). If we focus our attention to sport tourism, Aragon and Catalonia are the two most preferred areas (21.2% and 15.2%, respectively). For the case of nature-related tourism, Catalonia is the region with the largest share of tourists (14.1%). Apart from Aragon (13.2%), Castile and Leon also accounts for a large share of tourists (11.8%). From a more aggregate perspective, the North-East (22.7%) and the East (21.5%) areas of Spain are the ones with the greatest number of tourists. These two zones concentrate almost half of the total domestic flows for nature and sport purposes.

				Sport		Nature	
NUTS 1	**NUTS 2**	**Number of individuals**	**Share (%)**	**Number of individuals**	**Share (%)**	**Number of individuals**	**Share (%)**
North-West (1)	Cantabria	346	5.19	64	3.72	282	5.70
	Galicia	307	4.60	51	2.97	256	5.17
	Asturias	460	6.90	84	4.89	376	7.60
	Total	1,113	16.70	199	11.58	914	18.48
North-East (2)	Aragon	1,024	15.37	368	21.43	656	13.26
	Basque Country	182	2.73	67	3.90	115	2.32
	La Rioja	105	1.57	23	1.33	82	1.65
	Navarre	207	3.10	23	1.33	184	3.72
	Total	1,518	22.78	481	28.01	1,037	20.97
Community of Madrid (3)	Madrid	226	3.39	81	4.71	145	2.93
Centre (4)	Castilla-La Mancha	353	5.29	53	3.08	300	6.06
	Castile and Leon	713	10.70	130	7.57	583	11.79
	Extremadura	202	3.03	22	1.28	180	3.64
	Total	1,268	19.03	205	11.93	1,063	21.50
East (5)	The Balearic Islands	108	1.62	45	2.62	63	1.27
	Catalonia	958	14.38	261	15.20	697	14.09
	Valencian Community	368	5.52	94	5.29	274	5.54
	Total	1,434	21.52	400	23.29	1,034	20.91
South (6)	Andalusia	781	11.72	239	13.91	542	10.96
	Murcia	77	1.11	33	1.92	44	0.88
	Total	858	12.88	272	15.84	586	11.85
The Canary Islands (7)	The Canary Islands	244	3.66	79	4.60	165	3.33
Total		6,661	100	1,717	100	4,944	100

Table 2.3.- Inbound tourists per region, disaggregated by travel purpose.

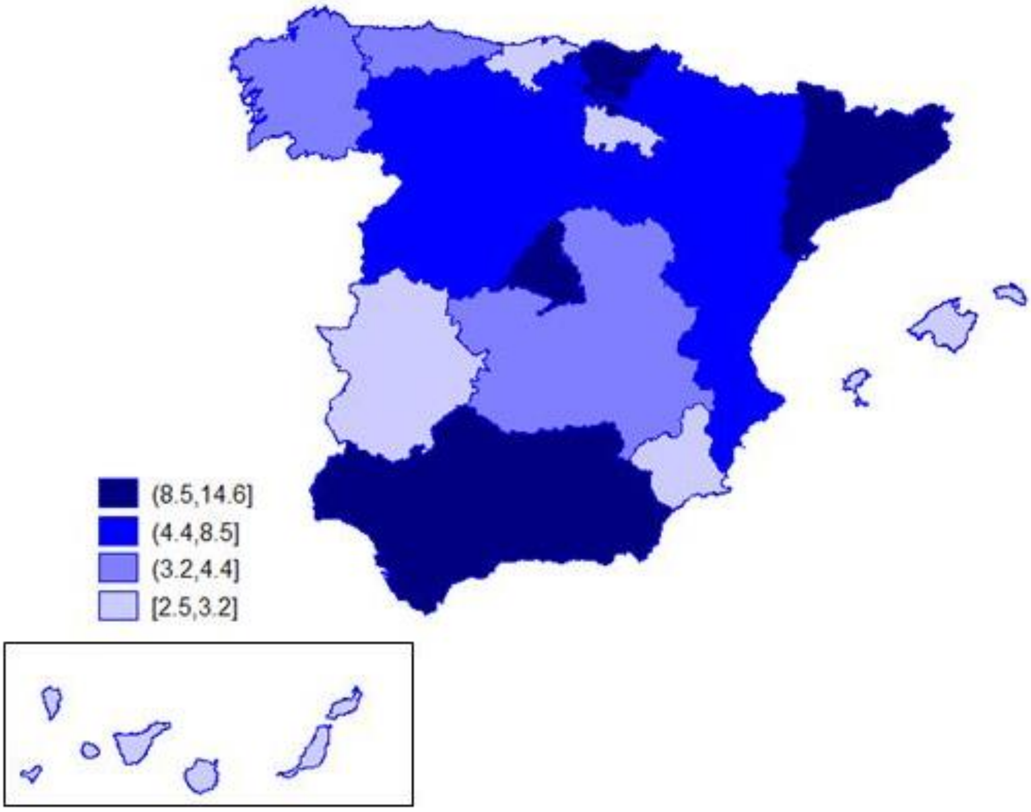


Figure 2.1.- Percentage of nature-based outbound tourists by region

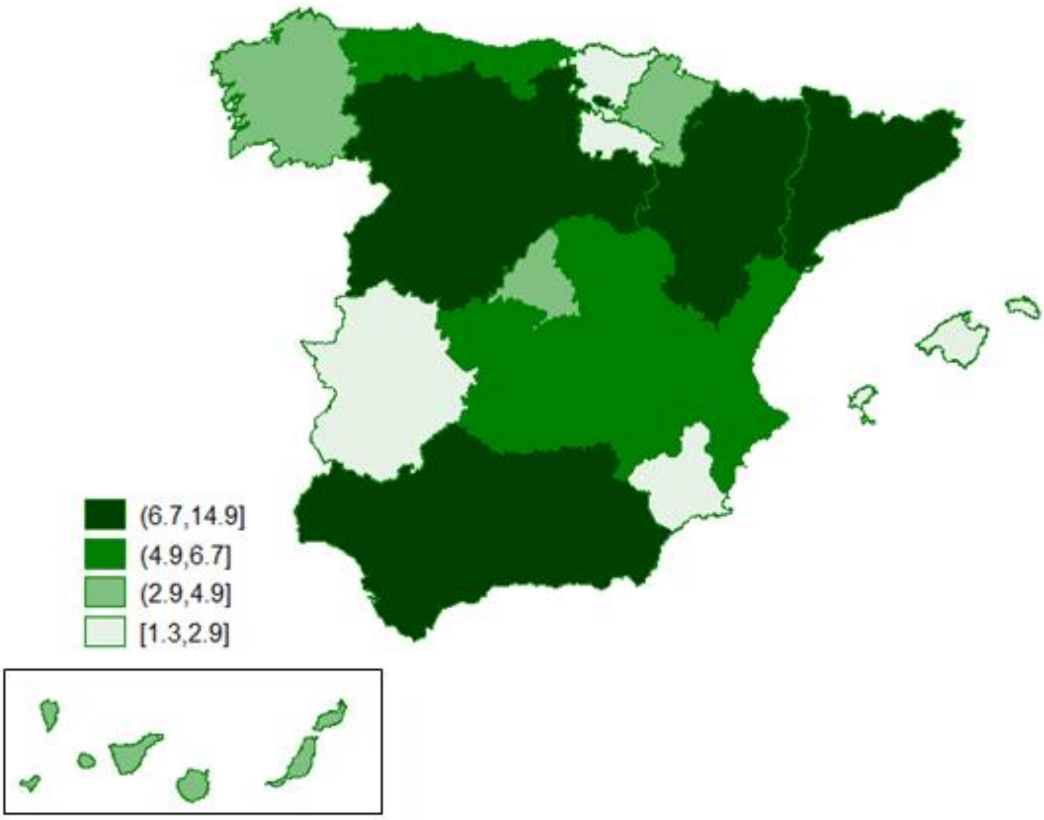


Figure 2.2.- Percentage of nature-based inbound tourists by region

4.2 Destination attributes

As introduced in Section 3, our destination choice modelling assumes that individuals choose where to go based on place-based attributes *á la Lancaster*. The following regional factors are considered in the analysis.

Distance

Although formally this is not a destination characteristic, the rationale for including it here is that, from the viewpoint of the tourist, any alternative destination is distant or nearby from her perspective. Therefore, distance can be seen as a destination feature.

We use Euclidean distance between the origin and each possible destination, which is the most used measure (e.g. Massidda and Etzo, 2012; Marrocu and Paci, 2013; Yang et al., 2018; Gosens and Rouwendal, 2018; Hasnat et al., 2019). Related studies in recreational demand compute travel costs to have a monetary measure of the effect of geographical distance. However, since we are working with trips that involve long distances, that would require defining the chosen mode of transport. This is endogenous and jointly determined with the choice of destination. If we restricted the sample only to tourists using road-based transport modes, as done by Bujosa and Rosselló (2013), this would limit the scope of our analysis since neither the Balearic nor the Canary Islands could be considered. The common practice of defining travel costs as the product of distance and €0.19 per kilometer (Bujosa and Riera, 2009; Bujosa and Rosselló, 2013) is merely a scale adjustment. Alternatively, one might consider using travel time rather than the Euclidean distance. As before, here we have the same problems since we must define the mode of transport. Even if we focused on road-based means of transport, the fastest routes would require going through toll motorways in some cases but not in others, which would introduce discretion by the researcher. This is a common problem in the literature. Because of these reasons, although possibly not perfectly, we use the Euclidean distance as the cleanest way to measure geographical distance between the origin and potential destinations.

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 (Spanish Autonomous Communities, equal to 17). The latter hinders the direct calculation of distances between the origin and the destination since they are not defined at the same regional level. If we computed the Euclidean distance in kilometers between NUTS 2 regional centroids, that would set to zero the distances for all the trips that take place within Autonomous Communities (NUTS 2). Given the heterogeneity in size between Spanish regions, that would equally assume zero distance for *true* intra-regional trips (a tourist from Murcia that travels within Murcia) and for *apparent* intra-regional trips (a tourist from Jaen that travels to Cádiz)[53]. Furthermore, that would reduce the variability of the distance variable.

[53] 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.

To alleviate this, for each tourist in the sample we compute a Euclidean weighted measure of distance that considers individuals' place of origin probabilistically. We proceed as follows. First, we calculate the distance between the centroids of all Spanish provinces (NUTS 3). Second, we compute bilateral distances between each province (p) and each Autonomous Community (c) in the following way:

$$d_{c,p} = \sum_{p \in c} pop_p / pop_c * d_{p,p'} \qquad (2.6)$$

for p=1,…,50 and c=1,…,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. Accordingly, distances between the origin and destinations consider the likelihood of the tourists living in each province based on population weights. Since our data covers 32 months and official statistics about population size are updated biannually, the weights are adjusted over the study period.

Finally, since our analysis is performed at the NUTS 2 aggregation level and due to the difficulty of analyzing choices with a choice set with 50 alternatives, we take the weighted distance mean within Autonomous Communities so that $d_{c,c'} = \sum_{\forall p \in c} \mathrm{d}_{c,p} * 1/n$, where n indicates the number of provinces in each Autonomous Community[54].

The resulting weighted distance ($d_{c,c'}$) is labelled *DIST*. This measure closely mimics the one used by Chandra et al. (2014) to model cross-border travelling. Consistent with the discussion in Section 2, we expect distance to exert, on average, a negative effect on tourists' utilities. Notwithstanding this, our empirical modelling will allow for heterogeneity in this effect (see Subsection 5.2).

Climate

Trips are normally planned some time in advance so that the actual weather at each destination is difficult to forecast. Hence, we consider the expected (average) temperature and rainfall at each region in each month. Average temperature and rainfall per destination and month during the 2010-2015 period were obtained from the Spanish National Meteorology Institute. Based on this data, 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 individual's place of residence at month t:

$$r\_TEMP_{imt} = \frac{T_{imt}}{T_{ijt}} \qquad (2.7)$$

[54] Only 2% of the observations take value zero for DIST, which correspond with the regions where NUTS 2 equals NUTS 3 (e.g. Cantabria). If we used the distance between the NUTS 2 centroids, the share of zero values would raise to 5.88%.

where T_{imt} is the temperature at region m during period t and T_{ijt} is the temperature at the place of origin.

Accordingly, the higher (lower) the ratio, the warmer (colder) the destination relative to the origin[55]. Apart from being consistent with the 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), this ratio further captures the non-linearity of relative temperature differences. 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 11ºC relative to an origin with 10ºC (11/10=1.1), or a destination with 21ºC relative to an origin with 20ºC (1.05). In this way, contrary to alternative approaches such as using the absolute difference in temperatures (Lorde et al., 2015; Agiomirgianakis et al., 2017), the ratio captures that a difference in temperature between the origin and the destination does not have the same effect on utility depending on the level. However, if we fix the temperature at origin, then marginal increases in the ratio correspond to marginal increases in temperature at the destinations, which facilitates interpretation[56]. This linearity in marginal changes in temperatures at destination fixing the one at the origin is relaxed later (see subsection 5.2). We expect this variable to positively influence utilities (i.e. tourists are assumed to prefer, on average, warmer destinations).

Second, we define a dummy variable denoted by *RAIN* that takes value one if expected rainfall at month t excess 60 liters per square meter (or equivalently 60 mm). This threshold was chosen for two reasons. First, it is the 75th percentile of the rainfall distribution during the study period. Second, Mieczkowski's subindex of 3 (Mieczkowski, 1985) precisely equals less than 60 mm rainfalls per month, having this threshold also been used by Eugenio-Martín and Campos-Soria (2010). Therefore, we consider 60 mm to be a valid cut-off point to distinguish rainy from non-rainy destinations. We expect this variable to exert a negative effect on utility.

Prices

In line with related applications, we use the regional consumer price indexes as a proxy of prices. Specifically, we employ the price index for accommodation and tourism-related services (denoted as $TCPI$). Compared with the general consumer price index, this tourism-specific index has the advantage of better controlling for price changes in the items tourists are assumed to purchase. This data is drawn from the Spanish National Institute of Statistics. The year 2011 is the base period.

[55] One might wonder whether average temperature per region and month are valid for measuring climate conditions, especially for the case of ski tourism where the relevant temperatures might be the ones at mountain resorts. For the New Hampshire ski areas in the USA, Hamilton et al. (2007) report that temperatures at the distant city of Bolton significantly predict tourism attendance to ski areas. Therefore, we consider average temperature per region to be a valid indicator of climate conditions, even for winter sports.

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

Also consistent with previous studies, we alternatively define the ratio of *TCPI* between each possible destination j and the one at the place of origin k for month t as follows:

$$r\_TCPI_{ijt} = \frac{TCPI_{ijt}}{TCPI_{ikt}} \tag{2.8}$$

Consistent with economic theory, we expect *TCPI* and *r_TCPI* to be negatively related with choice probabilities, *ceteris paribus*.

Tourism spots

The use of aggregate zones (NUTS 2) instead of individual attractions requires to control for the variability in utilities across the individual alternatives that compose aggregate alternative j (see Bekhor and Prashker (2008) for a discussion). The number of tourism sightseeing spots per region (TOU_SPOTS) seems to be a relevant variable to measure the number of municipalities of interest within each region. A municipality is considered by the Spanish National Institute of Statistics to be a tourism spot if it specifically concentrates tourism affluence. Importantly, this definition is prior to our study period, thereby being an exogenous indicator. We refer the reader to Annex 2 in the Supplementary Material for the list of the 106 Spanish municipalities considered as tourism spots. This variable is expected to have a positive effect on the probability of a destination being chosen.

Ski kilometers

We computed the sum of available kilometres for alpine, Nordic and indoor ski at each region at each month during the ski season (November-April). This variable takes value zero for the rest of months 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.

National parks

Since 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. 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 other natural areas without such categorization where tourists can recreate[57]. This data is drawn from the Ministry of Ecologic Transition.

Coast

For those whose seek to practise aquatic sports, the presence of coast in the region appears to be a relevant factor. We define a dummy variable denoted by *COAST* that takes value 1 if the region has a coastline and 0 otherwise.

Table 2.4 presents summary statistics of the attributes introduced above along with notation, description and data source. Table 2.5 provides additional information by breaking down attribute mean values per region.

While Andalusia is the region with the highest number of tourism spots (24), Cantabria, the Community of Madrid, La Rioja and Navarre have only one. The Basque Country, the Valencian Community, Castilla-La Mancha and the Balearic and the Canary Islands do not have any ski run, whereas Aragon and Catalonia provide winter tourists with large snow tracks. Similarly, whereas the Canary Islands, Castile and Leon and Catalonia are the regions with the greatest number of natural and national parks, La Rioja, the Balearic Islands and Murcia do not have any park with such distinction. However, in terms of protected natural surface, Andalusia stands as the region with the largest area for nature recreation. As for climate, the average temperature across Spanish regions and months is 17.3ºC, varying from the minimum 4ºC in Castile and Leon in February to 28ºC in the Region of Murcia in August. In general terms, the coldest region is Castile and Leon with an average annual temperature of 12ºC, whereas Murcia is the warmest, with 19.4ºC. Regarding rainfall, the Basque Country, Cantabria and Asturias are the regions with the highest rainfall. Conversely, the Canary Islands is the driest[58].

[57] This includes Natural Reserves, Protected Landscapes and *Protected Areas Natura Red 2000*, among others.
[58] Further information about the distributions of the attributes is provided in Annex 1 in the Supplementary Material.

Attribute	Description	Mean	SD	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 temperature 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 rainfall is 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 2011)	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	0.58	0.44	0	1	Google Maps	NO

Table 2.4.- 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)

		Varies over time (mean values)					Fixed over time			
NUTS 1	**NUTS 2**	*r_TEMP*	*RAIN*	*TCPI*	*SKI_KM*	*DIST*	*TOU_SPOTS*	*NAT_PARKS*	*SIZE_NAT*	*COAST*
North-West (1)	Cantabria	0.98	0.61	103.5	9.50	440.49	1	6	1,509	1
	Galicia	0.91	0.46	102.1	5.72	605.47	9	5	3,594	1
	Asturias	0.88	0.61	103.0	10.36	324.04	5	6	2,358	1
North-East (2)	Aragon	0.94	0.00	101.3	144.03	375.32	5	4	1,682	0
	Basque Country	0.88	0.61	101.4	0.00	406.64	3	7	1,676	1
	La Rioja	0.92	0.00	101.2	6.86	362.72	1	0	1,013	0
	Navarre	0.86	0.49	101.6	11.04	390.30	1	2	830	0
Community of Madrid (3)	Madrid	1.00	0.00	100.7	10.54	350.04	1	1	1,208	0
Centre (4)	Castilla-La Mancha	0.96	0.00	101.1	0.00	379.54	5	7	5,828	0
	Castile and Leon	0.76	0.00	100.6	27.17	398.37	10	13	7,603	0
	Extremadura	1.10	0.27	101.3	0.00	520.04	5	1	3,165	0
East (5)	The Balearic Islands	1.17	0.20	104.5	0.00	555.46	2	0	747	1
	Catalonia	1.07	0.13	102.1	159.68	473.97	9	12	10,258	1
	Valencian Community	1.21	0.00	102.4	0.00	424.99	9	15	2,446	1
South (6)	Andalusia	1.20	0.21	101.5	42.15	566.58	2	9	26,083	1
	Murcia	1.27	0.00	101.9	0.00	474.25	2	0	621	1
The Canary Islands (7)	The Canary Islands	1.20	0.00	100.6	0.00	1,822	11	15	3,020	1

Table 2.5.- Attribute mean values per destination

5. EMPIRICAL MODEL

5.1. Econometric Modelling

Consistent with our theoretical framework, each of the J possible destinations is characterized by a set of observable attributes. We assume each individual considers the full choice set and chooses the destination that gives her the highest utility, given her characteristics and preferences for the attributes[59].

Our destination choice model is based on Random Utility Maximization theory (hereafter RUM), developed by McFadden (1974)[60]. Under this framework, the conditional utility function is the sum of a systematic and a random component[61]. Therefore, RUM states that the latent utility of individual i for choosing alternative j (U_{ij}^*) is given by two components:

$$U_{ij}^* = V_{ij} + \varepsilon_{ij} \tag{2.9}$$

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 attributes, taste variations and measurement errors. The choice alternatives (destinations) are defined in terms of regions (NUTS 2) rather than the spatial points where the tourist stays (city, municipality). As introduced before, this emerges because of the methodological difficulties of defining larger choice sets.

The systematic part of the utility function, which can be interpreted as the conditional mean of U_{ij}^* (up to a constant term), is assumed to be an additively separable linear-in-parameters function of the K attributes of each alternative j (X_{kj}) so that:

$$V_{ij} = X_{kj}'\beta_k \tag{2.10}$$

The probability that decision maker i chooses alternative j from the choice set S is:

$$\begin{aligned} P(i\ chooses\ j) = P_{ij} &= P(U_{ij}^* > U_{im}^*)\ \forall\, m \subset S \quad m \neq j \\ &= P(V_{ij} + \varepsilon_{ij} > V_{im} + \varepsilon_{im})\ \forall\, m \subset S \quad m \neq j \\ &= P(\varepsilon_{im} - \varepsilon_{ij} < V_{ij} + V_{im})\ \forall\, m \subset S \quad m \neq j \end{aligned} \tag{2.11}$$

Depending on the assumption about the distribution of the random terms we obtain different discrete choice models. If the random terms are IID type I Extreme Value

[59] Given the limited analytical capacity of people (Simon, 1955), this assumption might be restrictive since tourists may tend to simplify and make decisions following a hierarchical decision strategy. This may be especially true when faced with several alternatives and a high number of attributes (Nicolau and Mas, 2008).

[60] RUM is the most widely applied framework for the modelling of recreational and tourism demand (Thiene et al., 2017)

[61] Marschak (1960) was the first who introduced the idea that utilities were not purely deterministic but contained random elements. RUM was originally proposed by Thurstone (1927) and Luce (1959), but McFadden (1974) was the one who developed it properly.

(Gumbel) distributed across the J alternatives, we obtain the standard Multinomial Logit Model (MNL) (Ben-Akiva and Lerman, 1985), whose probability expression is given by:

$$P_{ij} = \frac{\exp(\lambda V_{ij})}{\sum_{j=1}^{J} \exp(\lambda V_{ij})} \tag{2.12}$$

where λ is a positive scale parameter that is inversely proportional to the standard deviation of the Gumbel error terms (Ben-Akiva and Lerman, 1985; Swait and Louviere, 1993)[62]. Because λ and β are not separably identified, λ is normalized to 1 for identification. The IID assumption of the error term precludes the possibility of correlation in the random component of utility across alternatives. We relax this later.

The Multinomial Logit Model has been widely used in many studies about tourism destination choice (e.g. Lyons et al., 2009). However, this model presents some shortcomings. First, it exhibits the well-known Independence of Irrelevant Alternatives (IIA) property, by which the ratio of probabilities for any two alternatives j and k does not change if a third (irrelevant) alternative is included in the choice set. In other words, it implies proportional substitution across alternatives, which imposes strong restrictions on cross-attribute elasticities and has been object of criticism since the so-called *red-bus blue-bus* problem highlighted by Debreu (1960). When this is not the case, coefficient estimates are still consistent but biased (Train, 2009). Second, it assumes homogeneity in respondents' preferences, not addressing potential taste heterogeneity[63].

These limitations have motivated researchers to develop a variety of alternative models[64]. Among them, the Random Parameter Logit (hereafter RPL) has become the most used. 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:

$$\beta_{ki} = b_k + \delta' Z_i + \sigma_k \nu_{ik} \tag{2.13}$$

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 certain probability distribution independent from ε_{ij}, and σ_k is the standard deviation of the distribution of β_{ki} around b_k.

---

[62] The variance of the Gumbel error terms is $Var(\varepsilon_{ij}) = \pi^2/6\lambda^2$. As $\lambda \to 0$ the MNL model predicts equal choice probabilities whereas $\lambda \to \infty$ the MNL predicts choice probabilities fully deterministically. Some authors refer to this model as the Conditional Logit and reserve the MNL to a model based on individual-specific characteristics only.

[63] The literature agrees that unobserved preference heterogeneity needs to be accounted for.

[64] Examples of models that relax the IIA hypothesis are the McFadden's Generalized Extreme Value (GEV) (McFadden, 1978) – which collapses the Nested Multinomial Logit model (NMNL) developed by Ben-Akiva and Lerman (1985) as a special case –, the Heteroskedastic Extreme Value (Bhat, 1995) and the Multinomial Probit model (Daganzo, 1979). However, they cannot represent all RUM-consistent behavior (McFadden, 2001).

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:

$$P_{ij} = \int P_{ij}|\beta_i \;\; f(\beta_i|\Omega)\, d\beta_i = \int \frac{\exp(V_{ij})}{\sum_{j=1}^{J} \exp(V_{ij})} \;\; f(\beta_i|\Omega)\, d\beta_i \tag{2.14}$$

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

The simulated log likelihood to be maximized for cross-sectional data is:

$$Log\ L \; = \sum_{i=1}^{N} \ln \frac{1}{R} \sum_{r=1}^{R} \frac{\exp(V_{ij})}{\sum_{j=1}^{J} \exp(V_{ij})} \tag{2.15}$$

where R is the number of replications (draws). The simulation proceeds as follows. For a given value of the hyper-parameters Ω, a value for β_i is drawn from the assumed distribution. The conditional on β_i logit formula (simulated probability that an individual i chooses alternative j) is computed. The procedure is repeated several times. The higher the number of draws, the lower the bias in the maximum simulated likelihood estimator (Lee, 1992; Hensher and Greene, 2003). Specifically, the estimation procedure uses Halton draws since Train (1999) and Bhat (2001) provide evidence that they are better suited than purely random ones[65].

Its flexibility to accommodate taste heterogeneity and correlation between attributes makes the RPL model the most applied model for discrete choices. Moreover, this model does not exhibit the IIA property as the ratio of probabilities now depends on all the data, therefore including the attributes of alternatives other than j and k.

RPL with correlated parameters

Recent research in discrete choice modelling has started to use correlated random parameters (Mariel and Meyerhoff, 2018; Wakamatsu et al., 2018). In spite of this, most empirical studies that estimate a RPL model impose the random parameters to be uncorrelated (i.e. they specify the random coefficients to have mean zero and a diagonal covariance matrix[66]). However, this restriction suffers from some limitations. First, from a theoretical perspective, the use of correlated RPL is better since it does not impose constraints on the model estimation. In this vein, McFadden and Train (2000) have

[65] Halton sequences are a type of quasi-random numbers generated by dividing a unitary interval into as many parts as a random prime number, and then dividing each obtained into the same number of segments and so on. After the sequence is generated, the numbers are introduced into a function of inverse distribution to be used during the simulations.

[66] Interestingly, the first application of this methodology by Revelt and Train (1998) is one of the exceptions.

stablished that the RPL specification allows the representation of any well-behaved RUM discrete choice model, but a necessary condition is that no constraints are imposed on the model.

Second, the uncorrelated RPL implies that the scale is constant. As introduced before, the scale refers to the magnitude of the random component relative to the deterministic one, and it is inversely related to the standard deviation of the Gumbel error term. Under a RUM framework, it is highly likely that the weight of the random component differs over people (scale heterogeneity). Although this is a relatively old problem (see earlier research by Louviere et al. (1999) and Swait and Louviere (1993)), scale heterogeneity has received growing attention in the recent years (see Greene and Hensher, 2010; Keane and Wasi, 2013)[67].

Fiebig et al. (2010) and Greene and Hensher (2010) proposed a new econometric model for the purpose of disentangling taste heterogeneity from scale heterogeneity (the Generalized Multinomial Logit Model). However, Hess and Rose (2012) and Hess and Train (2017) have recently warned that both sources of heterogeneity cannot be separately identified. According to these authors, the GMNL model is only appropriate if scale heterogeneity is the only source of correlation between the attributes. However, this might not be the case if preferences for an attribute are correlated with preferences for another. For instance, Hess et al. (2017) show how preferences for quality of services, safety and cost are correlated in the context of travellers' choices of route by car and public transport. These authors 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[68].

Due to these reasons, we estimate a RPL with correlated parameters. Under this formulation, the full covariance matrix of the random coefficients is a lower triangular matrix with nonzero off diagonal elements (denoted by Γ). These below diagonal elements are additional parameters to be estimated. For notational convenience, let us write the full vector of random parameters as follows:

$$\beta_i = b + \Delta\,\mathrm{Z_i} + \Sigma\, \nu_i \qquad (2.16)$$

where $Var\,(\beta_i |\, X_j, Z_i) = \Sigma = \Gamma\Gamma^t$

Scarpa et al. (2008) and Mariel and Meyerhoff (2018) show that the RPL with correlated parameters outperforms the uncorrelated one since it allows for complex substitution patterns.

[67] Under scale heterogeneity, the utility for individual i is more random (less deterministic) than for individual i'. Hence, all utility coefficients for individual i will be smaller in magnitude than the corresponding ones for individual i'. For an in-depth discussion of the roots of scale heterogeneity and its effect on parameter estimates see Hess et al. (2009).

[68] Nevertheless, the estimated correlations between the attributes cannot be directly interpreted as scale heterogeneity since it includes other behavioural phenomena. Mariel and Meyerhoff (2018) show how positive correlation between the marginal utilities of two attributes can lead to a negative estimated correlation through scale heterogeneity.

RPL with Error components (RPL-ECM)

As discussed in Section 2, the tourism literature has shown that unobserved destination-specific attributes such as destination image play a role in destination choice (e.g. Hong et al., 2006). To control for residual utility not captured in the attributes, empirical modelling usually includes a full set of Alternative-Specific Constants (hereafter ASC). In the RPL context, the joint specification of random parameter together with ASC has 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 whole sample. Following the example introduced before, it assumes that all visitors hold the same destination image of a destination j. Since they control for non-observable preference heterogeneity, one might think they should be allowed to be randomly distributed like 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 (Scarpa et al., 2005; 2007; Brownstone and Train, 1999) so that the utility function is given by:

$$U_{ij} = ASC_j + X_{kj}'\beta_{ik} + \vartheta_n E_n + \varepsilon_{ij} \tag{2.17}$$

where ASC_j are a set of alternative-specific constants (to be specified in the next section), E_n are random Error Components 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 Error Components (henceforth EC) 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). In this way, the random component of the utility is given by $\epsilon_{ij} = \vartheta_n E_n + \varepsilon_{ij}$.

The set of ASCs and the ECs are not separable identified when specified at the j-level. The 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 js. Hence, the introduction of ECs in the model makes utility to be correlated over alternatives because $Cov(E_n, E_l) = E[(\vartheta_n E_n + \varepsilon_{ij})(\vartheta_l E_l + \varepsilon_{ij})]) = \vartheta_{kn}' W \vartheta_l$, where W is the covariance matrix of the ECs, which for the sake of parsimony is assumed to be diagonal.

The necessity of controlling for correlated unobserved heterogeneity across alternatives in recreational demand model has been recently highlighted by Bujosa (2014), who shows how heterogeneity in substitution patterns across alternatives significantly affects the marginal sensitivities for site attributes. Therefore, the RPLc-ECM model captures unobserved heterogeneity at the individual and the alternative level[69].

Several remarks are in order. First, the inclusion of the ECs mimics in certain way the structure of a Nested Logit model (NL) in that it allows for correlation in utilities of

[69] This model has also been referred to as the Normal Error Component Logit Mixture (NECLM) (Walker et al., 2007). A complete description of the model formulation, the log likelihood function and an application can be found in Greene and Hensher (2007).

destination belonging to the same component (nest), but with the additional advantage that it does not impose the IIA property, as the NL does across nests (see Train (2009, p. 139-140) for further details). Therefore, the RPL-ECM captures richer and more intuitive patterns of correlation than the NL (see Herriges and Phaneuf (2002) for a comparison). What is more, the RPL-ECM model is computationally more efficient than the analogous Generalized Extreme Value Mixture (Gopinath et al., 2005).

Second, the model is specified in the level form (that is, $U_{ij}, j = 1, \dots, 17$) rather than in the difference form (i.e. $U_{ij} - U_{ik}$ for $j, k = 1, \dots, 17$ and $\forall\ j \neq k$). This is because the level form is convenient for model identification and parameter interpretation, although the subsequent derivation of choice probabilities is based on utility differences. In this sense, to compile with order and rank conditions (see Walker et al., 2007), the ECs are assumed to be homoscedastic. Third, since the random error of each alternative in the choice set is composed of a normally distributed EC and an iid Extreme Value component, the systematic portion of the utility better represents the utility of some alternatives than others, this way further controlling for potential scale heterogeneity.

5.2. Model specification

Consistent with our theoretical model, our proposed empirical model has the following generic form:

$$U_{ij} = f(ASC_j, X_j) + \xi_j + \varepsilon_{ij} \tag{2.18}$$

where U_{ij} is the utility that individual i gets for travelling to destination j, ASC_j is a set of alternative-specific constants, ε_{ij} is the iid idiosyncratic error term, ξ_j is a set of ECs and X_j is the vector of destination-specific attributes introduced in Section 4.

We expect the number of kilometres for skiing (*SKI_KM*) and the presence of coast in each region (*COAST*) to be more valued by those whose main motivation is practising winter and aquatic sports, respectively. Therefore, the vector of regional attributes is expanded to include two interaction terms between these two attributes and these two motivations (*SKI_KM*winter_sports* and *COAST*aquatic*). Hence, the vector of attributes is given by:

$$X_j = (DIST_j, r\_TEMP_j, RAIN_j, TCPI_j, TOU\_SPOTS_j,\ NAT\_PARKS_j,$$
$$SIZE\_NAT_j, SKI\_KM_j, SKI\_KM_j * winter\_sport_i, COAST_j, COAST_j * aquatic_i)$$

In the main analysis, we use *TCPI* at each destination j as our proxy for prices instead of *r_TCPI*. This is because it facilitates the interpretation of the own- and cross-price elasticities to be derived later. Furthermore, for the case of domestic tourism, the use of relative price indexes might be less relevant since inflation differences between regions are of lower magnitude than for international destinations. Nonetheless, we repeat the benchmark analysis replacing *TCPI* by *r_TCPI* (see subsection 6.1).

Although our model specification includes those attributes that are supposed to be the determinants of destination choice, there are several destination-specific features we are not accounting for. To control for unobservable regional differences, a set of Alternative Specific Constants (ASC) are included in the empirical model. For the sake of parsimony, instead of defining them at the NUTS 2 level (17 Spanish Autonomous Communities) we opt for defining them at the NUTS1 level (7 regions)[70]. This practise of defining group specific constants is common in the recreational demand literature (Parsons and Hauber, 1998; Hynes et al., 2008). This is because the inclusion of a full set of ASCs along with time-invariant destination attributes (in our case *TOU_SPOTS*, *NAT_PARKS*, *SIZE_NAT*, *COAST*) produces an identification problem. By using the NUTS 1 regional aggregation, destinations are grouped based on geographical proximity. This follows the practise of Bekhor and Prashker (2008). Note that size effects are captured in this set of ASCs.

As for the ECs, we define four components. The first EC relates to regions located in the North of Spain. The second EC refers to regions in the centre without sea. The third EC gathers regions in the South-East part of the country (Mediterranean regions). The fourth EC is defined for the Canary Islands. Table 2.6 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 2.6.- ASC and ECs

Accordingly, the model to be estimated is:

$$\begin{aligned} U_{ij} = {} & \alpha_1\, REG1_j + \alpha_2\, REG2_j + \alpha_3\, REG3_j + \alpha_4\, REG4_j + \alpha_5\, REG5_j + \alpha_6\, REG6_j \\ & + \beta_{1i}\, DIST_{ij} + \beta_{2i}\, r\_TEMP_{ij} + \beta_3\, RAIN_j + \beta_4\, TCPI_j \\ & + \beta_5\, TOU\_SPOTS_j + \beta_6\, NAT\_PARKS_j + \beta_7\, SIZE\_NAT_j + \beta_8\, SKI\_KM_j \\ & + \beta_9\, SKI\_KM_j * winter\_sports_i + \beta_{10}\, COAST_j + \beta_{11}\, COAST_j * aquatic_i \\ & + \vartheta_1 E1_{ij} + \vartheta_2\, E2_{ij} + \vartheta_3\, E3_{ij} + \vartheta_4\, E4_{ij} + \varepsilon_{ij} \end{aligned} \tag{2.19}$$

[70] In doing so, we assume that NUTS2 regions that belong to the same NUTS1 regional aggregation share common unobservable factors. This grouping has the advantage that it only requires 6 parameters to be estimated rather than 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. In doing so we assume that both positive and negative values for these two parameters exist in the population. The rest (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 is unstable when many coefficients are allowed to vary, especially when working with cross-sectional data (Revelt and Train, 1998)[71].

There are several reasons for treating the price coefficient as fixed instead of randomly distributed. First, the election of the distribution for the cost attribute is not an easy task. On the one hand, since the cost parameter is necessarily negative, a normal distribution is not appropriate. To overcome this, most scholars opt for defining it to be log-normally distributed (with negative sign) to ensure that the price coefficient is always negative. However, in such a case Greene (2012, p.126) warns that mixing normal and lognormal distributions in an unrestricted covariance matrix would lead to a correlation parameter that would be difficult to rely on. Second, our cost variable is a price index that proxies the cost of tourism services. Since it is not a monetary measure, it would make little sense to specify the price index as a random parameter.

As discussed in Sections 2 and 3, apart from the stochastic component, we allow tourists' preferences for distance and temperatures to depend on a set of choice invariant observed characteristics. These mean shifters are grouped into six blocks: i) age, ii) income, iii) party size, iv) time effects, v) climate conditions at the origin, and vi) trip purposes (goals).

- Age: the marginal utility for temperature and distance might vary according to the 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 income, the lower the dissuasive effect of distance. Hence, high-income earners are expected to be willing to travel farther away (Nicolau, 2010a; Mathews et al., 2018). We include two dummy variables for medium and high household income (denoted by *inc2* and *inc3*), leaving low income (*inc1*) as the reference category[72].
- Party size: as argued in Section 2, trip party size is expected to exert a significant effect on the disutility of distance. The variable *parsize* is defined as the number of individuals that participate in the trip.
- Time effects: preferences for climate and distance might depend on time constraints. For instance, the disutility of covering long distances might require a minimum length of stay at the destination in exchange. First, we consider a dummy variable for whether the trip is a weekend one (Saturday and Sunday),

[71] 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 is allowed to be correlated.

[72] *inc1* takes value one if household income is below 1,500 euros per month. *inc2* takes value 1 if household income is between 1,500 and euros per month. *inc3* takes value 1 for households whose income is higher than 3,500 euros per month.

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 quarters (*q1*, *q2*, and *q4,* respectively). The third quarter acts as the reference category.

- Climate conditions at the origin: although in the utility function we consider the temperature at each destination relative to that at 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 in the following manner:

$$d\_warmorigin\_q_{jt} = \begin{cases} 1 \quad \text{if } \mathrm{T}_{jt} > \sum_{j=1}^{J}\sum_{t=1}^{T}\frac{T_{jt}}{J*T} \; \forall \, t \in q, \qquad for\, q = 1,2,3,4 \\ \qquad\qquad 0 \quad \text{otherwise} \end{cases} \tag{2.20}$$

These four dummy variables capture the effect of origins that are warmer than the average at each quarter[73].

- Trip purposes: consistent with the literature, the marginal (dis)utilities for distance and climate conditions are assumed to vary depending on trip purposes. We specifically consider the following ones: the practice of winter sports (*winter_sports*), mountaineering, trekking or visiting natural areas (*mou_trek_nat*)[74], 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:

$$\begin{aligned}\beta_{1i} = {} & b_1 + \delta_1\, age_i + \delta_2\, inc2_i + \delta_3\, inc3_i + \delta_4\, parsize_i + \delta_5\, weekend_i + \delta_6\, Q1_i + \\ & \delta_7\, Q2_i + \delta_8\, Q4_i + \delta_9\, d\_warmorigin\_1_i + \delta_{10}\, d\_warmorigin\_2_i + \\ & \delta_{11}\, d\_warmorigin\_3_i + \delta_{12}\, d\_warmorigin\_4_i + \delta_{13}\, winter\_sports_i + \\ & \delta_{14}\, mou\_trek\_nat_i + \delta_{15}\, rural_i + \delta_{16}\, aquatic_i + \delta_{17}\, advent_i + \sigma_1 v_{i1}\end{aligned} \tag{2.21}$$

$$\begin{aligned}\beta_{2i} = {} & b_2 + \theta_1\, age_i + \theta_2\, inc2_i + \theta_3\, inc3_i + \theta_4\, parsize_i + \theta_5\, weekend_i + \theta_6\, Q1_i + \\ & \theta_7\, Q2_i + \theta_8\, Q4_i + \theta_9\, d\_warmorigin\_1_i + \theta_{10}\, d\_warmorigin\_2_i + \\ & \theta_{11}\, d\_warmorigin\_3_i + \theta_{12}\, d\_warmorigin\_4_i + \theta_{13}\, winter\_sports_i + \\ & \theta_{14}\, mou\_trek\_nat_i + \theta_{15}\, rural_i + \theta_{16}\, aquatic_i + \theta_{17}\, advent_i + \sigma_2 v_{i2}\end{aligned} \tag{2.22}$$

In principle, 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

[73] The cut-off points (mean temperature 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. We refer the reader to Annex 1 Figure A7 in the Supplementary Material for a smooth kernel density plot for temperature per quarter.

[74] We have combined the dummy variable for trekking and mountaineering (*mou_trek*) and the one for visiting natural areas (*nat_areas*) since they are similar.

for *r_TEMP*. However, to avoid imposing such restrictions, in our benchmark analysis we use the same vector of moderators (Z_i) for the two random parameters.

6. RESULTS

Table 2.7 reports the estimates for a baseline MNL model, the RPL model with correlated parameters (RPLc) and the correlated RPL with Error Components (RPLc-ECM). The three models have been estimated in NLOGIT 5 (ChoiceMetrics, 2012). For the RPLc and the RPLc-ECM models, we used Halton draws with 1,000 replications[75]. The standard errors have been computed using the delta method. To reduce the scale of X_j, *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. Due to its important shortcomings, the MNL model is only presented for comparison purposes. Hence, we focus on the results from the RPLc and the RPLc-ECM models. To discriminate between them, we rely on model fit criteria. The McFadden pseudo ρ^2 (McFadden, 1974) is the most used measure to assess the goodness of fit in discrete choice modelling[76]. Compared with the baseline MNL model, there is a substantial improvement in the pseudo ρ^2 for the RPLc and RPLc-ECM models (0.26 vs 0.40). Therefore, it seems that allowing for preference heterogeneity and unrestricted scale (free correlation) between the attributes leads to a better model fit. Both the log-likelihood and the Akaike Information Criterion (AIC) indicate that the RPLc-ECM model provides a slightly better fit than the RPLc model. Accordingly, the subsequent interpretation of results is based on this model specification.

Distance exerts, on average, a negative effect on utility, in line with the findings of Lyons et al. (2009), Van Nostrand et al. (2013), Chandra et al. (2014), De Valck et al. (2017) and Gosens and Rouwendal (2018). The spread parameter (standard deviation of the random component) is statistically significant. Since the mean effect of distance is almost centered at zero, this implies that preferences for distance are distributed to both sides of zero. While for some people distance is a dissuasive factor, for others it is a desirable feature. These results are like the ones by Nicolau (2010a), who also find a high degree of heterogeneity in Spanish tourists' response to distance.

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 corroborates findings by Bigano et al. (2006) and Lyons et al. (2009), who note that warmer destinations are more likely to be chosen. However,

---

[75] Using a large number of draws as we do does not only reduce simulation error making parameter estimates more precise but also avoids identification problems (Chiou and Walker, 2007). For the maximum simulated estimator to be consistent, efficient and asymptotically equivalent to the maximum likelihood estimator, it is required that the number of draws increase at a faster rate than the square root of the number of observations (Train, 1998).

[76] It is a likelihood ratio test that fluctuates between zero and one and is computed as follows:
$\rho^2 = 1 - (L_j - K_j)/L(0)$, where L_j is the log-likelihood at convergence, L(0) is the log-likelihood for constants only and K is the number of parameters to be estimated. Louviere et al. (2000) indicate that values between 0.2 and 0.4 constitute a good model fit.

both the statistical significance and the magnitude of the standard deviation of the random component of *r_TEMP* suggest that the marginal utility of this variable is not homogeneous across the sample (to be developed below).

Regarding the rest of attributes, the higher the prices at destination, the less preferred the destination is. This is consistent with Bujosa et al. (2015) and Van Nostrand et al. (2013). The dummy variable for rainfall (*RAIN*) is negative and significant, in line with our expectations and the results by Maddison (2001) and Lyons et al. (2009). This implies that destinations with average rainfall over 60 litres per square kilometre are negatively valued. The number of tourism spots (*TOU_SPOTS*) is positively related with the likelihood of tourists travelling to that region. Additionally, both the number of national and natural parks (*NAT_PARKS*) and the surface of protected natural areas (*SIZE_NAT*) positively affect a destination being chosen. The positive coefficient of the *SKI_KM* variable indicates that, on average, nature-based tourists value the availability of kilometres for skiing. When we look at the interaction term with *winter_sports*, we see that this positive utility is higher for this type of tourists. Interestingly, regions with coast (*COAST*) are negatively valued, on average, *ceteris paribus*. However, the marginal utility for those who seek to practise aquatic sports (*COAST*aquatic*) becomes positive and significant. Therefore, the presence of coast is only relevant for this segment.

All ASCs are positive and 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 regions have some residual features that increase their attractiveness. In this sense, it appears that *REG1* (Galicia, Asturias and Cantabria) is the preferred area, followed by *REG4* (Castile and Leon, Castilla-LaMancha and Extremadura). On the other hand, the standard deviations of the ECs are statistically significant for the Centre block of regions (*EC2*) and for the Canary Islands (*EC4*). This implies that, after controlling for unobserved common factors through the ASCs, there remains other sources of heterogeneity between macroregions. Overall, it seems that unobserved features for the North and Mediterranean area are better captured by the ASCs, while the corresponding ones for the Centre and the Canary Islands exhibit larger variation and need to be controlled for by the ECs[77].

As introduced before, the statistical significance of the standard deviation of the random parameters for *DIST* and *r_TEMP* indicates that the marginal utility for temperature differentials and distance is heterogeneous across the sample. Apart from the usual t-test, we have additionally tested this using the Lagrange Multiplier test proposed by McFadden and Train (2000) (see Annex A3 in the Supplementary Material for more details). We now proceed to comment on the taste shifters for the two random coefficients.

[77] Nonetheless, *REG4* exhibits its own mean effect apart from the one from *EC2*.

	MNL		RPLc		RPLc-ECM	
Variable	Coef.	SE	Coef.	SE	Coef.	SE
$REG1$	–5.042***	0.2176	3.019***	0.4003	3.431***	0.4167
$REG2$	–6.236***	0.2497	1.912***	0.4189	2.118***	0.4340
$REG3$	–6.674***	0.2639	1.046**	0.4211	–0.236	0.5237
$REG4$	–6.561***	0.2464	1.535***	0.4141	1.773***	0.4294
$REG5$	–6.335***	0.2183	1.177***	0.3835	1.480***	0.3913
$REG6$	–7.067***	0.2512	0.319	0.4151	0.594	0.4195
$DIST$	–0.661***	0.0090	–0.496***	0.0560	–0.519***	0.0598
r_TEMP	0.347***	0.1277	1.818***	0.7045	2.189***	0.7573
$RAIN$	–0.182***	0.0536	–0.262***	0.0652	–0.328***	0.0725
$TCPI$	–0.005	0.0049	–0.012**	0.0058	–0.012**	0.0061
TOU_SPOTS	0.106***	0.0081	0.114***	0.0093	0.122***	0.0095
NAT_PARKS	0.047***	0.0069	0.052***	0.0085	0.054***	0.0089
$SIZE\_NAT$	0.001*	0.0007	0.002***	0.0009	0.002**	0.0009
SKI_KM	0.001***	0.0001	0.001***	0.0002	0.001***	0.0002
$SKI\_KM * winter\_sports$	0.007***	0.0003	0.009***	0.0005	0.009***	0.0005
$COAST$	–1.280***	0.1514	–1.144***	0.1648	–1.336***	0.1870
$COAST * aquatic$	2.530***	0.1097	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	–12,575.9		–11,194.7		–11,170.0	
AIC	25,186.0		22,497.6		22,456.0	
Pseudo-R2	0.269		0.406		0.408	
N	6,661		6,661		6,661	

Table 7.- Parameter estimates for MNL, RPLc and RPLc-ECM models
*** p<0.01, ** p<0.05, * p<0.1

The negative marginal utility of distance is moderated by age (i.e. elderly people are less deterred by travelling longer distances). This contradicts Lyons et al. (2009), who find that older people are averse to travelling farther away. This finding can be explained by elderly people having lower opportunity costs of time. Another explanation follows Van Nostrand et al. (2013), who show that the willingness to cover longer distances is related to the length of the stay. These authors show that retired people tend to stay for longer at a vacation destination so that they are less deterred by distance[78]. Similarly, income also appears to moderate the disutility of distance. This result is in line with Nicolau (2008). A possible explanation is that high-income people can afford high-speed modes of transport, so the same distance imposes lower costs to them (Van Nostrand et al., 2013). Conversely, travel party size increases the marginal disutility of distance. This might be explained by the fact that larger travel groups impose higher transportation costs. Similarly, weekend trips make tourists to be more deterred by travelling long distances. Since this type of trips is of a short duration, it makes sense that individuals prefer nearby destinations under 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 of distance in the second and third quarter. This suggests that when the place of origin is relatively warmer than the average, distance turns to a higher impediment.

Regarding the role of trip purposes, the estimates show that those who seek mountaineering, trekking and visiting natural areas are more discouraged to travel farther away. Nonetheless, the practise of aquatic or adventure sport moderates the disutility of distance, especially in the former case. This implies that these two trip purposes alleviate the distance decoy effect. This result is consistent with previous research that finds remarkable differences in sensitivity to distance depending on motivations (Nicolau, 2010a; Swait et al., 2020).

[78] We have tested including the square of age. However, this term is never significant, suggesting that the moderating effect of age is linear.

The marginal utility of *r_TEMP* is not related with age, income or party size. Conversely, those who travel in a 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). This suggests that, *ceteris paribus*, the pursuit of a warmer destination has less relevance in autumn and winter, although still warm regions are preferred. Interestingly, those 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 given threshold (Maddison, 2001; Bigano et al., 2006; Lyons et al., 2009). Contrariwise, 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, being this preference higher in the second and third quarters, and ii) mild climates are preferred in the first and fourth quarters. To facilitate the interpretation, Table 2.8 presents how the marginal utility of *r_TEMP* differs by whether the origin exhibits below-mean or above-mean temperatures. Based on equations (2.19) and (2.22), the figures for the first quarter are obtained as follows:

$$\frac{\partial\, U_{ijt}}{\partial\, r\_TEMP_{ijt}} \frac{\partial\, r\_TEMP_{ijt}}{\partial\, q1_{it}} = \begin{cases} b_2 + \theta_6 & if\ d\_warmorigin\_1 = 0 \\ b_2 + \theta_6 + \theta_9 & if\ d\_warmorigin\_1 = 1 \end{cases} \tag{2.23}$$

The marginal utilities for the remaining quarters are computed in the same fashion.

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

Table 2.8.- Estimated marginal utilities of *r_TEMP* per quarter

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

Regarding the role of goals, 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 less relevant. Conversely, those who seek to practise adventure/risky sports (*advent*) show a higher preference for warmer locations. Contrary to our expectations, 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 2.9[79]. The terms 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 2.9.- Variance-covariance matrix estimates for RPLc and RPLc-ECM models

As shown, the covariance between the random parameters is not statistically significant. This indicates that the two marginal utilities are independent. However, note that these estimates are conditional on the vector of taste shifters Z_i. We have re-estimated the RPLc-ECM model without them (i.e. allowing the random parameters to be merely random apart from a mean effect). Results are shown in Annex 4 Table A4. In this case, the covariance between the random parameters becomes significant (and positive), and the correlation amounts to 0.32. Accordingly, it seems that the marginal utilities of *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 observable sources of preference heterogeneity is able to capture the shared correlation between the marginal utilities of *DIST* and *r_TEMP*.

6.1. Robustness checks

Several alternative model specifications were examined. First, we tested whether our results change depending on the distribution assumed for the random parameters. Although the normal distribution is the most used, the choice of the parameter distribution is an empirical issue that remains to be investigated (Daly et al., 2012). Hence, a triangular, a uniform, a truncated normal and a Weibull distribution were tested as alternatives. We refer the reader to Annex 4 Tables A5 and A6 in the Supplementary Material for the estimates. Our results are not driven by the distribution of unobserved heterogeneity.

Second, we replaced *TCPI* by *r_TCPI*. This was done to check whether the results are sensitive to controlling for prices at destination or for relative prices. The estimates are reported in Annex 4 Table A7. The results are consistent with economic theory (utility is negatively related with destinations that are relatively more expensive than the origin) and remain largely unchanged. Similarly, we replaced our measure of distance that considers tourists' place of origin probabilistically by the Euclidean distance between NUTS 2 centroids (*DIST_alt*), which therefore assumes that the distance between provinces belonging to the same Autonomous Communities are all zero. Results are presented in Annex 4 Table A8.

[79] Even under the normality assumption, when Γ is not diagonal, the elements in their main diagonal are not exactly the standard deviations of the random parameters. Hence, the estimated variance-covariance matrix needs to be recovered from the estimates using Cholesky decomposition. See Greene (2012) and Hensher et al. (2015) for further details.

Third, both the RPLc and the RPLc-ECM models were estimated imposing the restriction that the parameters for the mean shifters that are non-significant in Table 7 are zero. Results are displayed in Annex 4 Table A9. The magnitude and the direction of the effects are not affected by imposing such restriction.

Finally, 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 from the analysis, resulting in 6,202 individuals in the retained subsample. Parameter estimates are reported in Annex 4 Table A10. Results remain consistent with the main analysis, showing that our findings are not affected by the inclusion of these islands.

6.2. Individual-specific marginal utilities

One of the appealing features of the RPL model is the possibility of recovering individual-specific estimates of the marginal utilities (MUs) for the attributes. The estimates of the structural parameters in our Random Utility model might provide an incomplete picture of the marginal utilities. The *unconditional* mean of β_{ki} (only conditioning on Z) is simply:

$$E[\beta_{ki}|Z_i] = b_k + \delta_k{}'Z_i \tag{2.24}$$

According to Greene (2004), the *unconditional* mean estimator in (2.24) 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 ECs 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:

$$h(\beta|y_j,\Omega) = \frac{P_{ij}(y_j|\beta)f(\beta|\Omega)}{P_{ij}(y_j|\Omega)} = \frac{\frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})} f(\beta_i|\Omega)}{\int \frac{\exp(V_{ij})}{\sum_{j=1}^{J}\exp(V_{ij})} f(\beta_i|\Omega)\, d\beta_i} \tag{2.25}$$

Since the denominator in (2.25) 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 (2.25) 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 (i.e. $P_{ij}(y_j|\beta)$) needs to firstly eliminate the ECs from the expression by integrating over its standard normal distribution.

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

$$E(\beta_i|y_{ij}, X_{kj}, Z_i) = \frac{\int_{\beta_i}\int_{E_i} \beta_i P_{ij}(y_j|\beta, X_{kj}, E_i)\, f(\beta_{i,} E_i\,|Z_i) dE_i\, d\,\beta_i}{\int_{\beta_i}\int_{E_i} P_{ij}(y_j|\beta, X_{kj}, E_i)\, f(\beta_{i,} E_i\,|Z_i) dE_i\, d\,\beta_i} \qquad (2.26)$$

where $f(\beta_{i,} E_i\,|Z_i)$ 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[80]. 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[81]. Since the integral in (2.26) does not have a closed form solution, the conditional means of the parameters are approximated by Simulated Maximum Likelihood[82].

Figures 2.3 and 2.4 show a kernel plot of the estimated individual-specific parameter estimates for *DIST* and *r_TEMP*[83].

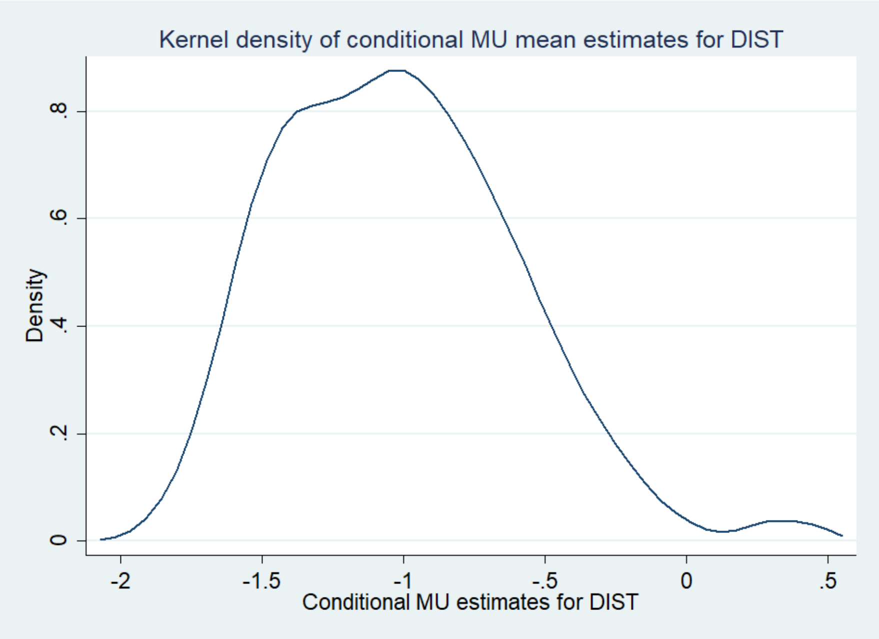


Figure 2.3.- Kernel density of conditional MU mean estimates for *DIST*

[80] Greene (2012) warns that these estimates are conditioned on the observable information for individual i. Put another way, the $\hat{\beta}_i$ 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.

[81] 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).

[82] Alternatively, the estimation could be performed under a Bayesian 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.

[83] 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.

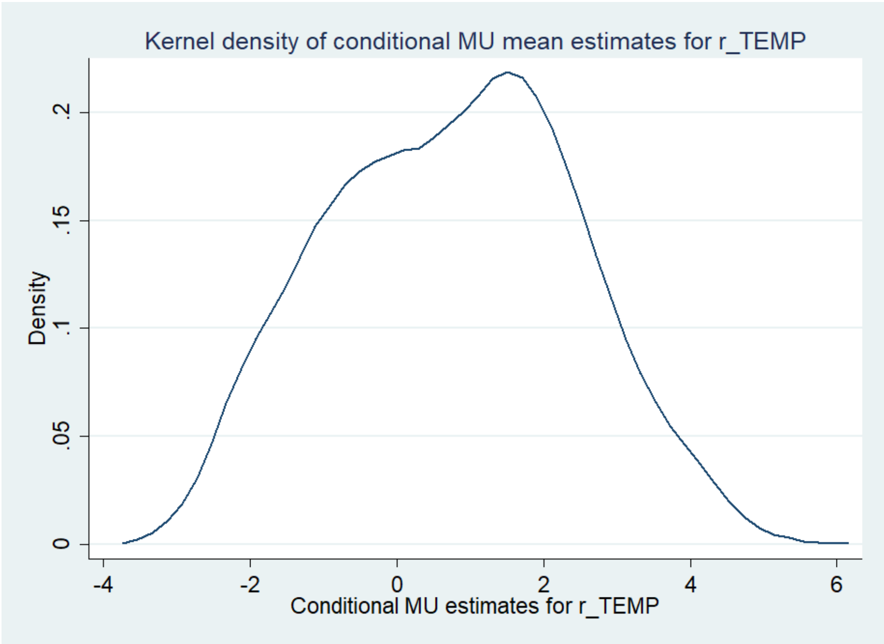


Figure 2.4.- Kernel density of conditional MU mean estimates for *r_TEMP*

As it is evident from Figures 2.3 and 2.4, the conditional mean estimates of the marginal utilities of *DIST* and *r_TEMP* are quite heterogeneous. Regarding distance, although a small share attaches positive utility, in general it can be regarded as a dissuasive factor. 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; Scarpa et al., 2013).

A recommended diagnosis test on the model specification and estimation consists on comparing the sample average of the conditional distribution (i.e. $\sum_{i=1}^{N} \widehat{\beta_{ik}}|y_{ij}, X_{kj}, Z_i$) with the sample average of the *unconditional* one using point estimates $\widehat{b_k}$ and $\widehat{\delta_k}$ (i.e. $\sum_{i=1}^{N} \widehat{\beta_{ik}}|Z_i$) (Train, 2009). Table 2.10 gives the mean values (and standard deviations) of the two estimated distributions (equations (2.24) and (2.26), respectively):

	$\widehat{\beta_{ik}}\|y_{ij}, X_{kj}, Z_i$		$\widehat{\beta_{ik}}\|Z_i$	
	Mean	SD	Mean	SD
DIST	-0.9525	0.4742	-1.0195	0.3551
r_TEMP	0.7054	1.6420	0.7338	1.6393

Table 2.10.- Mean and SD of conditional and 'unconditional' MUs for *DIST* and *r_TEMP*

The two-sample means are similar, which provides some reliability on the proper specification of our model. On this issue, Train (2009) argues that in correctly specified models the mean conditional distribution of tastes tends to equal the population distribution of tastes.

In sum, the derivation of the means of the conditional MUs illustrate how looking only at the estimate of the mean of the underlying distribution (b_1 and b_2 in equations (2.20) and (2.21)) might be not informative of preferences in the presence of taste heterogeneity. At it will become evident in subsection 6.4, the ratio of the marginal utilities for two attributes is not the same as the ratio of the mean marginal utilities for those attributes. In this way, by retrieving the most likely position of each individual in the distribution of sensitivities, we are able to identify extreme preferences that can be considered in subsequent analysis.

6.3. Attribute non-attendance

For modelling destination choice, we have assumed that all individuals consider all the attributes included in our model. However, it might happen that for some tourists distance and/or the ratio of temperatures are not relevant (i.e. they attach a zero value to these attributes in their utility function). As shown in Figures 2.3 and 2.4, when we allow for taste heterogeneity, some individuals have estimated values of the marginal utilities in the neighbourhood of zero. This implies that for these tourists those attributes have not been *attended* in their decision-making process, an issue normally referred to as *attribute non-attendance* (ANA)[84]. There is a large body of research on ANA in the discrete choice literature, especially in the context of choice experiments (Scarpa et al., 2010; Scarpa et al., 2013; Hess et al., 2013; Thiene et al., 2019).

Hess and Hensher (2010) propose a way to infer ANA from observed choice behaviour. These authors compute the coefficient of variation (CV) of the individual-specific conditional mean and standard deviation estimates for each attribute ($CV_{ki} = \frac{\sigma_{ki}}{\beta_{ki}}$). This coefficient of variation is interpreted "as the noise-to-signal ratio on the variability of taste intensity for attribute k as displayed by respondent i choice behaviour" (Scarpa et al., 2013, p. 170). When this noise-to-signal ratio is higher than a given threshold, the individual-specific preference distribution is said to be *over-dispersed,* so that the choice is consistent with that individual not having paid attention to that attribute k. Hess and Hensher arbitrarily choose 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). Individuals with estimated marginal utilities close to zero will have large values of CV_{ki}.

We follow Hess and Hensher and compute the coefficient of variation for each individual for both *DIST* and *r_TEMP* based on the estimates of the conditional distribution of the parameter estimates obtained from the RPLc-ECM model (Figures 2.3 and 2.4). We refer the reader to Greene (2004; 2012, pp. 147-148) and Hess (2010, p.136) for the derivation of the estimator of the conditional variance.

Although only 0.7% of the sample 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 share of individuals that do not *attend* the ratio of temperatures is larger than for the case of

[84] Whether or not they have *truly* attended or not the attribute cannot be addressed strictly. Rather, we explore whether individuals have *a very low sensitivity* to that attribute (i.e. a virtually zero marginal utility).

distance. This is consistent with Figures 2.3 and 2.4, where the percentage of tourists in the neighbourhood of zero is higher for *r_TEMP* than for *DIST*.

6.4. The marginal rate of substitution between distance and temperatures

In general, there is a natural positive relationship between distance and the ratio of temperatures: regions with a different climate relative to the place of residence (either warmer or cooler) are located farther away. Since we have shown that, on average, tourists attach positive utility to warmer destinations and negative utility to distant regions, they seem to face a trade-off between the two. To examine this, we compute the ratio between the conditional mean estimates, which can be understood as the marginal rate of substitution (MRS)[85]. RUM theory assumes compensatory behaviour, so the ratio of the partial derivatives of the latent utility with respect to *r_TEMP* and *DIST* measures how individuals are willing to trade distance in exchange for a temperature gain.

$$MRS_{r\_TEMP,DIST_{\,i}} = \frac{MU_{r\_TEMP}}{MU_{DIST}} = \frac{\frac{\partial\, U_{ij}{}^{*}}{\partial\, r\_TEMP_i}}{\frac{\partial\, U_{ij}{}^{*}}{\partial\, DIST_i}} = \frac{\beta_{2,i}}{\beta_{1,i}} \tag{2.27}$$

This marginal rate of substitution can be understood as a ‘willingness to pay’ if we consider distance as a payment vehicle (also referred to as ‘willingness to travel’). This is similar to Christie et al. (2007), De Valck et al. (2017), Whitehead and Wicker (2018) and Cronin et al. (2019). This ratio captures the distance (in hundreds of kilometres) individuals are willing to cover to gain a marginal increase (decrease) in relative temperatures. Since the ratio depends on the conditional mean estimates of the MUs for each attribute, the MRS is individual-specific. 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.

The distribution of the MRS is obtained from the distribution of the numerator and the denominator. In our case, the ratio of two normally distributed parameters has a discontinuous distribution with a singularity problem when the denominator takes value zero. To avoid this issue and following our previous discussion in subsection 6.3, for the calculation of the MRS we have omitted those who do not *attend* either of the two attributes (i.e. those for whom the coefficient of variation is higher than two in absolute terms)[86]. This leaves the sample with a total of 4,368 valid observations (65.5% of the original sample).

---

[85] Since the estimated marginal utilities are confounded by the scale factor (λ), the computation of the ratio removes the scale issue.

[86] This way to proceed is consistent with Daly et al. (2012). These authors provide mathematical proof that shows 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 exists. Their theorem applies to independent parameters. For the case of jointly normal variables, independence directly follows from a lack of correlation (i.e. $Cov(\beta_{1i}, \beta_{2i})$ =0). In our data, this is not an issue since, conditional on Z, the MUs in the population for distance and relative temperatures have been shown to be uncorrelated.

Figure 2.5 plots the histogram of the MRS, which is distributed to both sides of zero[87]. 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) 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 attend both attributes are willing to travel 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 of cooler climate conditions. Overall, the MRS shows how the willingness to travel longer distances increases as individuals attach higher importance to a different climate (either warmer or cooler).

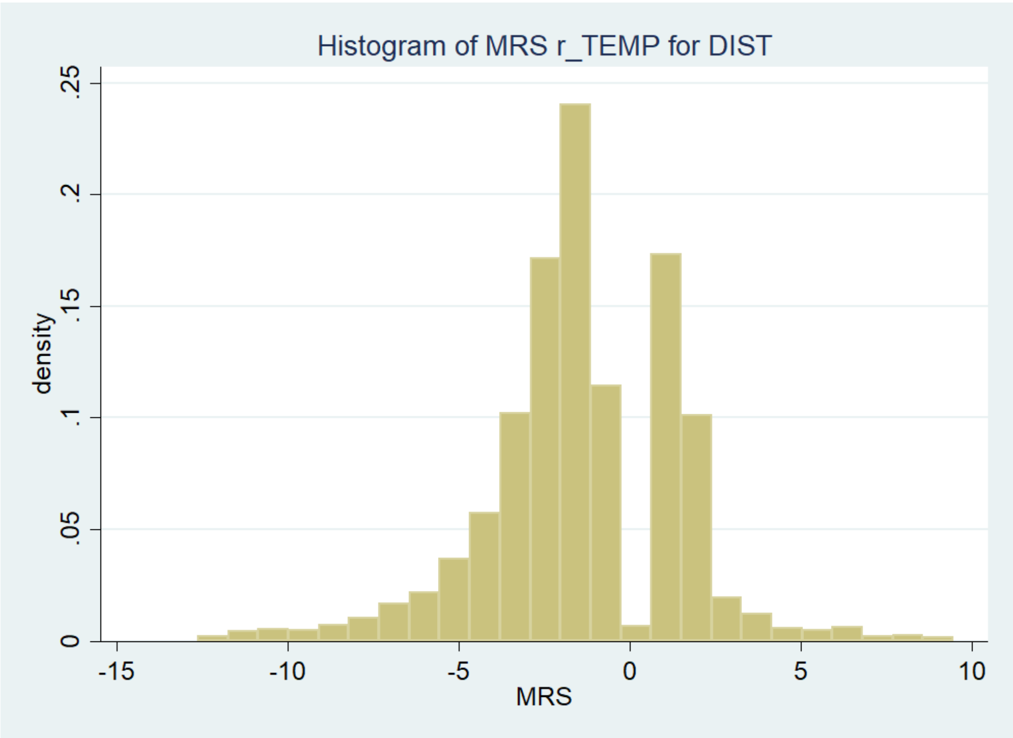


Figure 2.5.- MRS *r_TEMP* for *DIST*

6.5. Elasticities

Apart from considering taste heterogeneity, scale heterogeneity and cross-correlation between alternatives, another advantage of our modelling approach is that the RPLc-ECM model 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

[87] Since the obtained distribution has long tails, Figure 5 restricts the estimated MRS to lie on the interval (–12.87, 9.69), which gathers 99.7% of the data ($\mu \pm 3\sigma$).

evaluate tourists' reassignments under a shock in attribute k in a region j, *ceteris paribus*. More specifically, we aim to evaluate how changes in the price index for tourism services and 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:

$$\theta_{im(k|j)} = P_{im}(1(j = m) - P_{ij})\, \beta_{ik} \tag{2.28}$$

where P_{ij} and P_{im} are the choice probabilities as defined in Equation (2.13), 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:

$$\theta_{ij(k|j)} = P_{ij}\ (1 - P_{ij})\beta_{ik} \tag{2.29}$$

And the cross marginal effect is:

$$\theta_{im(k|j)} = P_{im}(-P_{ij})\, \beta_{ik} \tag{2.30}$$

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). The elasticities are obtained as follows:

$$\eta_{im\ (k|j)} \frac{\partial \ln P_{im}}{\partial \ln X\ (k|j)} = \frac{X_{k|j}}{P_{im}} \vartheta_{im(k|j)} = X_{k|j}(1(j = m) - P_{ij})\, \beta_{ik} \tag{2.31}$$

Tables 2.11 and 2.12 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 the corresponding attribute 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 2.6 and 2.7 depict the own elasticities with respect to *r_TEMP* and *TCPI*, respectively. The corresponding values can be found in the main diagonal in Tables 2.11 and 2.12. Starting with own *r_TEMP* elasticities (Figure 2.6), darker colours refer to higher values. All the regions exhibit positive own elasticities (i.e. a marginal increase in temperature relative to origin increases the likelihood of that region being chosen). Interestingly, regions in the North-West area (Cantabria, Galicia and Asturias), Madrid, Andalusia, La Rioja and Aragon have elasticities higher than the unity, suggesting that the effect of temperature increases is larger in these areas. Based on this, there is no clear spatial pattern in the own *r_TEMP* elasticities.

Turning to the cross *r_TEMP* elasticities, we pay attention to the substitution patterns between Northern and Southern regions. Interestingly, 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), Bujosa et al. (2015) 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 Annex 1 in the Supplementary Material) and different shares of outbound tourists (see Figure 1). 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. As can be seen in the third column in Table 2.5, the average values of *r_TEMP* vary by region so that percentage increases depend on the origin level.

Concerning the own *TCPI* elasticities (on-diagonal values in Table 2.12, depicted in Figure 2.7), 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. Darker colours in Figure 7 mean more elastic regions. The North-West regions are the ones that seem to be more sensitive to a price increase in tourism services, closely followed by the Valencian Community, Murcia and Castilla-La Mancha.

As for the cross *TCPI* elasticities (off-diagonal values in Table 2.12), 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 tends to be larger for neighbour 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 choice probabilities for the other region 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 only assume percentage increases in average temperatures in the region being analysed keeping everything constant. That is, both the temperatures 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.

	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 2.11.- *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 2.12.- *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.

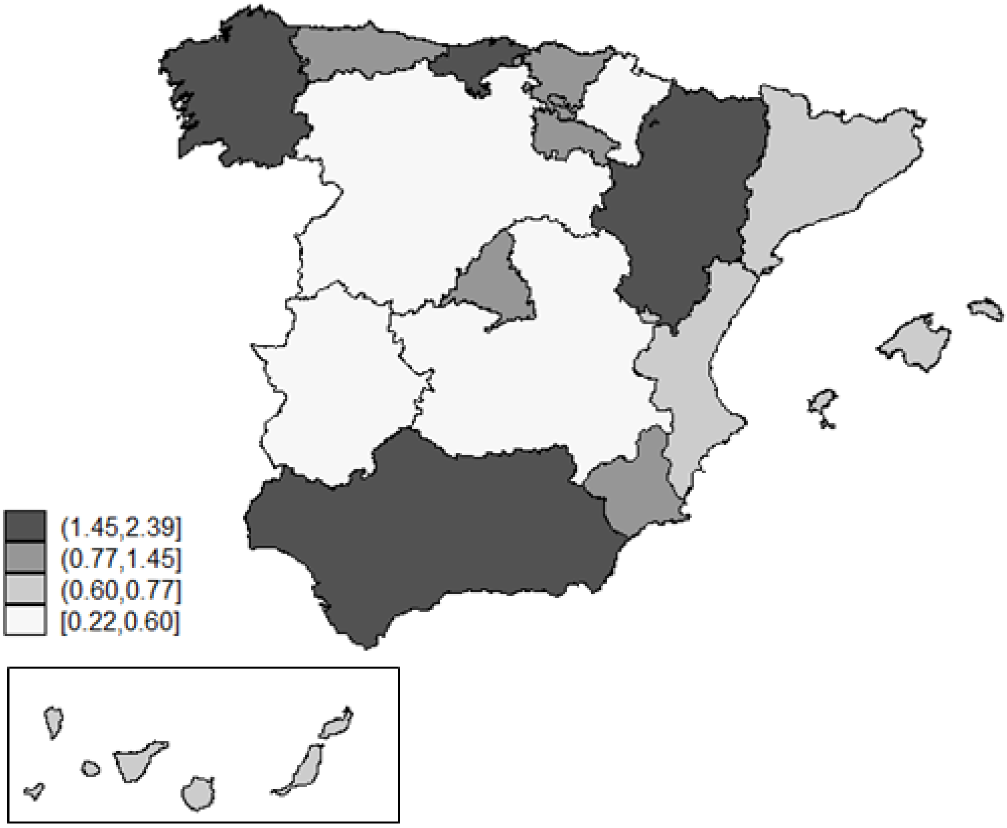


Figure 2.6.- Own choice-*r_TEMP* elasticities

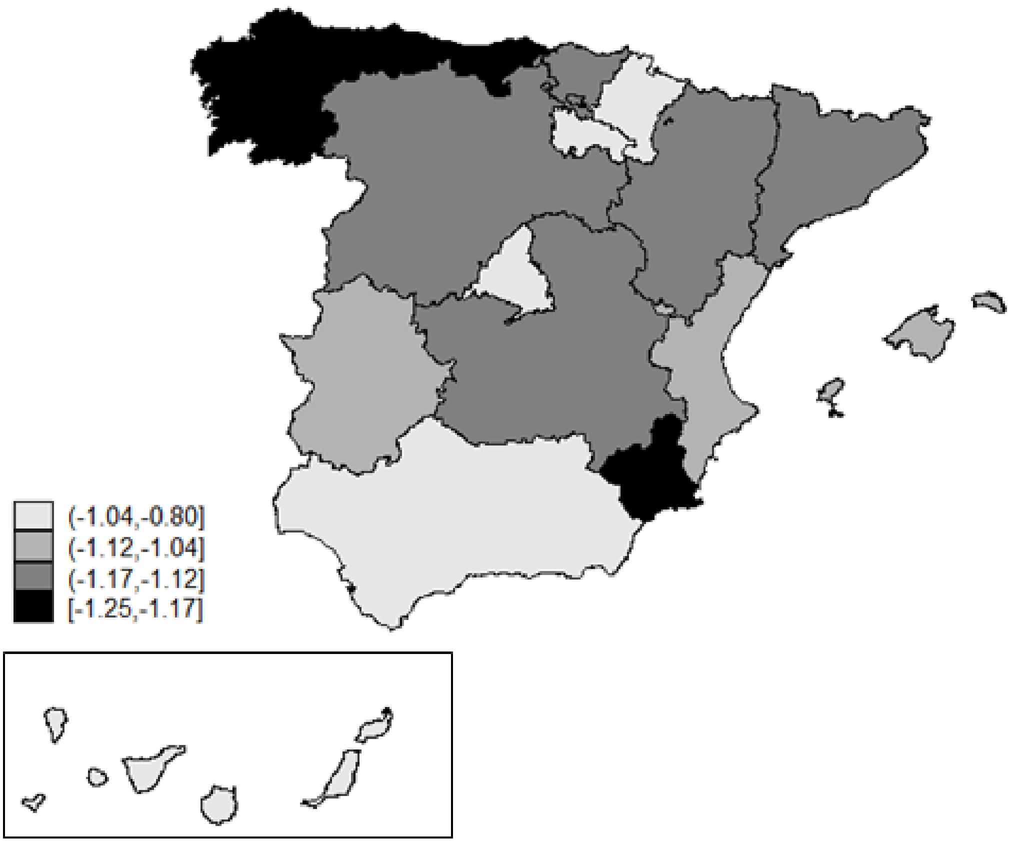


Figure 2.7.- Own choice-*TCPI* elasticities

7. CONCLUSIONS

In this paper, we have analysed the regional attributes that drive domestic tourist trips for nature-based purposes in Spain. We have combined monthly microdata with both time-varying and time-invariant regional characteristics, which has allowed us to control for exogenous sources of variation in destination attractiveness. Contrary to previous studies that analyse aggregate tourism flows, we have studied *individual* preferences for place-based attributes. We have paid attention to the effect of distance and relative temperatures in recreational site choice probabilities, while controlling for some other destination-specific amenities. In doing so, we have provided new evidence on the drivers of heterogeneity in marginal utilities for these two attributes.

We have estimated a correlated Random Parameter Logit with Error Components model that controls for unobservable heterogeneity in preferences for the attributes and for the destinations. The inclusion of Error Components in the model enables us to control for shared unobserved characteristics between similar destinations. Our parameter estimates are robust to the distribution assumed for the random parameters and to different specifications. A significant improvement in model fit is found when moving from a baseline Multinomial Logit model to the proposed correlated Random Parameter Logit with Error Components model.

Our results point to the existence of substantial preference heterogeneity for distance and temperature differentials between the origin and the destination. Whereas on average distance is as a dissuasive factor, this distaste for distant regions is moderated by age and income. Conversely, travel party size, travelling in weekends and in the fourth quarter reinforce the negative effect of distance. Interestingly, trips with the aim of practising aquatic or adventure sports make tourists more prone to travelling farther away, whereas trekking and mountaineering as the main trip purpose are associated with a higher distaste for distance.

As for the temperature at the destination relative to the origin, on average recreationists prefer warmer regions, especially in the third quarter. Remarkably, our estimates show that those from origins with above mean temperatures are largely deterred by 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, mountaineering and visiting rural areas. In addition, site choice probabilities are negatively related to high rainfall.

Our results also show 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 when choosing where to travel. Conversely, the probability of choosing a destination is negatively influenced by tourism prices. Interestingly, whereas inland regions are on average preferred for nature-based tourism, coastal destinations are best suited for those who like to practise aquatic sports.

Based on the model estimates, we have further shown that whereas only 0.7% of the sample attaches zero value to distance in their utilities, almost 34% of the sample appears to ignore the relative temperature of the destination with respect to that at the origin. For those who consider both attributes in their choices, we have found that, on average, they are willing to cover about 160 kilometres for a warmer climate. Interestingly, about 70% trade distance for temperature gains. However, the remaining 30% travel long distances to reach cooler locations. Furthermore, the own price elasticities indicate that all destinations are highly 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.

Our paper contributes to the literature on recreational demand and tourism economics in several ways. First, given the mixed evidence on the (dis)taste for distance and temperature in destination choice, we have proposed a model specification that allows the marginal utilities for these two dimensions to be heterogeneous. We have linked the distribution of preferences in the sample to a set of sociodemographic, temporal and trip-related variables. Therefore, we have not only modelled this taste heterogeneity but also identified the factors that shift sensitivities.

Second, contrary to most applications that only use temperatures at the destination, we have implemented a measure of relative temperatures that explicitly acknowledges that temperature rises may be differently valued depending on the level at the origin. Although this has been explored in studies concerned with explaining aggregate flows, this is the first application that considers relative temperatures for modelling individual choices. Moreover, among the set of taste shifters, we have considered indicators of quarterly above mean temperature at origin. Therefore, we provide some new evidence on the non-linearities in the preference for temperature.

Third, our econometric model deals with unobserved heterogeneity, allowing for shared correlation between the taste for temperatures and distance on the one hand, and shared correlation in unobservables between regions belonging to the same geographical area on the other. We have shown that the marginal utilities for distance and relative temperatures are unconditionally positively related. However, conditional on the taste shifters, the correlation vanishes. This suggest that modelling the sources of preference heterogeneity allows us to capture the shared correlation between the marginal utilities of these two factors. From the model estimates, we have derived the conditional estimates of the marginal utilities by conditioning on all available information for each individual. We have further computed the marginal rates of substitution of distance for temperature gains. To the best of our knowledge, this is the first study that offers some evidence on recreationists' substitution rates of distance for temperature.

Our results have some implications. Given the growing importance of the domestic market for the economic development of some regions, our findings could be valuable for public agencies in charge of the development and promotion of each destination. Our results might enhance their understanding of the factors that attract prospective tourists to their regions. Prompting nature-based tourism could also alleviate the intrinsic seasonality of tourism revenues. Each area has its own attractions and local public authorities need to highlight their appealing features based on what potential tourists are looking for when recreating. All Spanish regions choice probabilities are highly price elastic, which suggests that tourists are very sensitive to tourism price inflation. This is particularly true for Murcia, Asturias, Cantabria and Galicia, who exhibit 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, we believe our modelling approach offers an improved theoretically consistent way of analysing domestic leisure trips that can be extended to other types of recreational trips.

REFERENCES

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

Albalate, D. and Fageda, X. (2016). "High speed rail and tourism: empirical evidence from Spain". *Transportation Research Part A*, 85, 174-185.

Alegre, J. and Juaneda, C. (2006). "Destination loyalty: consumers' economic behavior". *Annals of Tourism Research*, 33(3), 684–706.

Alegre, J., Mateo, S. and Pou, L. (2010). "An analysis of households' appraisal of their budget constraints for potential participation in tourism". *Tourism Management*, 31, 45-56.

Amelung, B., Nicholls, S., and Viner, D. (2007). "Implications of global climate change for tourism flows and seasonality". *Journal of Travel Research*, 45, 285-296.

Ankomah, K.L., Crompton, J.L., and Baker, D. (1996). "Influence of cognitive distance in vacation choice". *Annals of Tourism Research*, 23(1), 138–150.

Arentze, T.A. and Timmermans, H.J.P. (2009). "A need-based model of multi-day, multi-person activity generation". *Transportation Research Part B*, 43, 251-265.

Athey, S. and Imbens, G.W. (2007). "Discrete choice models with multiple unobserved choice characteristics". *International Economic Review*, 48(4), 1159-1192.

Atzori, R., Fyall, A. and Miller, G. (2018). "Tourist responses to climate change: potential impacts and adaptation in Florida's coastal destinations". *Tourism Management*, 69, 12-22.

Balaguer, J. and Cantavella-Jordá, M. (2002). "Tourism as a long-run economic growth factor: the Spanish case". *Applied Economics*, 34(7), 877-884.

Basala, S.L. and Klenosky, D.B. (2001). "Travel-style preferences for visiting a novel destination: a conjoint investigation across the novelty-familiarity continuum". *Journal of Travel Research*, 40, 172-182.

Becken, S. (2012). "Measuring the effect of weather on tourism: a destination- and activity-based analysis". *Journal of Travel Research*, 52(2), 156-167.

Becken, S. and Wilson, J. (2013). "The impacts of weather on tourist travel". *Tourism Geographies*, 15(4), 620-639.

Beerli, A., Díaz-Meneses, G. and Moreno-Gil, S. (2007). "Self-congruity and destination choice". *Annals of Tourism Research*, 34(3), 571-587.

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.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete choice analysis: theory and application to travel demand*. The MIT Press, Cambridge.

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.

Bhat, C.R. (1995). "A heteroscedastic extreme value model of intercity travel mode choice". *Transportation Research Part B*, 29(6), 471–483.

Bhat, C.R. (2001). "Quasi-random maximum simulated likelihood estimation of the Mixed Multinomial Logit model". *Transportation Research Part B*, 35(7), 677-93.

Bhat, C.R., Astroza, S. and Bhat, A.C. (2016). "On allowing a general form for unobserved heterogeneity in the multiple discrete-continuous probit model: formulation and application to tourism travel". *Transportation Research Part B*, 86, 223-249.

Biagi, B. and Pulina, M. (2009). "Bivariate VAR models to test Granger causality between tourist demand and supply: Implications for regional sustainable growth". *Papers in Regional Science*, 88, 231–244.

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.

Boskin, M.J., Dulberger, E.R., Gordon, R.J., Griliches, Z. and Jorgenson, D.W. (1998). "Consumer prices, the Consumer Price Index, and the cost of living". *Journal of Economic Perspectives*, 12(1), 3-26.

Brownstone, D. and Train, K. (1999). "Forecasting new product penetration with flexible substitution patterns". *Journal of Econometrics*, 89, 109-129.

Bujosa, A. (2014). "Substitution patterns across alternatives as a source of preference heterogeneity in recreation demand models". *Journal of Environmental Management*, 144, 212-217.

Bujosa, A. and Riera, A. (2009). "Environmental diversity in recreational choice modelling". *Ecological Economics*, 68, 2743-2750.

Bujosa, A., Riera, A. and Torres, C.M. (2015). "Valuing tourism demand attributes to guide climate change adaptation measures efficiently: the case of the Spanish domestic travel market". *Tourism Management*, 47, 233-239.
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.
Campo-Martínez, S., Garau-Vadell, J.B. and Martínez-Ruiz, M.P. (2010). "Factors influencing repeat visits to a destination: the influence of group composition". *Tourism Management*, 31, 862-870.
Castelltort, M. and Mäder, G. (2010). "Press media coverage effects on destinations - A Monetary Public Value (MPV) analysis". *Tourism Management*, 31, 724-738.
Cao, J., Zhang, J., Wang, C., Hu, H. and Yu, P. (2020). "How far is the ideal destination? distance desire, ways to explore the antinomy of distance effects in tourist destination choice". *Journal of Travel Research*, 59(4), 614-630.
Capacci, S., Scorcu, A.E. and Vici, L. (2015). "Seaside tourism and eco-labels: the economic impact of blue flags". *Tourism Management*, 47, 88-96.
Chandra, A., Head, K. and Tappa, M. (2014). "The economics of cross-border travel". *The Review of Economics and Statistics*, 96(4), 648-661.
Chen, C., Lin, Y. and Petrick, J.F. (2013). "Social biases of destination perceptions". *Journal of Travel Research*, 52(2), 240-252.
Chiou, L. and Walker, J.L. (2007). "Masking identification of discrete choice models under simulation methods". *Journal of Econometrics*, 141(2), 683-703.
Christie, M., Hanley, N. and Hynes, S. (2007). "Valuing enhancements to forest recreation using choice experiment and contingent behaviour methods". *Journal of Forest Economics*, 13 (2–3), 75–102.
Correia, A., Santos, C. M., and Pestana-Barros, C. (2007). "Tourism in latin America: a choice analysis". *Annals of Tourism Research*, 34(3), 610–629.
Crompton, J.L. (1979). "Motivations for pleasure vacation". *Annals of Tourism Research*, 6(4), 408-424.
Crompton, J.L. (1992). "Structure of vacation destination choice sets". *Annals of Tourism Research*, 19, 420–434.
Crompton, J.L. and Ankomah, P.K. (1993). "Choice set propositions in destination choice". *Annals of Tourism Research*, 20, 461-476.
Cronin, C.J., Guilkey, D.K. and Speizer, I.S. (2019). "Measurement error in discrete health facility choice models: an example from urban Senegal". *Journal of Applied Econometrics*, 34, 1102-1120.
Crouch, G.I., Huybers, T. and Oppewal, H. (2016). "Inferring future vacation experience from past vacation choice: a latent class analysis". *Journal of Travel Research*, 55(5), 574-587.
Daganzo, C. (1979). *Multinomial Probit: the theory and its applications to demand forecasting.* Academic Press, New York.
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.
Davis, B. and Mangan, J. (1992). "Family expenditure on hotels and holiday". *Annals of Tourism Research*, 19, 669-691.
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.
Decrop, A. (2010). "Destination choice sets: an inductive longitudinal approach". *Annals of Tourism Research*, 37(1), 93-115.
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.
Dellaert, B.G.C., Ettema, D.F. and Lindh, C. (1998). "Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists' sequential choices of travel components". *Tourism Management*, 19(4), 313-320.
Dwyer, L. and Forsyth, P. (2011). "Methods of estimating destination price competitiveness: a case of horses for courses?" *Current Issues in Tourism*, 14(8), 751-777.
Dwyer, L., Forsyth, P. and Rao, P. (2000). "The price competitiveness of travel and tourism: a comparison of 19 destinations". *Tourism Management*, 21, 9-22.

Eugenio-Martin, J.L. (2003). “Modelling determinants of tourism demand as a five-stage process: a discrete choice methodological approach”. *Tourism and Hospitality Research*, 4(4), 341-354.
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.
Eugenio-Martin, J.L. and Campos-Soria, J.A. (2011). "Income and the substitution pattern between domestic and international tourism demand". *Applied Economics*, 43(20), 2519-2531.
Eugenio-Martin, J.L. and Campos-Soria, J.A. (2014). “Economic crisis and tourism expenditure cutback decision”. *Annals of Tourism Research*, 44, 53–73.
Eymann, A. and Ronning, G. (1992). *Discrete choice analysis of foreign travel demand*, in Vosgerau, H.J., ed, European Integration in the World Economy. Studies in International Economics and Institutions, Springer, Berlin, 641–664.
Eymann, A. and Ronning, G. (1997). "Microeconometric models of tourists' destination choice". *Regional Science and Urban Economics*, 27(6), 735–761.
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.
Falk, M. and Lin, X. (2018). "Sensitivity of winter tourism to temperature increases over the last decades". *Economic Modelling*, 71, 174-183.
Fiebig, D.G., Keane, M.P., Louviere, J. and Wasi, N. (2010). "The Generalized Multinomial Logit model: accounting for scale and coefficient heterogeneity". *Marketing Science*, 29(3), 393-421.
Fisichelli, N.A, Schuurman, G.W., Monahan, W.B. and Ziesler, P.S. (2015). "Protected area tourism in a changing climate: will visitation at us national parks warm up or overheat?" *PLoS ONE*, 10(6), e0128226.
García, J., Juaneda, C., Raya, J.M. and Sastre, F. (2015). "A study of traveller decision-making determinants: prioritizing destination or travel mode?". *Tourism Economics*, 21(6), 1149-1167.
Garín-Muñoz, T. (2006). "Inbound international tourism to Canary Islands: a dynamic panel data model". *Tourism Management*, 27, 281-291.
Garín-Muñoz, T. (2009). "Tourism in Galicia: domestic and foreign demand". *Tourism Economics*, 15(4), 753-769.
Garín-Muñoz, T. and Montero-Martín, L.F. (2007). "Tourism in the Balearic Islands: a dynamic model for international demand using panel data". *Tourism Management*, 28, 1224-1235.
Gkritzali, A., Gritzalis, D. and Stavrou, V. (2018). "Is Xenios Zeus still alive? Destination image of Athens in the years of recession". *Journal of Travel Research*, 57(4), 540-554.
Goh, C. (2012). "Exploring impact of climate on tourism demand". *Annals of Tourism Research*, 39(4), 1859-1883.
Gómez-Martín, B. (2005). “Weather, climate and tourism, a geographical perspective”. *Annals of Tourism Research*, 32(3), 571-591.
Gopinath, D., Schofield, M., Walker, J.L. and Ben-Akiva (2005). "Comparative analysis of discrete choice models with flexible substitution patterns". In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington, DC.
Gosens, T. and Rouwendal, J. (2018). "Nature-based outdoor recreation trips: duration, travel mode and location". *Transportation Research Part A*, 116, 513-530.
Gössling, S., Scott, D., Hall, C.M., Ceron, J.P. and Dubois, G. (2012). “Consumer behaviour and demand response of tourists to climate change”. *Annals of Tourism Research*, 39(1), 36-58.
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. 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.
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.
Grigolon, A.B. and Kemperman, A.D.A.M (2013). "Facet-based analysis of vacation planning processes: a binary mixed logit panel model". *Journal of Travel Research*, 52(2), 192-201.
Hamilton, L.C., Brown, C. and Keim, B.D. (2007). “Ski areas, weather and climate: time series models for New England case studies”. *International Journal of Climatology*, 27, 2113-2124.

Hasnat, M.M., Faghih-Imani, A., Eluru, N. and Hasan, S. (2019). "Destination choice modelling using location-based social media data". *Journal of Choice Modelling*, 31, 22-34.
Hausman, J. (2003). "Sources of bias and solutions to bias in the Consumer Price Index". *Journal of Economic Perspectives*, 17(1), 23-44.
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. and Greene, W.H. (2003). "The Mixed Logit model: the state of practise". *Transportation*, 30, 133-176.
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.
Hensher, D., Rose, J. and Greene, W. (2015). *Applied Choice Analysis*, second ed. Cambridge University Press, Cambridge, MA.
Herriges, J.A. and Phaneuf, D.J. (2002). "Inducing patterns of correlation and substitution in repeated logit models of recreation demand". *American Journal of Agricultural Economics*, 84(4), 1076-1090.
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.
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., Murphy, P., Le, H. and Leong, W. (2017). “Estimation of new monetary valuations of travel time, quality of travel and safety for Singapore”. *Transportation Research Record*, 2664, 79-90.
Hess, S. and Rose, J.M. (2012). "Can scale and coefficient heterogeneity be separated in random coefficients models?" *Transportation*, 39(6), 1225-1239.
Hess, S., Rose, J.M. and Bain, S. (2009). "Random scale heterogeneity in discrete choice models". Working Paper, 1-25.
Hess, S., Stathopoulos, A., Campbell, D., O’Neill, V. and Caussade, S. (2013). “It’s not I don’t care, I just don’t care very much: confounding between attribute non-attendance and taste heterogeneity”. *Transportation*, 40, 583-607.
Hess, S. and Train, K. (2017). "Correlation and scale in mixed logit models". *Journal of Choice Modelling*, 23, 1-8.
Hewer, M., Scott, D.J and Fenech, A. (2016). "Seasonal weather sensitivity, temperature thresholds, and climate change impacts for park visitation". *Tourism Geographies*, 18(3), 297-321.
Hewer, M., Scott, D.J. and Gough, W.A. (2017). "Differences in the importance of weather and weather-based decisions among campers in Ontario Parks (Canada)". *International Journal of Biometeorology*, 61(10), 1805-1818.
Hill, R.J. and Melser, D. (2008). "Hedonic imputation and the price index problem: an application to housing". *Economic Inquiry*, 46(4), 593-609.
Hong, S., Kim, J., Hang, H. and Lee, S. (2006). "The roles of categorization, affective image and constraints on destination choice: an application of the NMNL model". *Tourism Management*, 27, 750-761.
Hsu, T.K., Tsai, Y.F. and Wu, H.H. (2009). "The preference analysis for tourist choice of destination: a case study of Taiwan". *Tourism Management*, 30, 288-297.
Huber, J. and Train, K. (2001). “On the similarity of classical and bayesian estimates of individual mean partworths”. *Marketing Letters*, 12(2), 259-269.
Hyde, K.F. and Laesser, C. (2009). "A structural theory of the vacation". *Tourism Management*, 30, 240-248.
Hynes, S., Hanley, N. and Scarpa, R. (2008). "Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models". *American Journal of Agricultural Economics*, 90(4), 1011-1027.
Instituto Nacional de Estadística (INE) (2017). *Encuesta de Ocupación en Alojamientos de Turismo Rural.* Disponible en: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176963&menu=resultados&idp=1254735576863
Instituto Nacional de Estadística (INE) (2019). *Cuenta Satélite del Turismo en España. Revisión estadística 2019*. Disponible en: https://www.ine.es/prensa/cst_2018.pdf
Kah, J.A., Lee, C.K. and Lee. S.H. (2016). "Spatial-temporal distances in travel intention-behavior". *Annals of Tourism Research*, 57, 160-175.
Kaoru, Y. (1995). “Measuring marine recreation benefits of water quality improvements by the nested random utility model”. *Resource and Energy Economics*, 17(2),119–136.

Keane, M. (1997). “Quality and pricing in tourism destinations”. *Annals of Tourism Research*, 24, 117–130.
Keane, M. and Wasi, N. (2013). "Comparing alternative models of heterogeneity in consumer choice behavior". *Journal of Applied Econometrics*, 28, 1018-1045.
Kim, D. and Jang, S.S. (2013). "Price placebo effect in hedonic consumption". *International Journal of Hospitality Management*, 35, 306–315.
Kim, Y., Kim, C., Lee, D., Lee, H. and Andrada, R. (2019). “Quantifying nature-based tourism in protected areas in developing countries by using social big data”. *Tourism Management*, 72, 249-256.
Kim, J. and Seo, Y. (2019). "An evolutionary perspective on risk taking in tourism". *Journal of Travel Research*, 58(8), 1235-1248.
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.
Konu, H., Laukkanen, T. and Komppula, R. (2011). "Using ski destination choice criteria to segment Finnish ski resort customers". *Tourism Management*, 32, 1096-1105.
Kozak, M. (2002). “Comparative analysis of tourist motivations by nationality and destinations”. *Tourism Management*, 23(2), 221-232.
Lam, T. and Hsu, C.H.C. (2006). "Predicting behavioral intention of choosing a travel destination". *Tourism Management*, 27, 589-599.
Lancaster, J.K. (1966). “A new approach to consumer theory”. *Journal of Political Economy*, 74(2), 132–157.
Lee, L. (1992). “On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete choices”. *Econometrica*, 8, 518-552.
Lee, C. and Chang, C. (2008). "Tourism development and economic growth: a closer look at panels". *Tourism Management*, 29, 180-192.
Lepp, A. and Gibson, H. (2008). "Sensation seeking and tourism: tourist role, perception of risk and destination choice". *Tourism Management*, 29, 740-750.
Li, Q., Li, C., McCabe, S. and Xu, H. (2019). "Always best or good enough? The effect of 'mind-set' on preference consistency over time in tourist destination making". *Annals of Tourism Research*, 75, 186-201.
Li, H., Song, H. and Li, L. (2017). "A dynamic panel data analysis of climate and tourism demand: additional evidence". *Journal of Travel Research*, 56(2), 158-171.
Lise, W. and Tol, R.S.J. (2002). “Impact of climate on tourism demand”. *Climatic Change*, 55, 429-449.
Lithgow, D., Martínez, M.L., Gallego-Fernández, J., Silva, R. and Ramírez-Vargas (2019). "Exploring the co-occurrence between coastal squeeze and coastal tourism in a changing climate and its consequences". *Tourism Management*, 74, 43-54.
Liu, T. (2016). "The influence of climate change on tourism demand in Taiwan national parks". *Tourism Management Perspectives*, 20, 269-275.
Lorde, T., Li, G. and Airey, D. (2015). "Modelling Caribbean tourism demand: an augmented gravity approach". *Journal of Travel Research*, 55(7), 946-956.
Louviere, J.J., Meyer, R.J., Bunch, D.S., Carson, R., Dellaert, B., Hanemann, W.M., Hensher, D. and Irwin, J. (1999). "Combining sources of preference data for modelling complex decision processes". *Marketing Letters*, 10(3), 205-217.
Louviere, J.J., Hensher, D.A. and Swait, J.D (2000). *Stated choice methods. Analysis and application*. Cambridge University Press.
Luce, R.D. (1959). *Individual Choice Behavior*. Wiley, New York.
Lyons, S., Mayor, K. and Tol, R.S.J. (2009). "Holiday destinations: understanding the travel choices of Irish tourists". *Tourism Management,* 30, 683-692.
Maddison, D. (2001). “In search of warmer climates? The impact of climate change on flows of British tourists”. *Climatic change*, 49, 193-208.
Mansfeld, Y. (1992). “From motivation to actual travel”. *Annals of Tourism Research*, 19, 399–419.
Mariel, P., de Ayala, A., Hoyos, D. and Abdullah, S. (2013). "Selecting random parameters in discrete choice experiment for environmental valuation: a simulation experiment". *Journal of Choice Modelling*, 7, 44-57.
Mariel, P. and Meyerhoff, J. (2018). "A more flexible model or simply more effort? on the use of correlated random parameters in applied choice studies". *Ecological Economics*, 154, 419-429.
Marrocu, E. and Paci, R. (2013). “Different tourists to different destinations. Evidence from spatial interaction models”. *Tourism Management*, 39, 71-83.

Marschak, J. (1960). *Binary choice constraints on random utility indicators*. In K. Arrow, ed., Stanford Symposium on Mathematical Models in the Social Sciences, Stanford Univ. Press, 312-329.
Masiero, L. and Nicolau, J.L. (2012). "Tourism market segmentation based on price sensitivity: finding similar price preferences on tourism activities". *Journal of Travel Research*, 51(4), 426-435.
Masiero, L. and Qiu, R.T.R. (2018). "Modelling reference experience in destination choice". *Annals of Tourism Research*, 72, 58-74.
Massidda, C. and Etzo, I. (2012). "The determinants of Italian domestic tourism: a panel data analysis". *Tourism Management*, 33, 603-610.
McCabe, S. (2000). “Tourism motivation process”. *Annals of Tourism Research*, 27(4), 1049–1052.
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. (1978). *Modelling the choice of residential location*. In: Karlquist, A. et al. (Eds.), Spatial Interaction Theory and Residential Location. North Holland, Amsterdam, pp. 75–96.
McFadden, D. (2001). "Economic choices". *American Economic Review*, 91(3), 351-378.
McFadden, D. and Train, K. (2000). "Mixed MNL models for discrete response". *Journal of Applied Econometrics*, 15, 447-470.
McKercher, B. (2018). “The impact of distance on tourism: a tourism geography law”. *Tourism Geographies*, 3, 1–5.
Mckercher, B., Chan, A. and Lam, C. (2008). "The impact of distance on international tourist movements". *Journal of Travel Research*, 47(2), 208-224.
McKercher, B. and Guillet, B.D. (2011). “Are tourists or markets destination loyal?” *Journal of Travel Research*, 50(1), 121-132.
Mckercher, B. and Lew, A.A. (2003). "Distance decay and the impact of effective tourism exclusion zones on international travel flows". *Journal of Travel Research*, 42(2), 159-165.
Mckercher, B. and Mak, B. (2019). "The impact of distance on international tourism demand". *Tourism Management Perspectives*, 31, 340-347.
Mieczkowski, Z. (1985). “The Tourism Climatic Index: a method of evaluating world climates for tourism”. *Canadian Geographer*, 29(3), 220–233.
Moreno Sánchez, A. (2010). “Mediterranean tourism and climate (change): a survey-based study”. *Tourism and Hospitality Planning & Development*, 7(3), 253-265.
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.
Morey, E.R., Shaw, W.D. and Rowe, R.D. (1991). “A discrete choice model of recreational participation site choice, and activity valuation when complete trip data are not available”. *Journal of Environmental Economics and Management*, 20, 181-201.
Morley, C. (1992). “A microeconomic theory of international tourism demand”. *Annals of Tourism Research,* 19, 250–267.
Morley, C. L. (1994). “The use of CPI for tourism prices in demand modelling”. *Tourism Management*, 15(5), 342-346.
Morley, C.L. (1998). "A dynamic international demand model". *Annals of Tourism Research*, 25(1), 70-84.
Moulton, B.R. (1996). "Bias in the Consumer Price Index: what is the evidence?" *Journal of Economic Perspectives*, 10(4), 159-177.
Moutinho, L. (1987). “Consumer behaviour in tourism”. *European Journal of Marketing*, 21(10), 5–44.
Murdock, J. (2006). "Handling unobserved site characteristics in random utility models of recreation demand". *Journal of Environmental Economics and Management*, 51, 1-25.
Mussalan, G.Q. and Tajeddini, K. (2016). "Tourism in Switzerland: how perceptions of place attributes for short and long holiday can influence destination choice". *Journal of Hospitality and Tourism Management*, 26, 18-26.
Neuvonen, M., Pouta, E., Puustinen, J. and Sievänen, T. (2010). "Visits to national parks: effects of park characteristics and spatial demand". *Journal for Nature Conservation*, 18, 224-229.
Nicolau, J.L. (2008). "Characterizing tourist sensitivity to distance". *Journal of Travel Research*, 47, 43-52.
Nicolau, J.L. (2009). “The smile of the tourist: the relationship between price sensitivity and expenses”. *The Service Industries Journal*, 29(8), 1125-1134.
Nicolau, J.L. (2010a). "Variety-seeking and inertial behavior: the disutility of distance". *Tourism Economics*, 19(1), 251-264.

Nicolau, J.L. (2010b). "Anomaly in Spanish tourist sensitivity to price". *Tourism Economics*, 16(4), 915-923.
Nicolau, J.L. (2011). "Differentiated price loss aversion in destination choice: the effect of tourists' cultural interest". *Tourism Management*, 32, 1186-1195.
Nicolau, J.L. (2012). "Asymmetric tourist response to price: loss aversion segmentation". *Journal of Travel Research*, 51(5), 568-576.
Nicolau, J.L. and Más, F.J. (2005). "Stochastic modeling: a three-stage tourist choice process". *Annals of Tourism Research*, 32(1),49-69.
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.
Nicolau, J.L. and Más, F.J. (2008). "Sequential choice behavior: going on vacation and type of destination". *Tourism Management*, 29, 1023-1034.
Nicolau, J.L. and Masiero, L. (2013). "Relationship between price sensitivity and expenditures in the choice of tourism activities at the destination". *Tourism Economics*, 19(1), 101-114.
Nordhaus, W.D. (1998). "Quality change in price indexes". *Journal of Economic Perspectives*, 12(1), 59-68.
Nordström, J. (2005). "Dynamic and stochastic structures in tourism demand modeling". *Empirical Economics*, 30, 379-392.
Paci, R. and Marrocu, E. (2014). "Tourism and regional growth in Europe". *Papers in Regional Science*, 93, S25–S60.
Pagliara, F., Mauriello, F. and Garofalo, A. (2017). "Exploring the interdependences between High Speed Rail systems and tourism: some evidence from Italy". *Transportation Research Part A: Policy and Practice*, 106, 300-308.
Papatheodorou, A. (2001). "Why people travel to different places". *Annals of Tourism Research*, 28, 164–179.
Papatheodorou, A. (2002). "Exploring competitiveness in Mediterranean resorts". *Tourism Economics*, 8, 133-150.
Park, S. and Nicolau, J.L. (2018). "If you, tourist, behave irrationally, I'll find you!". *Tourism Management*, 69, 434-439.
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.
Pattuelli, R., Mussoni, M. and Candela, G. (2013). "The effects of World Heritage sites on domestic tourism: a spatial interaction model for Italy". *Journal of Geographic Systems*, 15, 396–402.
Pearce, P.L. and Lee, U.I. (2005). "Developing the travel career approach to tourist motivation". *Journal of Travel Research*, 43, 226–237.
Perles-Ribes, J.F., Ramón-Rodríguez, A.B., Such-Devesa, M.J. and Moreno-Izquierdo, L. (2019). "Effects of political instability in consolidated destinations: the case of Catalonia (Spain)". *Tourism Management*, 70, 134-139.
Pestana-Barros, C., Butler, R. and Correia, A. (2008). "Heterogeneity in destination choice. Tourism in Africa". *Journal of Travel Research*, 47(2), 235-246.
Plog, S. (2002). "The power of psychographics and the concept of venturesomeness". *Journal of Travel Research*, 40, 244–251.
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.
Pongkijvorasin, S. and Chotiyaputta, V. (2013). "Climate change and tourism: Impacts and responses. A case study of Khaoyai National Park". *Tourism Management Perspectives*, 5, 10-17.
Qiu, R.T.R., Masiero, L. and Li, G. (2018). "The psychological process of travel destination choice". *Journal of Travel and Tourism Marketing*, 35(6), 691-705.
Rao, A.R. (2005). "The quality of price as a quality cue". *Journal of Marketing Research*, 42(4), 401–405.
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.
Richardson, R.B. and Loomis, J.B. (2004). "Adaptative recreation planning and climate change: a contingent visitation approach". *Ecological Economics*, 50(1-2), 83-99.
Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P. and Martens, P. (2014). "Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: evidence from Aruba". *Tourism Management*, 41, 245-256.
Riera, A. (2000a). "Modelos de elección discreta y coste del viaje. Los espacios naturales protegidos en Mallorca". *Revista de Economía Aplicada*, 24, 181-201.

Riera, A. (2000b). “Mass tourism and the demand for protected natural areas: a Travel Cost approach”. *Journal of Environmental Economics and Management*, 39(1), 97-116.
Rosselló-Nadal, J. (2014). "How to evaluate the effects of climate change on tourism". *Tourism Management*, 42, 334-340.
Rosselló-Nadal, J., Riera, A., and Cárdenas, V. (2011). “The impact of weather variability on British outbound flows”. *Climatic Change*, 105, 281-292.
Rosselló-Nadal, J., Riera, A. and Sansó-Rosselló, A. (2004). "The economic determinants of seasonal patterns". *Annals of Tourism Research*, 31(3), 697-711.
Rugg, D. (1973). “The choice of journey destination: A theoretical and empirical analysis”. *Review of Economics and Statistics,* 55(1), 64–72.
Rutty, M., Scott, D., Johnson, P., Jover, E., Pons, M. and Steiger, R. (2015). “Behavioural adaptation of skiers to climatic variability and change in Ontario, Canada”. *Journal of Outdoor Recreation and Tourism*, 11, 13–21.
Santana-Gallego, M., Fourie, J. and Rosselló, J. (2020). "The effect of safety and security issues on international tourism". *Annals of Tourism Research*, https://doi.org/10.1016/j.annals.2019.02.004
Scarpa, R., Ferrini, S, and Willis, K. (2005). *Performance of error component models for status quo effects in choice experiments*. In: Bateman, I.J. (Ed.), The Economics of Non-Market Goods and Resources. Springer, Dordrecht, pp. 247–273 (chapter 13).
Scarpa, R., Thiene, M. and Hensher, D.A. (2010). “Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter?” *Land Economics*, 86(4), 817-839.
Scarpa, R., Thiene, M. and Train, K. (2008). "Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps". *American Journal of Agricultural Economics*, 90(4), 994-1010.
Scarpa, R., Willis, K.G. and Acutt, M. (2007). “Valuing externalities from water supply: status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments”. *Journal of Environmental Planning and Management*, 50(4), 449-466.
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.
Scott, D., Simpson, M.C. and Sim, R. (2012). “The vulnerability of Caribbean coastal tourism to scenarios of climate change related sea level rise”. *Journal of Sustainable Tourism*, 20(6), 883-898.
Seabra, C., Reis, P. and Abrantes, J.L. (2020). "The influence of terrorism in tourism arrivals: a longitudinal approach in a Mediterranean country". *Annals of Tourism Research*, https://doi.org/10.1016/j.annals.2019.102811
Seddighi, H.R. and Theocharous, A.L. (2002). "A model of tourism destination choice: a theoretical and empirical analysis". *Tourism Management*, 23, 475-487.
Sequeira, T.N. and Nunes, P.M. (2008). “Does tourism influence economic growth? A dynamic panel data approach”. *Applied Economics*, 40, 2431-2441.
Serquet, G. and Rebetez, M. (2011). “Relationship between tourism demand in the Swiss Alps and hot summer air temperatures associated with climate change”. *Climatic Change*, 108, 291-300.
S.G de Análisis, Perspectiva y Coordinación (SGAPC) (2017). “El turismo de naturaleza en España”. *Análisis y prospectiva serie Medio Ambiente*, 9, 1-14. Edita: Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. NIPO: 013-17-154-8. Disponible en:
https://www.miteco.gob.es/es/biodiversidad/temas/conservacion-de-la-biodiversidad/seriemedioambienten9_turismodenaturalezaenespana_tcm7-464178_tcm30-481336.pdf
Simon, H.A. (1955). “A behavioral model of rational choice”. *Quarterly Journal of Economics*, 69(1), 99-118.
Smallman, C. and Moore, K. (2010). "Process studies of tourists' decision-making". *Annals of Tourism Research*, 37(2), 397-422.
Smith, J.W., Wilkins, E., Gayle, R. and Lamborn, C.C. (2018). "Climate and visitation to Utah's 'Mighty 5' national parks". *Tourism Geographies*, 20(2), 250-272.
Steiger, R., Scott, D., Abegg, B., Pons, M. and Aall, C. (2019). "A critical review of climate change risk for ski tourism". *Current Issues in Tourism*, 22(11), 1343-1379.
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.
Swait, J. and Louviere, J. (1993). "The role of the scale parameter in the estimation and comparison of Multinomial Logit Models". *Journal of Marketing Research*, 30(3), 305-314.

Tanford, S., Choi, C. and Jun Joe, S. (2019). "The influence of pricing strategies on willingness to pay for accommodations: anchoring, framing and metric comparability". *Journal of Travel Research*, 58(6), 932-944.
Taylor, C.E. and Knudson, D.M. (1976). “Area preferences of Midwestern campers”. *Journal of Leisure Research*, 7, 39-48.
Thiene, M., Franceschinis, C. and Scarpa, R. (2019). "Congestion management in protected areas: accounting for respondents' inattention and preference heterogeneity in stated choice data". *European Review of Agricultural Economics*, 46(5), 834-861.
Thiene, M., Swait, J. and Scarpa, R. (2017). "Choice set formation for outdoor destinations: the role of motivations and preference discrimination in site selection for the management of public expenditures on protected areas". *Journal of Environmental Economics and Management*, 81, 152-173.
Thrane, C. (2015). "Examining tourists' long-distance transportation mode choices using a Multinomial logit regression model". *Tourism Management Perspectives*, 15, 115-121.
Thurstone, L. (1927). “A law of comparative judgment”. *Psychological Review*, 34, 273- 286.
Train, K. (1998). "Recreation demand models with taste variation". *Land Economics*, 74, 230–239.
Train, K. (1999). "Halton sequences for mixed logit". Working Paper, Department of Economics, University of California at Berkeley.
Train, K. (2009). *Discrete choice models with simulation*. Second Edition. Cambridge University Press.
Trope, Y. and Liberman, N. (2010). "Construal-level theory of psychological distance". *Psychological Review*, 117(2), 440–463.
United National World Tourism Organization (UNWTO) (2018). *Statistical Data.* Available at: https://www.unwto.org/data
Um, S. and Crompton, J.L. (1992). “The roles of perceived inhibitors and facilitators in pleasure travel destination decisions”. *Journal of Travel Research*, 30(3), 18–25.
Van Nostrand, C., Sivaraman, V. and Pinjari, A.R. (2013). "Analysis of long-distance vacation travel demand in the United States: a multiple discrete-continuous choice framework". *Transportation*, 40, 151-171.
Van Raaij, W.F. and Francken, D.A. (1984). “Vacation decisions, activities and satisfactions”. *Annals of Tourism Research*, 11, 101–112.
Wakamatsu, M., Shin, K.J., Wilson, C. and Managi, S. (2018). “Exploring a gap between Australia and Japan in the economic valuation of whale conservation”. *Ecological Economics*, 146, 397–407.
Walker, J.L., Ben-Akiva, M. and Bolduc, D. (2007). "Identification of parameters in Normal Error Component Logit-Mixture (NECLM) models". *Journal of Applied Econometrics*, 22, 1095-1125.
Whitehead, J.C. and Wicker, P. (2018). "Estimating willingness to pay for a cycling event using a willingness to travel approach". *Tourism Management*, 65, 160-169.
Wong, I., Fong, L., and Law, R. (2016). "A longitudinal multilevel model of outbound travel behaviour and the dual cycle model". *Journal of Travel Research*, 55(7), 957–970.
Wong, I., Law, R. and Zhao, X. (2017). "When and where to travel? A longitudinal multilevel investigation on destination choice and demand". *Journal of Travel Research*, 56(7), 868-880.
Wu, L., Zhang, J. and Fujiwara, A. (2011). "Representing tourists’ heterogeneous choices of destination and travel party with an integrated latent class and nested logit model". *Tourism Management*, 32, 1407-1413.
Yang, Y., Fik, T. and Zhang, J. (2013). "Modelling sequential tourist flows: where is the next destination?". *Annals of Tourism Research*, 43, 297-320.
Yang, Y., Liu, Z. and Qi, Q. (2014). "Domestic tourism demand of urban and rural residents in China: does relative income matter?" *Tourism Management*, 40, 193-202.
Yang, Y., Liu, H., Li, X. and Harrill, R. (2018). "A shrinking world for tourists? Examining the changing role of distance factors in understanding destination choices". *Journal of Business Research*, 92, 350-359.
Yoon, Y. and Uysal, M. (2005). "An examination of the effects of motivation and satisfaction on destination loyalty: A structural model". *Tourism Management*, 26, 45–56.
Zhang, M., Zhang, G., Dursoy, D. and Fu, X. (2018). "Message framing and regulatory focus effects on destination image formation". *Tourism Management*, 69, 397-407.
Zimmer, Z., Brayley, R.U. and Searle, M.S. (1995). "Whether to go and where to go: identification of important influences on seniors' decisions to travel". *Journal of Travel Research*, 33(3), 3-10.

SUPPLEMENTARY MATERIAL

Modelling Heterogeneous Preferences for Nature-based Recreational Trips

ANNEX 1.- Histograms of the attributes

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.

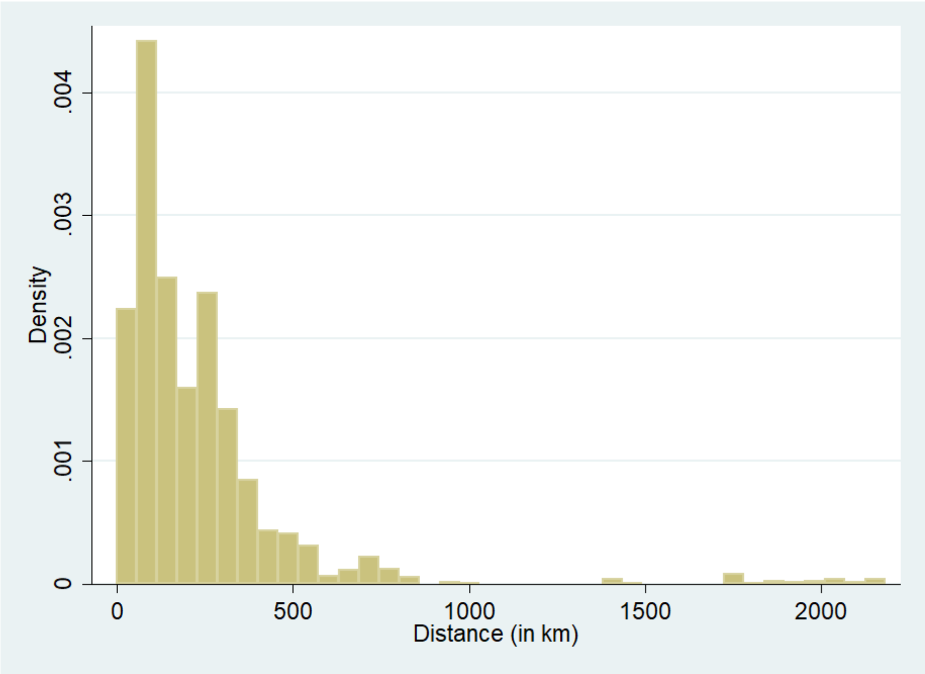


Figure A1.- Histogram of *DIST* (in kilometres)

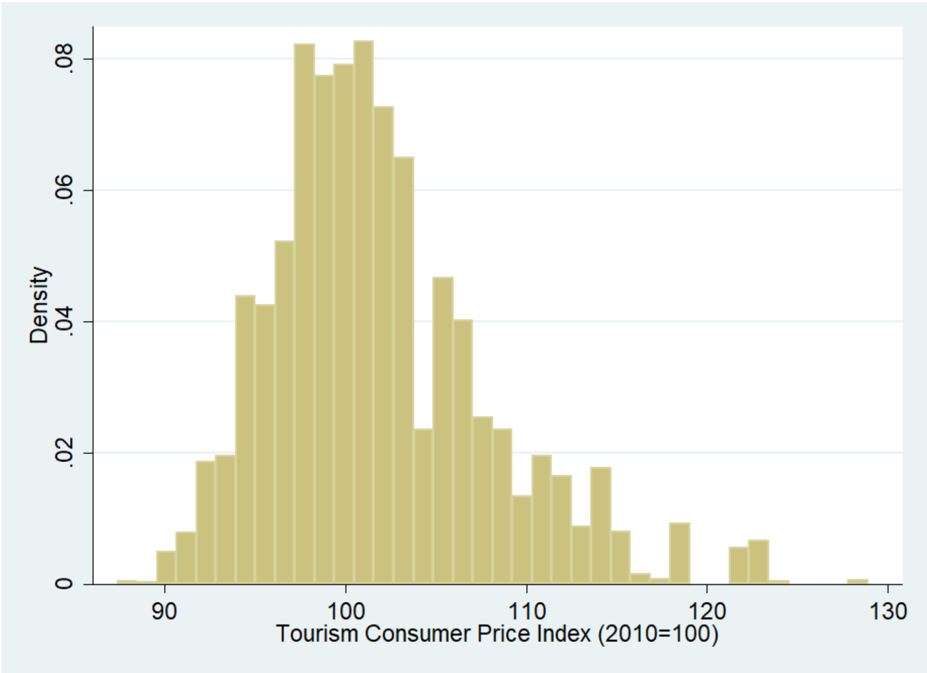


Figure A2.- Histogram of *TCPI*

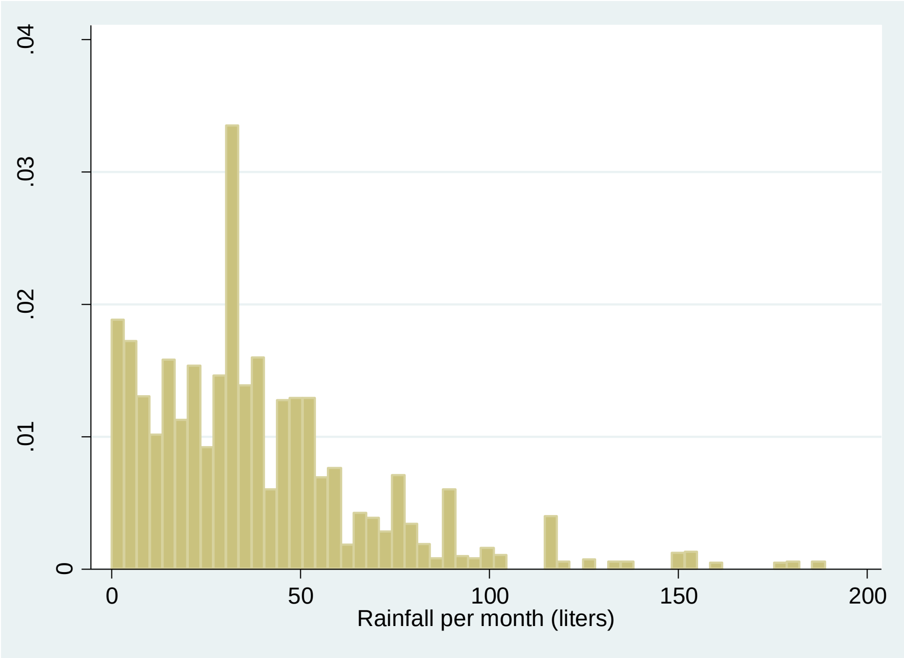


Figure A3.- Histogram of *RAINFALL* (liters per month)

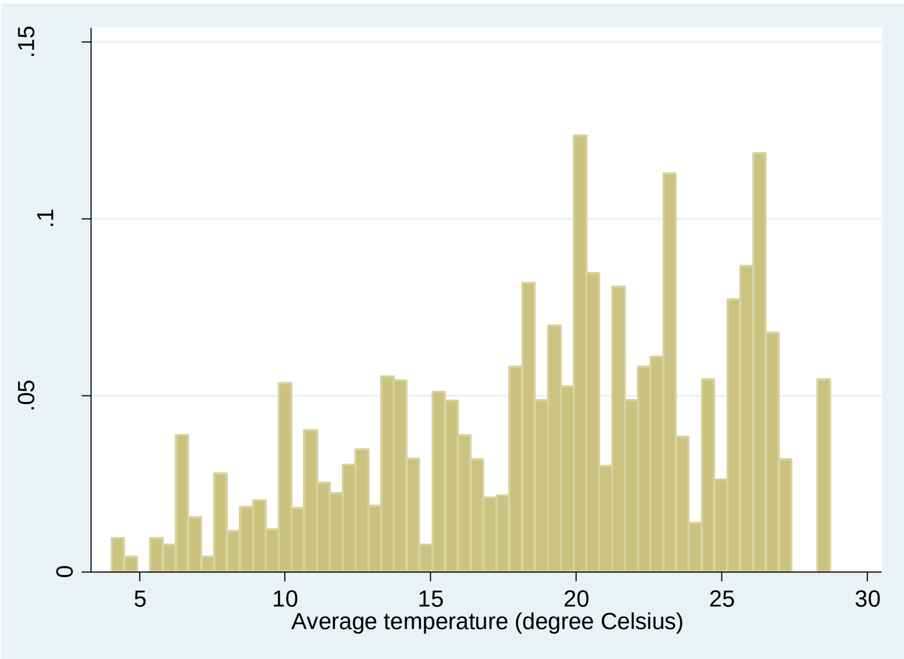


Figure A4.- Histogram of *TEMPERATURE* (degree Celsius)

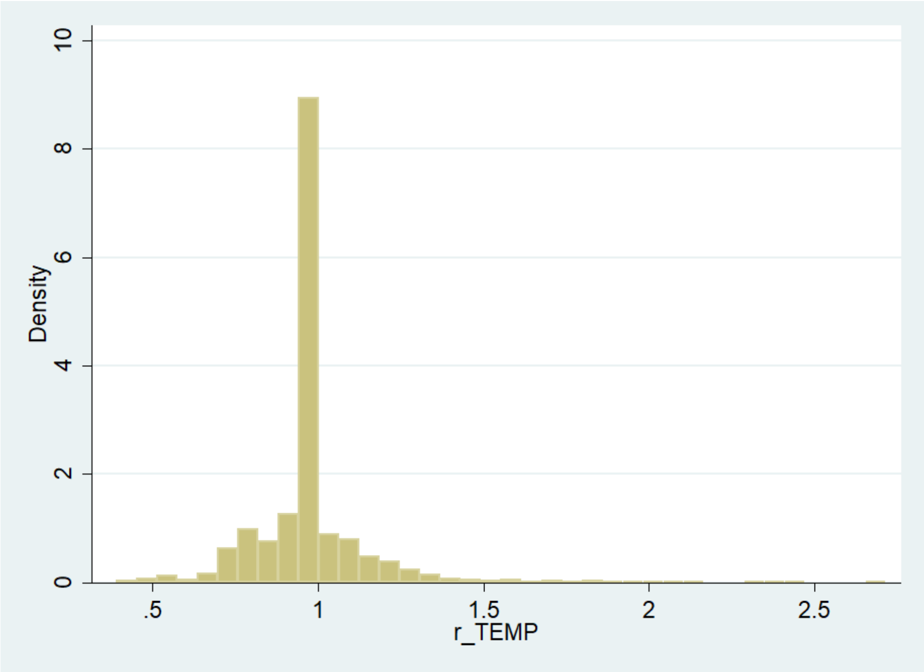


Figure A5.- Histogram of *r_TEMP*

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

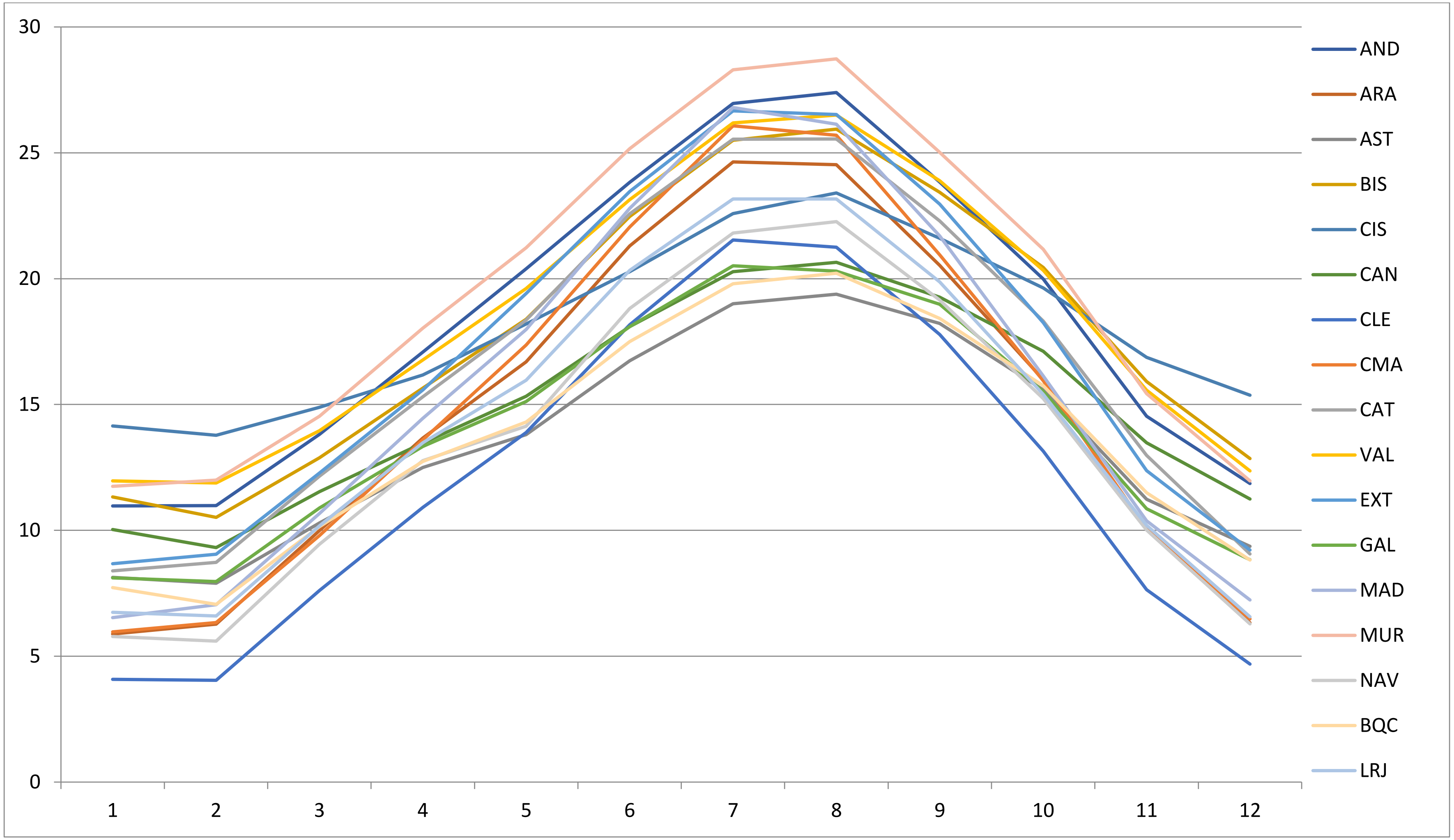


Figure A6.- Monthly average temperature by region

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.

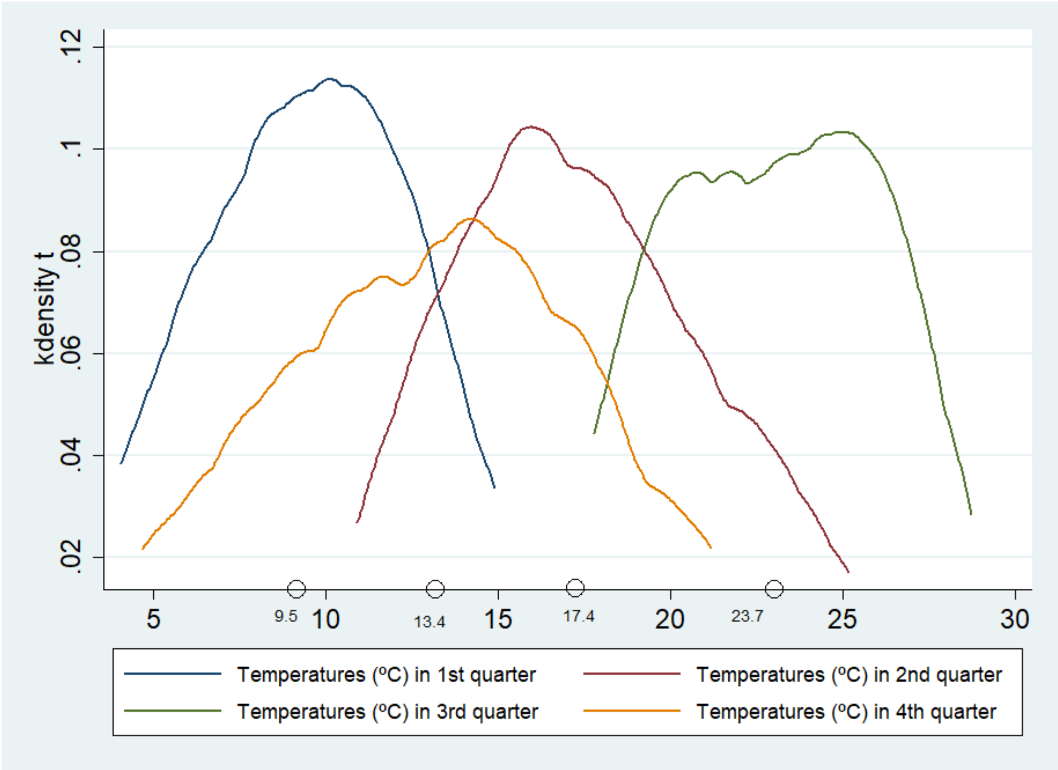


Figure A7.- Smooth kernel density plot for temperature per quarter

ANNEX 2.- List of Tourism Spots (Municipalities)

Municipality code	Tourism spot	NUTS 1	NUTS 2	NUTS 3
1059	Vitoria-Gasteiz	NORESTE	País Vasco	Álava
2003	Albacete	CENTRO	Castilla la Mancha	Albacete
3014	Alicante/Alacant	ESTE	Comunidad Valenciana	Alicante
3031	Benidorm	ESTE	Comunidad Valenciana	Alicante
3063	Dénia	ESTE	Comunidad Valenciana	Alicante
3065	Elche/Elx	ESTE	Comunidad Valenciana	Alicante
3133	Torrevieja	ESTE	Comunidad Valenciana	Alicante
4013	Almería	SUR	Andalucía	Almería
4064	Mojácar	SUR	Andalucía	Almería
4066	Níjar	SUR	Andalucía	Almería
4079	Roquetas de Mar	SUR	Andalucía	Almería
5019	Ávila	CENTRO	Castilla y León	Ávila
6015	Badajoz	CENTRO	Extremadura	Badajoz
6083	Mérida	CENTRO	Extremadura	Badajoz
7011	Calvià	ESTE	Islas Baleares	Islas Baleares
7014	Capdepera	ESTE	Islas Baleares	Islas Baleares
7040	Palma de Mallorca	ESTE	Islas Baleares	Islas Baleares
7051	Sant Llorenç des Cardassar	ESTE	Islas Baleares	Islas Baleares
8019	Barcelona	ESTE	Cataluña	Barcelona
8270	Sitges	ESTE	Cataluña	Barcelona
9059	Burgos	CENTRO	Castilla y León	Burgos
10037	Cáceres	CENTRO	Extremadura	Cáceres
10148	Plasencia	CENTRO	Extremadura	Cáceres
10195	Trujillo	CENTRO	Extremadura	Cáceres
11004	Algeciras	SUR	Andalucía	Cádiz
11006	Arcos de la Frontera	SUR	Andalucía	Cádiz
11012	Cádiz	SUR	Andalucía	Cádiz
11020	Jerez de la Frontera	SUR	Andalucía	Cádiz
11027	Puerto de Santa María	SUR	Andalucía	Cádiz
11035	Tarifa	SUR	Andalucía	Cádiz
12040	Castellón de la Plana	ESTE	Comunidad Valenciana	Castellón

12089	Peníscola/Peñíscola	ESTE	Comunidad Valenciana	Castellón
13034	Ciudad Real	CENTRO	Castilla la Mancha	Ciudad Real
14021	Córdoba	SUR	Andalucía	Córdoba
15030	A Coruña	NOROESTE	Galicia	La Coruña
15078	Santiago de Compostela	NOROESTE	Galicia	La Coruña
16078	Cuenca	CENTRO	Castilla la Mancha	Cuenca
17095	Lloret de Mar	ESTE	Cataluña	Gerona
18087	Granada	SUR	Andalucía	Granada
19257	Sigüenza	CENTRO	Castilla la Mancha	Guadalajara
20069	Donostia/San Sebastián	NORESTE	País Vasco	Guipúzcoa
22054	Benasque	NORESTE	Aragón	Huesca
22130	Jaca	NORESTE	Aragón	Huesca
22204	Sallent de Gállego	NORESTE	Aragón	Huesca
23028	Cazorla	SUR	Andalucía	Jaén
24089	León	CENTRO	Castilla y León	León
24115	Ponferrada	CENTRO	Castilla y León	León
25025	Naut Aran	ESTE	Cataluña	Lérida
25120	Lleida	ESTE	Cataluña	Lérida
25243	Vielha e Mijaran	ESTE	Cataluña	Lérida
26089	Logroño	NORESTE	La Rioja	La Rioja
27028	Lugo	NOROESTE	Galicia	Lugo
27051	Ribadeo	NOROESTE	Galicia	Lugo
27066	Viveiro	NOR	Galicia	Lugo
28079	Madrid	COMUNIDAD DE MADRID	Comunidad de Madrid	Madrid
29015	Antequera	SUR	Andalucía	Málaga
29025	Benalmádena	SUR	Andalucía	Málaga
29051	Estepona	SUR	Andalucía	Málaga
29054	Fuengirola	SUR	Andalucía	Málaga
29067	Málaga	SUR	Andalucía	Málaga
29069	Marbella	SUR	Andalucía	Málaga
29075	Nerja	SUR	Andalucía	Málaga
29084	Ronda	SUR	Andalucía	Málaga
29901	Torremolinos	SUR	Andalucía	Málaga
30016	Cartagena	SUR	Murcia	Murcia
30030	Murcia	SUR	Murcia	Murcia
31201	Pamplona/Iruña	NORESTE	Comunidad Foral de Navarra	Navarra
32054	Ourense	NOROESTE	Galicia	Orense

33012	Cangas de Onís	NOROESTE	Asturias	Asturias
33024	Gijón	NOROESTE	Asturias	Asturias
33036	Llanes	NOROESTE	Asturias	Asturias
33044	Oviedo	NOROESTE	Asturias	Asturias
33076	Villaviciosa	NOROESTE	Asturias	Asturias
34120	Palencia	CENTRO	Castilla y Leon	Palencia
35012	Mogán	CANARIAS	Canarias	Las Palmas
35015	Pájara	CANARIAS	Canarias	Las Palmas
35016	Las Palmas de Gran Canaria	CANARIAS	Canarias	Las Palmas
35019	San Bartolomé de Tijarana	CANARIAS	Canarias	Las Palmas
35024	Teguise	CANARIAS	Canarias	Las Palmas
35028	Tías	CANARIAS	Canarias	Las Palmas
35034	Yaiza	CANARIAS	Canarias	Las Palmas
36022	O Grove	NOROESTE	Galicia	Pontevedra
36051	Sanxenxo	NOROESTE	Galicia	Pontevedra
36057	Vigo	NOROESTE	Galicia	Pontevedra
37274	Salamanca	CENTRO	Castilla y León	Salamanca
38001	Adeje	CANARIAS	Canarias	Santa Cruz de Tenerife
38006	Arona	CANARIAS	Canarias	Santa Cruz de Tenerife
38028	Puerto de la Cruz	CANARIAS	Canarias	Santa Cruz de Tenerife
38038	Santa Cruz de Tenerife	CANARIAS	Canarias	Santa Cruz de Tenerife
39075	Santander	NOROESTE	Cantabria	Cantabria
40194	Segovia	CENTRO	Castilla y León	Segovia
41091	Sevilla	SUR	Andalucía	Sevilla
42173	Soria	CENTRO	Castilla y León	Soria
43038	Cambrils	ESTE	Cataluña	Tarragona
43148	Tarragona	ESTE	Cataluña	Tarragona
43905	Salou	ESTE	Cataluña	Tarragona
44009	Albarracín	NORESTE	Aragón	Teruel
44216	Teruel	NORESTE	Aragón	Teruel
45168	Toledo	CENTRO	Castilla la Mancha	Toledo
46131	Gandia	ESTE	Comunidad Valenciana	Valencia
46250	Valencia	ESTE	Comunidad Valenciana	Valencia
47186	Valladolid	CENTRO	Castilla y León	Valladolid
48020	Bilbao	NORESTE	País Vasco	Vizcaya
49021	Benavente	CENTRO	Castilla y León	Zamora
49275	Zamora	CENTRO	Castilla y León	Zamora
50297	Zaragoza	NOROESTE	Aragón	Zaragoza

Table A2.- List of tourism spots (municipalities)

ANNEX 3.- Lagrange Multiplier Test

Although the standard way to examine whether the marginal utility of an attribute is heterogeneous is by means of a zero-based t-test, another relevant diagnosis test is the Lagrange Multiplier test proposed by McFadden and Train (2000). The test is particularly suitable for wide distribution spreads, since Mariel et al. (2013) come upon evidence that the test power is larger for normal distributions as compared with uniform or triangular ones. The test consists of constructing the following artificial variables W_{ij} for each attribute k as follows:

$$W_{ijk} = \frac{(X_{ijk} - \overline{X_{\iota k}})^2}{2}, \qquad \text{with } \overline{X_{\iota k}} = \sum_j X_{ijk}\ P_{ij}$$

where X_{ijk} are the destination-specific attributes and P_{ij} is the estimated MNL (conditional logit) choice probability for individual i choosing region j.

The artificial variables are included in a Conditional Logit model together with the attributes X_{ijk}. The null of no random coefficients on attribute k is rejected if the parameter estimates for the artificial variable W_{ijk} are significantly different from zero. The joint significance of more than one random attribute can be assessed by using a Wald or a LR test statistic.

Table A3 presents the results of a conditional logit model with two artificial variables for testing whether *DIST* and *r_TEMP* should be allowed to be random. The two auxiliary variables are individually statistically significant. A LR test (Chi(2)=1588, p-value<0.01) further supports the necessity of allowing these two variables to be randomly distributed.

	Baseline MNL		Extended MNL	
Variable	Coef.	SE	Coef.	SE
$REG1$	-5.042***	(0.218)	1.678***	(0.151)
$REG2$	-6.237***	(0.250)	0.533***	(0.194)
$REG3$	-6.674***	(0.264)	-0.083	(0.213)
$REG4$	-6.561***	(0.246)	0.222	(0.196)
$REG5$	-6.335***	(0.218)	0.138	(0.132)
$REG6$	-7.068***	(0.251)	-0.509***	(0.195)
$DIST$	-0.662***	(0.009)	-0.928***	(0.024)
r_TEMP	0.347***	(0.128)	1.456***	(0.257)
$RAIN$	-0.182***	(0.054)	-0.257***	(0.057)
$TCPI$	-0.005	(0.005)	-0.002	(0.005)
TOU_SPOTS	0.106***	(0.008)	0.113***	(0.009)
NAT_PARKS	0.047***	(0.007)	0.055***	(0.007)
$SIZE\_NAT$	0.001*	(0.001)	0.001	(0.001)
SKI_KM	0.001***	(1.73e-04)	0.001***	(1.83e-04)
$COAST$	-1.281***	(0.151)	-1.278***	(0.156)
$SKI\_KM * winter\_sports$	0.008***	(3.66e-04)	0.008***	(4.25e-04)
$COAST * aquatic$	2.530***	(0.110)	2.690***	(0.124)
W_DIST			0.072***	(0.002)
W_r_TEMP			-0.847***	(0.228)
N		6,661		

Table A3.- Results for baseline and extended MNL model for LM test
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

ANNEX 4.- Robustness Checks

	RPLc-ECM	
Variable	Coef.	SE
$REG1$	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
$SIZE\_NAT$	0.001	9.0e-04
SKI_KM	0.001***	2.3e-04
$SKI\_KM * winter\_sports$	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_TEMP)	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 A4.- Parameter estimates for RPLc-ECM without taste shifters
*** p<0.01, ** p<0.05, * p<0.1

Here we examine the robustness of our estimates to the distribution assumed for the random parameters. Table A4 presents the results from the estimation of equations (2.19), (2.21) and (2.22) assuming a triangular and a uniform distribution. Table A5 shows the results under the assumption that the random components follow a truncated normal and a Weibull distribution.

	Triangular		Uniform	
	$\beta_{ki} = \beta_k + \delta_k{}'Z_i + \sigma_k v_{ik}$ $v_{ik} \sim Triangle[-1,1]$		$\beta_{ki} = \beta_k + \delta_k{}'Z_i + \sigma_k v_{ik}$ $v_{ik} \sim U[-1,1]$	
Variable	Coef.	SE	Coef.	SE
$REG1$	3.023***	0.3564	2.690***	0.2996
$REG2$	1.711***	0.3767	1.356***	0.3247
$REG3$	–0.296	0.4572	–0.626	0.2749
$REG4$	1.365***	0.3738	0.998***	0.4140
$REG5$	1.072***	0.3314	0.728***	0.3221
$REG6$	0.2058	0.3646	–0.112***	0.3147
$DIST$	–0.513***	0.0559	–0.538***	0.0524
r_TEMP	2.224***	0.7499	2.205***	0.7506
$RAIN$	–0.333***	0.0722	–0.337***	0.0727
$TCPI$	–0.011*	0.0061	–0.011*	0.0061
TOU_SPOTS	0.120***	0.0095	0.119***	0.0096
NAT_PARKS	0.056***	0.0089	0.059***	0.0090
$SIZE\_NAT$	0.002**	9.4e-04	0.002**	9.4e-04
SKI_KM	0.001***	2.3e-04	0.001***	2.4e-04
$SKI\_KM * winter\_sports$	0.009***	5.5e-04	0.009***	5.4e-04
$COAST$	–1.346***	0.1866	–1.397***	0.1872
$COAST * aquatic$	2.801***	0.1264	2.826***	0.1263
SD $DIST$	1.106***	0.0395	0.824****	0.0271
SD r_TEMP	4.543***	0.8405	3.085***	0.5932
Cov($DIST$, r_TEMP)	0.612	0.8100	0.299	0.4176
ϑ_1	0.085	0.6634	0.048	0.7707
ϑ_2	1.006***	0.1354	1.060***	0.1303
ϑ_3	0.158	0.4251	0.018	0.5971
ϑ_4	2.018***	0.2335	1.978***	0.2316
$DIST$ Mean shifters				
age	0.002***	8.2e-04	2.6e-03***	7.4e-04
inc2	0.053**	0.0242	0.044**	0.0212
inc3	0.152***	0.0351	0.127***	0.0323
parsize	–0.093***	0.0100	–0.091***	0.0092
weekend	–0.567***	0.0278	–0.554***	0.0269
q1	–0.001	0.0486	9.7e-03	0.0442
q2	–0.001	0.0400	1.8e-03	0.0355
q4	–0.197***	0.0518	–0.187***	0.0470
d_warmorigin_1	–0.014	0.0487	–0.015	0.0446
d_warmorigin_2	–0.142***	0.0433	–0.140***	0.0389
d_warmorigin_3	–0.123***	0.0362	–0.107***	0.0321
d_warmorigin_4	0.101*	0.0609	0.095	0.0549
winter_sports	–0.118	0.0749	–0.129*	0.0762
mou_trek_nat	–0.184***	0.0281	–0.166***	0.0267

rural	–0.039	0.0243	–0.039*	0.0217
aquatic	0.276***	0.0268	0.260***	0.0240
advent	0.065**	0.0287	0.069***	0.0255
r_TEMP Mean shifters				
age	0.012	0.0092	0.011	9.2e-03
inc2	0.160	0.2765	0.147	0.2754
inc3	0.344	0.3548	0.291	0.3531
parsize	–0.159	0.1033	–0.147	0.1025
weekend	–0.705***	0.2604	–0.677***	0.2597
q1	–1.271**	0.5504	–1.212**	0.5515
q2	–0.763	0.6159	–0.748	0.6165
q4	–1.419**	0.5582	–1.357**	0.5580
d_warmorigin_1	2.022***	0.4246	1.970***	0.4223
d_warmorigin_2	1.018	0.6213	1.082*	0.6210
d_warmorigin_3	–3.026***	0.6136	–2.763***	0.6090
d_warmorigin_4	–0.321	0.5932	–0.268	0.5867
winter_sports	–0.572	0.4168	–0.517	0.4141
mou_trek_nat	–0.690**	0.2769	–0.652**	0.2761
rural	–0.803***	0.2860	–0.815***	0.2851
aquatic	0.294	0.4054	0.316	0.4048
advent	1.443***	0.3098	1.435***	0.3087
Log L	-11,182.5		–11,206.8	
AIC	22,481.1		25,529.8	
Pseudo-R2	0.407		0.406	
N	6,661		6,661	

Table A5.- Parameter estimates for RPLc-ECM assuming Triangular and Uniform distributions
*** p<0.01, ** p<0.05, * p<0.1

	Truncated normal		Weibull	
	$\beta_{ki} = \beta_k + \delta_k{}'Z_i + \sigma_k \nu_{ik}$ ν_{ik} ~ N (0, σ; -1.96,1.96)		$\beta_{ki} = \beta_k + \delta_k{}'Z_i + \sigma_k \nu_{ik}$ $\nu_{ik} \sim 2(-\log(u_i))^{\sqrt{5}}$ $u_i \sim U[0,1]$	
Variable	Coef.	SE	Coef.	SE
$REG1$	3.091***	0.3501	3.958***	0.4571
$REG2$	1.772***	0.3706	2.679***	0.4750
$REG3$	–0.211	0.4454	0.793	0.5371
$REG4$	1.424***	0.3684	2.376***	0.4729
$REG5$	1.136***	0.3265	2.077***	0.4381
$REG6$	0.273	0.3603	1.191***	0.4611
$DIST$	–0.520***	0.0558	–0.976***	0.0646
r_TEMP	2.190***	0.7517	–0.517	0.9411
$RAIN$	–0.331***	0.0722	–0.328***	0.0709
$TCPI$	–0.011*	0.0061	–0.012**	0.0060
TOU_SPOTS	0.120***	0.0095	0.122***	0.0093
NAT_PARKS	0.057***	0.0089	0.049***	0.0086
$SIZE\_NAT$	0.002**	9.4e-04	0.002***	9.1e-04
SKI_KM	0.001***	2.3e-04	0.001***	2.3e-04
$SKI\_KM * winter\_sports$	0.009***	5.4e-04	0.009***	5.3e-04
$COAST$	–1.364***	0.1868	–1.274***	0.1844
$COAST * aquatic$	2.810***	0.1258	2.740***	0.1234
SD $DIST$	0.526***	0.0178	0.275****	0.0117
SD r_TEMP	2.137***	0.3938	1.422***	0.2578
Cov($DIST$, r_TEMP)	0.119	0.1786	0.010	0.0592
ϑ_1	0.034	0.7013	0.028	1.062
ϑ_2	1.032***	0.1310	0.970***	0.1340
ϑ_3	0.051	0.4965	0.033	0.7084
ϑ_4	1.986***	0.2317	1.961***	0.2470
$DIST$ Mean shifters				
age	0.002***	8.1e-04	0.002***	8.4e-04
inc2	0.053**	0.0240	0.081***	0.0262
inc3	0.148***	0.0349	0.182***	0.0379
parsize	–0.092***	0.0099	–0.091***	0.0104
weekend	–0.565***	0.0276	–0.570***	0.0272
q1	–0.015	0.0500	–0.026	0.0519
q2	–0.012	0.0402	–0.012	0.0421
q4	0.201***	0.0512	–0.218***	0.0548
d_warmorigin_1	–0.005	0.0504	–0.019	0.0535
d_warmorigin_2	–0.140***	0.0443	–0.145***	0.0452
d_warmorigin_3	–0.118***	0.0355	–0.148***	0.0382
d_warmorigin_4	0.102*	0.0594	0.129	0.0643
winter_sports	–0.119	0.0761	–0.077	0.0710
mou_trek_nat	–0.181***	0.0280	–0.194***	0.0271
rural	–0.040*	0.0242	–0.042	0.0262
aquatic	0.269***	0.0267	0.295***	0.0276
advent	0.067**	0.0281	0.063**	0.0285

r_TEMP Mean shifters				
age	0.012	0.0092	0.013	0.0091
inc2	0.1499	0.2771	0.230	0.2760
inc3	0.3260	0.3548	0.379	0.3512
parsize	–0.148	0.1035	–0.151	0.1020
weekend	–0.695***	0.260	–0.756***	0.2594
q1	–1.273**	0.5519	–1.245**	0.5460
q2	–0.756	0.6180	–0.713	0.6107
q4	–1.411**	0.5604	–1.404**	0.5550
d_warmorigin_1	2.018***	0.4255	1.958***	0.4159
d_warmorigin_2	1.037*	0.6222	0.938	0.6154
d_warmorigin_3	–2.909***	0.6144	–3.261***	0.6186
d_warmorigin_4	–0.292	0.5922	–0.325	0.5914
winter_sports	–0.539	0.4179	–0.703*	0.4195
mou_trek_nat	–0.688**	0.2772	–0.731***	0.2733
rural	–0.807***	0.2875	–0.794***	0.2857
aquatic	0.284	0.4052	0.232	0.3919
advent	1.453***	0.3104	1.463***	0.3042
Log L	–11,182.5		–11,179.3	
AIC	22,481.1		22,474.6	
Pseudo-R2	0.407		0.407	
N	6,661		6,661	

Table A6.- Parameter estimates for RPLc-ECM assuming Truncated Normal and Weibull distributions
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef.	SE
$REG1$	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_warmorigin_1	–0.024	0.0538
d_warmorigin_2	–0.142***	0.0455
d_warmorigin_3	–0.132***	0.0382
d_warmorigin_4	0.114*	0.0646
winter_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_warmorigin_1	1.988***	0.4302
d_warmorigin_2	1.022	0.6238
d_warmorigin_3	–3.029***	0.6229
d_warmorigin_4	–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 A7.- Parameter estimates for RPLc-ECM replacing *TCPI* by *r_TCPI*
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef.	SE
$REG1$	2.031***	0.3036
$REG2$	0.659**	0.3264
$REG3$	0.092	0.4087
$REG4$	0.250	0.3227
$REG5$	0.288	0.2776
$REG6$	–0.530*	0.3146
$DIST\_alt$	–0.366***	0.0401
r_TEMP	1.022	0.7751
$RAIN$	–0.250***	0.0716
$TCPI$	–0.019***	0.0060
TOU_SPOTS	0.096***	0.0093
NAT_PARKS	0.037***	0.0084
$SIZE\_NAT$	0.003***	9.5e-04
SKI_KM	0.001***	2.3e-04
$SKI\_KM * winter\_sports$	0.009***	5.4e-04
$COAST$	–1.393***	0.1832
$COAST * aquatic$	2.790***	0.1289
SD $DIST\_alt$	0.285***	0.0129
SD r_TEMP	2.308***	0.3449
Cov($DIST\_alt$, r_TEMP)	0.275***	0.1002
ϑ_1	0.362	0.3623
ϑ_2	0.963***	0.1448
ϑ_3	0.664***	0.2258
ϑ_4	0.251	1.3015
$DIST\_alt$ Mean shifters		
age	0.001***	5.8e-04
inc2	0.052***	0.0178
inc3	0.112***	0.0257
parsize	–0.061***	0.0071
weekend	–0.383***	0.0196
q1	–0.031	0.0365
q2	–0.001	0.0292
q4	–0.141***	0.0383
d_warmorigin_1	0.019	0.0374
d_warmorigin_2	–0.084***	0.0310
d_warmorigin_3	–0.077***	0.0255
d_warmorigin_4	0.092**	0.0436
winter_sports	–0.044	0.0488
mou_trek_nat	–0.129***	0.0195
rural	–0.029*	0.0179
aquatic	0.194***	0.0194
advent	0.051***	0.0198
r_TEMP Mean shifters		
age	0.011	0.0098
inc2	0.175	0.2921
inc3	0.314	0.3803
parsize	–0.205	0.1098
weekend	–0.929***	0.2740

q1	–0.070	0.5555
q2	–0.092	0.6186
q4	–0.396	0.5627
d_warmorigin_1	1.773***	0.4397
d_warmorigin_2	0.881	0.6243
d_warmorigin_3	–2.537***	0.6079
d_warmorigin_4	–0.643	0.6061
winter_sports	–0.258	0.4510
mou_trek_nat	–0.794***	0.2922
rural	–0.943***	0.2996
aquatic	0.820**	0.4177
advent	1.591***	0.3257
Log L	–11,198.8	
AIC	22,513.6	
Pseudo-R2	0.406	
N	6,661	

Table A8.- Parameter estimates for RPLc-ECM replacing *DIST* by *DIST_alt*
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef.	SE
$REG1$	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_warmorigin_2	–0.135***	0.0353
d_warmorigin_3	–0.127***	0.0302
d_warmorigin_4	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_warmorigin_1	2.067***	0.4116
d_warmorigin_3	–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 A9.- Parameter estimates for RPLc-ECM imposing restrictions on some taste shifters
*** p<0.01, ** p<0.05, * p<0.1

	RPLc-ECM	
Variable	Coef.	SE
$REG1$	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_warmorigin_1	–0.021	0.0546
d_warmorigin_2	–0.106**	0.0463
d_warmorigin_3	–0.126***	0.0389
d_warmorigin_4	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_warmorigin_1	1.747***	0.4222
d_warmorigin_2	1.178*	0.6045
d_warmorigin_3	–2.956***	0.6194
d_warmorigin_4	–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	6,661	

Table A10.- Parameter estimates for RPLc-ECM without the Balearic and the Canary Islands
*** p<0.01, ** p<0.05, * p<0.1

Chapter 3.- Understanding Marginal Rates of Substitution Among Holiday Destination Attributes: A Discrete Choice Experiment

David Boto-García

botodavid@uniovi.es

Department of Economics. University of Oviedo

Avenida del Cristo s/n, Oviedo (Asturias). Faculty of Business and Economics, Oviedo (Spain).

Abstract:

This paper studies experimentally the marginal rates of substitution and Willingness to Pay for holiday destination attributes. By means of a Discrete Choice Experiment, we specifically examine how much individuals are willing to pay for accommodation dwelling, mode of transport, travel time and length of stay. We estimate a Latent Class Model that allows us to control for taste heterogeneity based on sociodemographic characteristics. The welfare loss due to a tourism daily tax is also examined. Our results show that respondents place positive utility to travelling by plane, high quality accommodation and longer stays. Specifically, they are willing to pay about €170 for plane travelling with respect to the use of car, €120 for staying at 4-star hotel relative to an apartment, and €760 for a 10-day trip relative to a 3-day one. We show that a daily tax of €1 per person would produce a larger welfare loss in coastal destinations.

JEL codes: C91, D12

Keywords: *Discrete Choice Experiment, latent class, Willingness to Pay, vacation choice, travel preferences*

Acknowledgements:

I am extremely grateful to José Baños, Marta Escalonilla and Emma Zapico for their assistance during the experimental sessions. I also wish to thank Esteban García Canal, Leví Pérez, Rosa María Rodríguez, and Fermín Rodríguez for their help with the logistics and for proving us with the rooms needed in the different locations where the experiment was conducted.

I am grateful to my PhD advisors José Baños and Antonio Alvarez for helpful comments and suggestions. This work could not have been done without the help, support and suggestions received from Petr Mariel.

This paper also benefited from suggestions and comments by Luís Valdés, Eduardo del Valle, Federico Perali, Martina Menon, Alessandro Bucciol, Luca Zarri, Vittoria Levatti, Juan Prieto, Ana Rodríguez, Josepa Miquel-Florensa and María José Suárez during the experimental design stage. I also appreciate comments and suggestions from seminar participants at the Universities of Oviedo and Verona that helped in improving the context of the work.

Financial support from the Spanish Ministry of Education, Culture and Sport (FPU 16/00031) is gratefully acknowledged.

1. INTRODUCTION

A holiday trip is a multi-faceted decision that involves choosing several factors such as the mode of transport, the accommodation dwelling, or the length of stay. The choice of a particular bundle of trip-features requires individuals to make trade-offs between the attributes. Understanding preferences for vacation together with the substitution rates at which individuals are willing to forego a trip feature to gain in exchange another one constitutes an issue of economic relevance.

So far, the factors that pull tourists to holiday destinations have been largely studied (e.g. Nicolau and Más, 2006; Wong et al., 2017). However, studies based on revealed preferences have the weakness that the researcher only observes the *chosen* trip. That is, when modelling the choice of a holiday trip, we do not observe the full range of existing alternatives. This typically requires scholars to make some assumptions about the choice set from which individuals choose. Furthermore, choices in real life might be affected by several factors that are unobserved from the researcher perspective. We, instead, aim to study preferences for holiday trips through an experimental setting that controls for the characteristics of the non-chosen alternatives. A precise definition of the context and the environment in which decisions are made allows for a better identification of preferences.

We conduct a Discrete Choice Experiment (hereafter DCE) in which respondents are presented with a series of six hypothetical choice scenarios, each of them characterized by three types of destinations (coastal, urban, and nature-based) plus a 'none-of-them' option. Each alternative is defined by a set of exogenous well-defined attributes. Participants are required to choose their preferred option in each scenario. Based on that, we recover the associated marginal utilities that rationalize the data according to Random Utility Maximization Theory. Importantly, the repeated nature of the choice task has the advantage that tastes are identified based on choices from different combinations of attribute levels in each choice situation.

We estimate a Latent Class Model (LCM) that distinguishes groups of individuals with different tastes for the attributes. Preference heterogeneity is modelled as a function of sociodemographic characteristics such as gender, age, income, and education level. We derive marginal rates of substitution in the form of Willingness to Pay (WTP) estimates, which provide an economic valuation of the preferences for the attributes. Therefore, our model aims to explain how individuals trade attributes when choosing among different recreational sites.

A particular feature of our study is that our data comes from a sample of real live couples recruited from the general population of four cities in Northern Spain. We have several motivations to select real life couples as our subject pool. Since preferences for vacation features might depend on trip companions and the season of the year, we opted for framing the choice experiment in the context of a summer trip with their sentimental partner[88]. Partners have a common past, an expected future and are accustomed to make decisions on behalf of the other. In this way, from the respondents' perspective,

[88] This is because most holiday trips take place during the summer period and in dyads.

being confronted with a choice decision for a joint trip knowing that their partner is also making his own decision enhances the *salience* of the choice task, making it more in line with a real-life situation, and thereby reducing hypothetical bias[89]. Remarkably, both partners make their choices individually and separately. Therefore, we elicit their *individual* preferences, without prejudice of the potential exercise of altruism towards partner's preferences. This issue is discussed later. This, in our view, enhances the practical relevance of our results.

Our paper contributes to the empirical literature on preferences for vacation attributes by shedding light on how individuals are willing to trade one attribute by another. Although there are some scholars that study willingness to pay for coastal (e.g. Schuhmann et al., 2016), cultural (e.g. Figini and Vici, 2012) and nature-based recreation (e.g. Wuepper, 2017), these studies focus on specific features such as beach quality or the tenure of World Heritage status. We, instead, assess marginal rates of substitution for a more generic tourist trip that allows the respondent to choose among coastal, urban and nature-based tourism alternatives. A second relevant feature is that we allow for preference heterogeneity in the form of latent classes, thereby identifying groups of individuals with different preferences. In this sense, our study is similar to Chen et al. (2019), who also conduct a choice experiment for addressing preferences for alternative package tours. However, we depart from them in that rather than deriving point estimates of the willingness to pay for each class we use both a weighted and a conditional estimator that gives information about the distribution of the WTP in the sample. We also conduct a simulation exercise to explore the welfare loss associated with a tourism daily tax.

Our model identifies three classes of individuals based on their preferences. Results show that young females with university education and low income mainly focus on the length of the stay and total cost in their vacation choices. By contrast, elderly males with high income and non-university education also attach positive value to the accommodation quality and travelling by plane. However, respondents do not seem to give importance to travel time. Individuals are willing to pay, on average, €170 for travelling by plane with respect to the use of car, €120 for lodging at a 4-star hotel (relative to an apartment) and €760 for a 10-day trip (relative to a 3-day one). Nevertheless, the WTP estimates are found to be heterogeneous in the sample. We also show that a daily tax of €1 per person would produce a larger welfare loss in coastal destinations.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 describes the experimental setting and the choice experiment. This is followed by a description of the econometric modelling that includes a discussion on the state of art in discrete choice modelling. Section 5 outlines the model specification and reports the estimation results. Finally, Section 6 discusses the main findings and concludes.

[89] This falls within the discussion in the economic literature about the validity of DCE for eliciting preferences in an incentive-compatible way. Apart from monetary incentives, evidence by Vossler et al. (2012) points to consequentiality as a major factor to enhance truthful preference elicitation. We consider the recruitment and participation of the two members of a couple for a choice experiment that involves preferences for a joint trip to be a valid way to increase the realism of the task.

2. LITERATURE REVIEW

There is a growing body of literature in tourism research that uses experimental techniques for analysing cause-effect relationships. A recent review of the state of art of field and laboratory experiments in tourism can be found in Viglia and Dolnicar (2020). Among the different typologies, DCEs stand as the most applied. Participants in a DCE are asked to choose their preferred alternative from a given exogenous choice set with two or more alternatives (apart from the 'none of them' option), each one characterized by a set of well-defined attributes and levels. Hence, DCE are a useful technique for examining the trade-offs consumers make between alternatives depending on their attributes (Hoyos, 2010). In comparison to survey data, it offers the great advantage that the choice set and the environment in which decisions are made are known and controlled.

Louviere and Hensher (1982) and Louviere and Woodworth (1983) were among the first who employed DCE. Over time, they have been widely used in marketing (e.g. Sándor and Wedel, 2001), transportation research (e.g. Hess et al., 2007) or environmental valuation (e.g. Johnston et al., 2017). In the tourism literature, DCE are nowadays a valuable and commonly employed methodology for eliciting tourist's preferences.

Most empirical studies concerned about individual preferences for leisure recreation have focused on nature-based tourism. In this regard, the recreational demand literature has made important contributions on anglers and recreationists' preferences for destination choice using DCE. Specifically, scholars have studied how recreational site choice probabilities relate to amenities and facilities (Juutinen et al., 2011; De Valck et al., 2017), the willingness to pay for quality improvements (Wuepper, 2017; León et al., 2015) or the heterogeneity in preferences based on sociodemographic characteristics (Hearnes and Salinas, 2002; Chaminuka et al., 2012) and trip motivations (Shoji and Tsuge, 2015; Swait et al., 2020), among others. Furthermore, much of the methodological advances in choice modelling in terms of the econometric modelling of unobserved heterogeneity have benefited from this body of literature (Boxall and Adamowicz, 2002; Swait et al., 2020).

However, these studies mainly focus on preferences for day trips. Since the degree of involvement and the importance of the trade-offs to be made increase with the budget share the trip entails, our study falls better within the body of research aimed at understanding preferences for holiday destinations. In this regard, the most closely related works to ours are Huybers (2003), Grigolon et al. (2012), Van Cranenburgh et al. (2014), Oppewal et al. (2015), Van Cranenburgh (2018) and Chen et al. (2019). Their different scopes, sample characteristics, and study purposes makes it difficult to compare them. Because of this, we subsequently proceed to review each of them.

Huybers (2003) conducts a DCE to a sample of Australian residents with the purpose of addressing the drivers of short-break domestic tourists' destination choices. Travel time, expenditure per person, amenities, crowdedness, presence of event/festival and environment are the six attributes considered. According to their estimates, tourists devote great importance to crowdedness and the quality of amenities, whereas travel time is not found to be statistically significant.

Vacation preferences of Dutch students are analysed by Grigolon et al. (2012). The portfolio choice considers four types of destinations (groups of cities based on geographic proximity), trip duration (daytrips, short and long holidays), travel party (alone, with partner, with family, with friends), mode of transport (car, bus/train, regular airlines and low-fare airlines) and accommodation (hotel/rented apartment, hostel, camping and friend's/relative's house). Since their purpose is to explore students' preferences for low-fare airlines, only transport-specific attributes (mode, cost, travel time, time of the time, and time to get to the station/airport) were systematically varied. As such, preferences for transport are conditional on the remaining characteristics of the trip, which are freely chosen. However, a main limitation of their experiment is that costs other than transportation are not considered. These authors conclude that the mode of transport and transportation costs are the most important facets of the travel decision, followed by the duration of the trip and travel party.

A novel DCE to study how destination and experience information affect holiday choices is conducted by Oppewal et al. (2015). Participants are presented first with a set of eight holiday destinations (place names) and then a set of eight types of experiences, or vice versa. After that, choices are made from a reduced choice set with only four options characterized by transport and accommodation attributes. Their results indicate that early exposure to a type of attribute enhances its importance, being the effect larger for the destination place names than for type of experiences.

Van Cranenburgh et al. (2014) study vacation behaviour under high travel cost conditions. Their aim is to see whether a rise in travel costs (three times larger than expected) would exert an impact on travel destination choice. They use real-life destinations (cities, regions, and countries) in a DCE that pivots the choice tasks using information on respondents' true consideration set. Basically, the choice set is endogenously created from respondents' actual trip plans. They show that an increase in travel costs is more negatively valued for air travel than for road transport. Overall, vacationers exhibit diminishing marginal disutilities of travel costs. Using the same data, Van Cranenburgh (2018) disentangles the heterogeneity in tourists' sensitivity to travel costs using latent class analysis. Four classes of tourists are found with different attitudes towards cost rises. Whereas middle age and elderly tourists are highly inclined to change their vacation to closer locations, about 12% of respondents (those with relatively high incomes) are not willing to change their choices even under a high travel cost scenario.

Finally, Chen et al. (2019) analyse heterogeneity in preferences for package tours using a DCE directed to mainland Chinese outbound tourists. Alternative packages were described according to the availability of free time, the number of optional activities, the meals and attractions included, the type of flight and total cost. Participants are grouped in three classes, mainly based on their budgets. Direct flights are strongly preferred among the medium and high budget segment. USA and Europe are preferred relative to Australia for one class, whereas the opposite holds for the other. Strangely, respondents in one of the two classes attach positive utility to cost, possibly because they associate it with higher quality[90].

By contrast, other studies have focused on specific types of tourism activities (e.g. coastal, cultural, urban tourism). In doing so, the choice experiment is designed for specific contexts and directed to consumers of the type of tourism being analysed, thereby reducing the generalization of results. Schuhmann et al. (2016) study tourists' willingness to pay for coastal and marine resources in Barbados. They find that beach-front lodging together with beach width and cleanness are the two factors to which visitors attach great economic value. Similarly, Talpur et al. (2018) show that beach recreationists in Pakistan are willing to pay entrance fees if water and beach quality were improved.

For the case of urban tourism, Dellaert et al. (1995) examine Dutch tourists' choices of activity packages for a weekend in Paris, finding that shopping and sightseeing are the most preferred activities. Additionally, they show that night-time activities are chosen independently from daytime ones. Also for a sample of tourists in the Netherlands, Dellaert et al. (1997) conduct a DCE to study preferences for short city breaks. Alternative destinations are described in terms of amenities (shopping facilities, restaurants and bars), hotel features (location, price, quality) and the mode of transport to reach there (bus or train, travel time, cost). Through several sub-designs, the authors conclude that destination and transportation choices are made together in a portfolio combination so that the separate analysis of these two dimensions may be misleading. Figini and Vici (2012) investigate whether enhancing the off-season cultural supply of Rimini (Italy) can help smoothing the high seasonality of this destination. Results from a DCE directed to offseason visitors show that promoting cultural tourism is a valuable option for raising tourism revenues in the offseason.

Some other applications of DCE in the tourism literature include preferences for rural houses accommodation (Albadalejo-Pina and Díaz-Delfa, 2009) and hotels (Kim and Park, 2017; Masiero et al., 2019), tourists' attitudes towards eco-efficient destinations (Kelly et al., 2007) and policies aimed at facing global warming (Bujosa et al., 2018), tourists' preferences for improvements in transport infrastructures (Bimonte et al., 2016), the role of personal interactions in Tourism Information Offices (Araña et al., 2016), the value of destination image (Carballo et al., 2015), conflicting preferences between local residents and tourists (Concu and Atzeni, 2012), how terrorism affects the image profile of destinations (Araña and Leon, 2008), the economic valuation of cultural heritage (Choi et al., 2010), the effect of fear of flying on flight choice (Fleischer et al., 2012), or potential engagement in space tourism (Crouch et al., 2009).

Table 3.1 provides a summary of several studies in the tourism and recreational literature that have conducted DCE to answer different research questions. They are presented in chronological order.

[90] Growing attention has been paid to price sensitivity in tourists' choices. For the case of tourists from Kuala Lumpur travelling to Sidney, Morley (1994) shows that a 10% price increase in airfares and hotel tariffs reduce demand by 12% and 4%, respectively. Masiero and Nicolau (2012) explore this further by classifying tourists into segments based on their price sensitivity. As expected, price is a dissuasive factor, with trip motivations explaining part of the differences in price elasticities across respondents.

Reference	Research question	Methodology	Empirical results
Morley (1994)	Effect of prices on tourist's destination choice	DCE/ MNL[a]	A 10% decrease in airfare increases the number of tourists travelling from Kuala Lumpur to Sidney by 12%. A 5% increase in hotel tariffs lead to a 2% decrease in tourism demand.
Dellaert et al. (1995)	Urban Dutch tourists' choice of activity packages for a weekend in Paris	DCE/ MNL[a]	Sightseeing during different periods of the day, having a drink at night and shopping are the most valued activities to perform.
Dellaert et al. (1997)	Identification of the attributes that drive Dutch tourists' choices for short trips	DCE/ CBLM[b], Probit, NLM[c] and JLM[d]	City sights and shopping facilities are the most valued attributes, whereas travel costs and distance are not significant for explaining choices.
Boxall and Adamowicz (2002)	To examine the role of psychological factors as underlying drivers of taste heterogeneity for recreational site choice	DCE/ LCM[e]	Wilderness trippers seek parks with high chances of entry, while escapists and nature nuts prefer areas that impose restrictions to the number of visitors. Heterogeneity in preferences for recreation is found to be related to motivational constructs.
Hearne and Salinas (2002)	Tourists' preferences for infrastructure, use restrictions and conservation in national parks and protected areas in Costa Rica	DCE/ MNL[a]	Tourists appreciate improved infrastructure, low entrance fees and more information. Restrictions in the access to some trails are more demanded by foreign visitors. WTP for greater information is estimated to be greater for foreign tourists than for Costa Rican ones.
Huybers (2003)	Determinants of short-break holidays	DCE/ MNL[a] and NLM[b]	Destinations offering natural or 'mixed' attractions are preferred over those with cultural attractions. Potential visitors prefer moderately busy destinations rather than crowded ones.
Kelly et al. (2007)	Visitor's preferences for several hypothetical strategies intended to promote eco-efficiency in tourism destinations	DCE/ MNL[a]	Both overnight and same-day visitors are willing to pay a fee for bus service availability. Resorts with protected landscape and high levels of recycled waste are preferred.
Louviere et al. (2008)	Preferences for island holidays (in a study about optimal experimental design)	DCE/ HMNL[f] and MXLM[g]	Tourists attach positive utility to overseas destinations, travelling in the peak season, organized tours, being close to tourist attractions, the availability of swimming pools and beaches, and meal inclusion in the accommodation. Conversely, utility decreases with costs and travel time.
Araña and León (2008)	The short-term impact of September 11th on tourists' preferences for travelling to destinations in the Mediterranean and to the Canary Islands	DCE/ MNL[a]	The attacks produced a shift in the perceived image of some destinations. The Canary Islands gained popularity at the cost of Tunisia and Turkey, which experienced a drop in their destination brand value.
Albadalejo-Pina and Díaz-Delfa (2009)	Tourist's preferences for rural houses accommodation	DCE/ MXLM[g]	The appeal of rural houses mainly depends on their natural surroundings and their intrinsic characteristics. The higher the price, the lesser the probability of the rural house being chosen. Frequent visitors are less deterred by high prices than first time visitors.

Crouch et al. (2009)	Preferences for hypothetical alternatives of space tourism	DCE/ MXLM[g]	Respondents are heterogeneous in their price sensitivity. Elderly people are not prone to engage in any space tourism alternatives. Risk aversion together with conservatism cancels out the effect of higher income availability.
Choi et al. (2010)	An understanding of tourists' economic valuation of cultural heritage in Australia	DCE/ MXLM[g]	Respondents show a higher preference for temporary exhibitions and the diversity of cultural events. However, tourists are neutral to the presence of galleries, the provision of audiovisual effects in the displays or making the exhibitions to travel around the country.
Juutinen et al. (2011)	The complex trade-offs in preferences between recreational and ecological aspects for visitors to national parks	DCE/ MXLM[g] and LCM[e]	Biodiversity is the most valued attribute. A high number of visitors produces welfare losses, especially among foreign visitors. Interestingly, preferences over alternative park features differ by nationality.
Grigolon et al. (2012)	The factors that drive student's destination and transport mode choices	DCE/ MNL[a]	Students primarily focus on transport mode and the associated transportation costs. Their preferred accommodation is at friend's or relative's house. They also prefer to travel with friends than alone or with their family.
Figini and Vici (2012)	How the cultural supply of Rimini (Italy) can enhance offseason tourism	DCE/ MNL[a]	Three segments of tourists are identified (leisure, business and cultural). They would like some improvements in the destination such as environmentally friendly investments or the pedestrianization of the seaside avenue.
Fleischer et al. (2012)	The effect of fear of flying on travel mode choice.	DCE/ MXLM[g]	Home carriers, scheduled carriers and non-stop flights are the attributes that mostly compensate for the fear of flying. The price elasticities of demand are lower for those who are averse to flying.
Masiero and Nicolau (2012)	To classify tourists into segments based on their price sensitivities	DCE/ MXLM[g] + cluster analysis	Four groups of tourists are identified. Although price is a dissuasive attribute for three segments, for the remaining group high prices could are perceived as a quality signal.
Chaminuka et al. (2012)	Preferences for ecotourism in South Africa	DCE/MNP[h]	Tourists appreciate village tours and crafts markets, but they are not interested in village-based accommodations. Domestic and high-income tourists are willing to pay higher fees that the ones proposed in the survey.
Concu and Atzeni (2012)	Residents' and tourists' preferences regarding a set of reforms on coastal development and environmental protection	DCE/ MNL[a], MXLM[g], MXLM-ECM[i]	Increasing environmental conservation only produces welfare gains to those that do not earn income from tourism activities. Congestion affects negatively local residents, while tourists do not declare to be affected by congestion.
Van Cranenburgh et al. (2014)	Vacation behavior under high travel costs conditions	SP-off-RP CE[j]	Vacationers exhibit diminishing marginal disutility of travel costs. An increase in travel costs is more negatively valued for air travel than for car or train travel.

Shoji and Tsuge (2015)	Offseason tourists' preferences for nature-based tours in sub-frigid climate zones	DCE/ LCM[e]	Tourists' preferences for nature-based tours are heterogeneous: while some individuals attach utility to wildlife observation, others prefer skiing and snowshoeing.
Oppewal et al. (2015)	How information affects holiday destination choices, paying attention to whether early exposure to the geographical destination or to experience-type information affects choices	DCE/ MNL[a]	Introducing the destinations' name early or later matters more for the final outcome than does the experience-type information. Tourists first decide where to go and then think about what to do there.
Carballo et al. (2015)	The economic value of destination image	DCE/ MNL[a]	Tourists attach high importance to destination image when choosing between Mediterranean destinations, being the Balearic and the Canary Islands more valued than Turkey, the Greek Islands or Cyprus.
León et al. (2015)	Preferences for policies aimed at managing tourist congestion and improvements in ecosystem services in two National Parks in Colombia	DCE/ MXLM[g] and MM-MNL[k]	Coral reef restoration is the policy that would most increase satisfaction, followed by the restoration of coastal and sandy ecosystems. Tourists seem to be concerned about the welfare of local populations.
Araña et al. (2016)	Visitor's preferences for service designs at tourist information offices	DCE/ LCM[e]	Visitors appreciate information services directly received from personal interaction instead of trough automated processes.
Bimonte et al. (2016)	Tourists' preferences for distance, time and costs with special attention to the environmental impact of a new airport	DCE/ MXLM[g]	Tourists prefer an improvement in existing ground connections rather than developing a new airport. They attach high importance to the environmental damage of transport facilities.
Schuhmann et al. (2016)	Visitors' preferences and willingness-to-pay for coastal amenities and marine characteristics in Barbados	DCE/ MXLM[g] and LCM[e]	Visitors highly appreciate beach-front lodging. They also show a strong aversion to beach litter and narrow beaches.
Kim and Park (2017)	Hotel choice for business and leisure travelers, focusing on the relevance of cognitive, affective and sensory attributes	DCE/ MNL[a] and MXLM[g]	Leisure travelers are the most sensitive to price, whereas business travelers value more safety and relax.
Wuepper (2017)	The economic premium of World Heritage status for a national park in Germany	DCE/ MXLM[g] and GMNL[l]	National park status and chalk cliffs are the most valuable attributes for visitors. The higher the travel costs to reach the park, the higher the value attached to World Heritage status.
De Valck et al. (2017)	Preferences for outdoor recreation, paying special attention to the effect of distance	DCE/ MXLM[g]	Respondents attach high value to natural landscapes and tranquility. Travel distance produces great disutility.
Van Cranenburgh (2018)	Dutch vacationer's short-term responses to a hypothetical scenario in which transport costs tripled current ones	DCE/LCCA[m]	Four classes of vacationers are found, with class membership being explained by age and income. Three of them were inclined to change their behavior by means of changing the destination or reducing the total expenditure there. The other group would not change their behavior.

Bujosa et al. (2018)	Tourists' preferences for pro-environmental policies	DCE/ MXLM[g]	Tourists' willingness-to-pay for a specific program increases with both increases in the expected temperature and the likelihood of occurrence.
Koo et al. (2018)	How safety risk influences flight choice	DCE/ MNL[a], LCM[e] and MXLM[g]	Although travelers do not declare safety as a key factor in airline choice, when presented with safety information they consider it when choosing. There are two classes of travelers: those who prioritized price and those who prioritize safety.
Talpur et al. (2018)	Recreationists' willingness-to-pay for beach quality improvements and the role of payment vehicles	DCE/ MXLM[g]	Willingness-to-pay estimates are larger when using entrance fees alone instead of entrance fees plus travel costs. Welfare estimates are larger when including an explicit payment vehicle.
Chen et al. (2019)	Mainland Chinese outbound tourists' preferences for all-inclusive package tours	DCE/ LCM[e]	Two different classes of tourists with different preferences. Most respondents preferred international flights and less designated shopping, thereby exhibiting desires for safety, time saving and freedom.
Crouch et al. (2019)	How site attributes affect the choice of a host city in the international conventions market	DCE/MNL[a]	The accessibility of the city, the risk of disrupting events and the quality and range of accommodation emerge as the most relevant attributes for convention site selection.
Masiero et al. (2019)	Preferences towards hotel location	DCE/ MNL[a] and MXLMg	Vicinity to the metro station and waterfront location are two highly appreciated features for hotel choice. An additional star in the online rating is valued about $18 on average.
Swait et al. (2020)	The influence of goal pursuit on site choice for recreation, paying special attention to whether sensitivity to distance is mitigated by trip purpose	DCE/ H-LCM[n]	Goals and their importance mitigate the disutility of distance and travel costs. 'Relaxation' and 'spending time with family' are more important for this purpose than 'contact with nature' or 'knowledge of territory'.

Table 3.1. Summary of DCE in tourism economics literature by chronological order.

[a] Multinomial Logit Model (MNL) [also referred to as Conditional Logit Model (CLM)]
[b] Component-Based Logit Model (CBLM)
[c] Nested Logit Model (NLM)
[d] Joint Logit Model (JLM)
[e] Latent Class Model (LCM)
[f] Heteroskedastic Multinomial Logit Model (HMNL)
[g] Mixed Multinomial Logit Model (MXLM)
[h] Multinomial Probit Model (MNP)
[i] Mixed Multinomial Logit Model with Error-Components (MXLM-ECM)
[j] Stated Preference of Revealed Preference Choice Experiment (SP-off-RP)
[k] Mixed Normal Multinomial Logit Model (MM-MNL)
[l] Generalized Multinomial Logit Model (GMNL)
[m] Latent Class Cluster Analysis (LCCA)
[n] Hybrid Latent Class Model (H-LCM)

3. EXPERIMENTAL SETTING

In this section we first describe the experimental design. Second, we provide a brief discussion on hypothetical bias. Third, we outline the data collection procedure. Finally, we show some descriptive statistics.

3.1. Experimental Design

Attributes, levels, and variable coding

As it is well-known in the discrete-choice literature, destination characteristics (attributes), their corresponding levels (i.e. the value each attribute takes) and the number of alternatives that define each choice scenario play a critical role (Caussade et al., 2005). For deciding which attributes and how many levels to consider, we conducted a qualitative discussion focus group with some experts in tourism. This is common practice and a requirement for appropriate experimental design. In addition, a review of related studies also helped to define them.

There are many dimensions individuals consider when organizing a holiday trip. However, the greater the number of attributes, the higher the cognitive burden of completing the choice task (Carson et al., 1994). Therefore, the analyst needs to find a balance between making the choice task realistic and avoiding omitted variable bias, but without making it too much complex (Lancsar and Louviere, 2008).

The different alternatives are defined by the following five attributes: i) the time required to reach the destination (travel time), ii) the mode of transport, iii) the length of the stay, iv) the type of accommodation, and vi) the total cost (including both transport and lodging costs). These choice attributes and their respective levels are displayed in Table 3.2. The definition of the level values is based on previous studies and discussions carried out in the focus group. We tried to consider attributes that were salient to most respondents. We avoided including overlapping attributes such as distance (given mode of transport, travel time and total cost) to reduce inter-attribute correlation. Although related studies consider a larger number of attributes (8 in De Valck et al. (2017) and Chen et al. (2019), 16 in Keane and Wasi (2013) and 31 in Crouch et al. (2019)), we decided to keep them to a reduced number since Louviere et al. (2008), in the specific context of holiday choice, show that choice consistency decreases as the number of attributes increases[91]. We performed a Monte Carlo simulation exercise (available upon request) in which alternative designs with more attribute levels were examined.

[91] Indeed, in the study by Crouch et al. (2019) only 12 out of the 31 attributes were found to be statistically significant.

Attribute	Levels	Acronym
Travel time	Less than 2 hours Between 2-5 hours More than 5 hours	*shortTT* *medTT* *longTT*
Mode of transport	Car Bus or Train Plane	*car* *bustrain* *plane*
Length of the stay	3 days 7 days 10 days	*3days* *7days* *10days*
Accommodation site	Full private apartment 2-star hotel 4-star hotel	*apartment* *2starhotel* *4starhotel*
Total cost (per couple)	€200 €600 €1,000 €1,400	*Cost*

Table 3.2.- Attributes and levels

The cost is the only attribute that is treated as continuous and the one with the highest number of levels (4)[92]. Since it gathers both accommodation and travel costs, we consider four levels to reflect different possible combinations of accommodation and mode of transport. The specific monetary values were derived from existing market prices at the time of the data collection. The selection of the cost vector is not trivial since it can exert non-negligible effects on decision heuristics and willingness-to-pay (WTP) estimates. In this regard, Kragt (2013) shows that respondents are more sensitive to relative than to absolute cost differences. Glenk et al. (2019) document that WTP tend to be larger as the cost vector increases, mainly due to anchoring effects.

The rest of attributes are dummy coded (see Subsection 5.1). One critique of the Random Utility framework is its linearity assumption, which implies that, independently of the potential existence of preference heterogeneity, the marginal utility of an attribute is constant over its whole domain[93]. By using dummy variables, we allow the effects of travel time, mode of transport, length of stay and accommodation to be different depending on the level. The specific attribute levels chosen (Table 3.2) mimic in some way the ones used in Keane and Wasi (2013) for a choice experiment for generic holiday packages. They specifically used 3 and 5 hours for travel time, 7 and 12 days for length of stay, and 2-star and 4-star hotel for type of accommodation.

Number of alternatives per choice set

A relevant decision for the proper design of a DCE is the number of alternatives per choice set. Under utility maximization, the higher the options, the higher the probability

[92] It is common practise that the cost attribute has the largest number of levels (Scarpa and Rose, 2008).

[93] Although this assumption is due to econometric tractability, recent research has started to allow for further flexibility by specifying non-linear (in parameters) MNL and RPL models (e.g. Hensher et al., 2011). The advantages of breaking the common linearity assumption are discussed in Andersen et al. (2012).

for respondents to find a suitable alternative (Oehlmann et al., 2017). However, the level of complexity affects choice consistency (DeShazo and Fermo, 2002). Furthermore, providing respondents with a large choice set might make them feel overwhelmed, a phenomenon known as ‘choice overload’[94]. Therefore, to avoid cognitive burdens, in our DCE respondents are presented with three alternatives per choice task plus a ‘none of them’ option (also referred as *status quo*). This is done to avoid forcing respondents to choose any of the alternatives if none of them is enough attractive. This is standard practise in DCE in general (Lancsar et al., 2017) and in the context of holiday choice in particular (Huybers, 2003; Grigolon et al., 2012; Bimonte et al., 2016) to ensure realism. The choice of three alternatives is based on the results by Rolfe and Bennet (2009), who find that presenting respondents with three alternatives is better for avoiding serial non-participation than two[95]. Indeed, the tourism literature indicates that travellers normally consider less than five options (Karl et al., 2015).

Experimental design labelling

Instead of using real destinations, alternatives are presented as ‘types’ of destinations. This was done to eliminate potential prior knowledge biases due to established preferences or destination image that would affect attribute evaluations. The alternatives are labelled according to three different types of generic destinations: coastal, urban, and nature-based tourism. There are two reasons for this. First, the labelling can help respondents to evaluate the alternatives. Second, we specifically aim to assess generic preferences for types of destinations. In this sense, the environmental economics literature has shown that preferences depend not only on the attributes of the goods but also on the ‘label’ under which they are presented (Czajkowski and Hanley, 2009).

Number of choice tasks

Like Bimonte et al. (2016) and Chen et al. (2019), respondents are confronted with six different choice tasks, creating a panel structure of our data. Chung et al. (2011) indicate that six are the optimal number of choice tasks per individual.

Choice card display

The left-to-right top-to-bottom reading pattern makes that the way attributes and alternatives are displayed can significantly affect individual’s choices (see Shi et al. (2013) for eye-tracking evidence). The standard matrix display choice cards present respondents with one row per attribute and one column per alternative (Boxall and Adamowicz, 2002; Grigolon et al., 2012; Araña et al., 2016; Bujosa et al., 2018; Logar and Brouwer, 2018). However, recent studies have proposed a transposed matrix, with one row per alternative and one column per attribute (e.g. Windle and Rolfe, 2014).

[94] Oehlmann et al. (2017) report that as the number of choice tasks increases, the probability of respondents opting for the ‘none of them’ rises, possibly due to fatigue effects. In the tourism context, Thai and Yuksel (2017) show that participants that choose from larger choice sets are significantly less satisfied with their choices and less sure about having chosen the best option. Similar results are also reported by Park and Jang (2013).

[95] This refers to the situation in which respondents choose the non-choice option for all choice situations (see Van Haefen et al., 2005).

Sándorf et al. (2018) test the effect of matrix display on the decision rule. They show that answers from the standard display condition are more compatible with attribute-based heuristics like Random Regret Minimization (RRM), whereas the transposed matrix is more appropriate for alternative-based heuristics like Random Utility Maximization (RUM). Therefore, they recommend presenting the alternatives matrix in the transposed form (i.e. one row per alternative and one column per attribute) for a DCE to be consistent with RUM. Based on this, choice cards are displayed in the transposed form to facilitate row wise comparison.

Context and framing

Since preferences for a trip might change depending on the season, it seems important before the series of choice tasks to describe a specific vacation context. Although hypothetical, we aim to make the portfolio choice as realistic as possible. Participants were required to imagine they will have 15 holiday days with partner during any period in the summer season (June-September) and plan to go on holidays together. Although for different study purposes, this contextualization is like the ones made by Thai and Yuksel (2017) and Mathews et al. (2017).

However, it is not only the context in which decisions are made but also the way in which the attributes are presented (framing effects). Studies by Kragt and Bennett (2012) and Logan and Brouwer (2018) have shown that the description of choice alternatives and attribute levels affect welfare measures. Respondents were told that the presented cost involves the total cost of accommodation (breakfast included) and transportation, and the travel time refers to the transit time between departure from home and arrival at the destination[96].

Importantly, the experimental setting focuses on individual tourism preferences for a trip with their sentimental partner[97]. In the choice task description, we emphasized that in making their individual decision they have full freedom to choose their preferred option. This was explicitly intended to identify their *individual preferences* rather than the *couple's* ones. Furthermore, since children usually exert great influence in family holiday decisions (Thornton et al., 1997; Nickerson and Jurowski, 2001), we also highlighted that they have to choose a destination for a couple trip in which neither children nor relatives are allowed to participate. The detailed instructions are provided in Annex 1.

Pilot test

As recommended by the DCE literature, we pre-tested the questionnaire, the choice task, and the attribute levels by a pilot study. It was conducted in February 2019. A total of 17 couples (i.e. 34 individuals) participated, being their participation not rewarded. It allowed us to ensure that the questions and the tasks were presented in a comprehensible way. Moreover, we checked the degree of complexity of the full experimental task by asking

[96] Although some attempts have been done to incorporate visual information, we keep a verbal description of the choice situation since visualizing information procedures reduce choice consistency and increase the time needed to reach decisions (Eppink et al., 2019).
[97] Travelling in couples has been shown to enhance couple's cohesion, flexibility, and functioning (Shahvali et al., 2020).

subjects to report the degree of difficulty encountered in answering, as done by Sever and Verbic (2018).

Experimental design

A D-efficient design was generated in NGENE (Choice Metrics, 2012) to find the optimal design among all the 324 possible designs (3x3x3x3x4). The priors used were obtained from the pilot study. In doing so, we imposed some constraints on the attribute level combinations to avoid dominant alternatives (i.e. cases in which one destination is clearly superior to the rest). This follows common practice (e.g. Juutinen et al., 2011). We refer the reader to Annex 2 for a discussion about efficient design theory and the importance of avoiding dominant alternatives. There we also report the NGENE code used. The D-efficient design generated 18 choice tasks separated into three blocks. Individuals were randomly assigned to one block so that they only completed the six choice tasks in that block. An example of a choice card translated into English is presented in Figure 3.1. The 18 choice tasks are available from the authors upon request.

Bear in mind that money expended on the trip will reduce the budget available for other purposes

BLOCK 2.- CHOICE CARD 3

	TRAVEL TIME	MODE OF TRANSPORT	LENGTH OF THE STAY	ACCOMMODATION	COST	MY CHOICE	MY PARTNER'S CHOICE
Option A: Coastal destination (sun and beach tourism)	Between 2 and 5 h	Bus or Train	3 days	4-star hotel	600 €/ couple		
Option B: Big city destination	Between 2 and 5 h	Plane	10 days	4-star hotel	1400 €/ couple		
Option C: Nature-based destination	More than 5 h	Car	7 days	2-star hotel	1000 €/ couple		
NONE OF THEM							

If you do not like any of the alternatives, recall that you have the option to select "NONE OF THEM"

Figure 3.1.- Example of the DCE

The data were collected in face-to-face interview and the choice task was presented to respondents in paper format in a randomized order (Börger, 2016; Oehlmann et al., 2017)[98]. Finally, all the choice cards included two reminders for i) the household budget

[98] Although some scholars have acknowledged the necessity of alternatives being randomly placed within the choice card to avoid left-right bias (e.g. Thai and Yuksel, 2017), options were always presented in the same order (coastal, urban, nature-based) to make respondents easier to find each option within the card.

constraint[99], and ii) if they do not like any of the three options, they could select 'none of them' (see Figure 3.1). This practise follows Logar and Brauwer (2018) and Czajkowski et al. (2019).

3.2. Hypothetical bias in choice experiments

One of the main criticisms about choice experiments is that it is subject to hypothetical bias (see Hensher (2010) for an overview). Hypothetical bias refers to the difference between how people think, feel, and behave, and what people say they are thinking, feeling and behaving. In our context, individuals may opt for an option in the lab whereas they choose another one in a real situation.

Several procedures have been proposed to deal with hypothetical bias. One of the most used is the so-called 'cheap talk', which consists of encouraging respondents about the importance of the research purposes ex ante, and that the results would make an impact on the goods and services available to them in real markets (Vossler et al., 2012). The validity of this method is mixed (Landry and List, 2007; Blumenschein et al., 2008). By contrast, other scholars opt for asking respondents their degree of certainty with the choices made (Beck et al., 2013). This imposes the difficulty of defining the threshold that separates certain from uncertain choices and potential problems of endogeneity. To deal with the latter, novel econometric methods that jointly model choices and the reported uncertainty have been recently developed (e.g. Beck et al., 2016).

In a similar study to ours, Van Cranenburgh et al. (2014) pivot the choice tasks around the respondents' true consideration set in real life (i.e. the self-declared alternatives that they have in their minds for a coming trip) to reduce hypothetical bias. By doing so, the choice set is adjusted to present respondents with alternatives that are truly relevant. However, their approach has some limitations[100]. We disregard their experimental design because we are interested in preferences over generic attributes and types of destinations rather over specific destinations, where unobserved heterogeneity plays a major role.

A meta-analysis conducted by List and Gallet (2001) shows that hypothetical bias is minimized when individuals are required to make decisions for private goods and to perform familiar tasks. Furthermore, choice experiments and surveys conducted personally in lab experiments have been shown to be more reliable and to produce superior data quality than those obtained through internet-based methods (Haener et al., 2001). In this respect, hypothetical bias seems to be reduced in our study since respondents perform a face-to-face choice task for a good that is familiar to them.

---

[99] The specific wording was: *Bear in mind that money expended on the trip will reduce the budget available for other purposes.*

[100] First, their approach does not allow them to construct an efficient design but a random one, which may produce important problems of dominance and implausible combinations of attributes. Second, only participants that intend to take a vacation in the near future are allowed to participate, this way leading to a potential self-selection effect. Third, their experiment requires respondents to make a total of fourteen choices (six in the RP part and eight in the SP one), which might lead to fatigue effects.

In our view, hypothetical bias is an issue that will always be present to a certain extent in a stated preference study. Even the most sophisticated econometric model that addresses choice uncertainty relies on stated answers. Therefore, although possibly not perfectly, a 'cheap talk' and a follow up survey together with presenting the choice task in a friendly manner could be valuable ways to minimize hypothetical bias[101].

3.3. Data collection

The DCE was conducted together with a parallel study on intra-household decision making. We recruited a fairly representative sample of stablished couples over 18, no matter whether they were married or whether they had children. Couples were recruited through flyers, brochures, social networks, and word-of-mouth from four cities in the North of Spain (Oviedo, Gijón, Avilés and Bilbao). The data were collected from couples to allow for the analysis of bargaining choice process that should lead to consensual holiday destination choice that is not included in this work. In the announcements we just stated that we were looking for stable couples to participate in a research study for a better understanding of preferences. We also indicated that each participant would receive a fixed amount of money for participation (€10) plus a variable sum of money depending on choices from a Public Good Game (not to be analysed here). We gave respondents this monetary incentive to reward them for their time. Everyone was paid individually and anonymously at the end of each experimental session. Our recruitment procedure follows the ones by Munro and Popov (2013), Abdellaoui et al. (2013) and Cochard et al. (2016).

The experimental protocol is as follows. Upon arrival at the lab, subjects were given a random ID code that identified the couple and the individual within it (e.g. 44B). They were asked to switch off their mobiles phones in order to avoid interruptions. Before the experiment started, participants were gathered in a large room and informed about two important issues. First, they were guaranteed that their answers to the questionnaire would remain unknown to their partner. In doing so we follow Ashraf (2009), who pointed to the necessity of hiding participant's answers to their partners as to maintain 'plausible deniability' when couples exited the experiment. Second, we conducted a brief introductory talk in which subjects were told about the aim of the study, the structure of the experimental session and that the information collected will only be used for research purposes. Neither in the recruitment process not during the introductory talk we told participants they were about to take part in an 'experiment'. This was done for minimizing the potential hypothetical bias, in line with the issues raised in the previous subsection.

Subsequently, participants were separated into two different rooms (one member of the couple in each one, at random). They were given the set of six portfolio choices described above together with an example for purposes of illustration. Confronted with the four-option choice set, they were required to indicate their preferred option for a trip with their partner.

[101] In line with this, the study by Haener et al. (2001) compares the predictive performance of SP, RP and a mixture of SP and RP data collection, concluding that a well-implemented SP survey can even be more reliable than a less well-implemented RP survey.

While completing the choice tasks participants could not communicate and therefore know the decisions their partner (spouse) was actually making. They had to make their choices alone as if they had the full power to decide where to go. This is similar to Huybers (2003), in which respondents were also required to make their choices assuming they were the main decision maker. As indicated before, the objective was to identify their individual preferences. In addition, everyone was seated at their own table, and tables were sufficiently separated to avoid anyone seeing others' answers. Each respondent was randomly allocated to one of the three versions of the DCE (blocks). Importantly, the two members of the couple were given the same sequence order.

After completing the DCE, respondents were asked to individually fill a questionnaire. The questionnaire comprises different blocks (see Annex 3). First, they were required to report a set of sociodemographic characteristics. Specifically, they report their gender, age, education, civil status, labor situation, net individual income (per month), nationality, number of children (if any) and (if so) their age. A second set of questions collects information about the relationship with their partner, not to be used in the current study. The third block contains questions about travel experience, frequency of travelling, general preferences for destinations and expenditure. A lack of interest in travelling could be an indicator of lower involvement in travel destination decision-making.

3.4. Descriptive statistics

Prior to the data collection, we conducted Monte Carlo simulation exercises with the purpose of identifying the minimum sample size for parameter identification subject to our budget constraint. In Chapter 4 we present estimates using 6 different sample sizes (N=30, 60, 120, 180, 240 and 300). Our estimates indicate that we needed at minimum 240 respondents for reliable parameter identification.

In total, we conducted 10 sessions with approximately 13 couples per session, on average. Due to time schedule constraints, some couples completed the choice tasks individually or in reduced groups, always separating the two partners in two large rooms[102]. Our DCE was successfully completed by 262 individuals. Therefore, a total of 131 couples participated in the study, of which 3 were same-sex couples. In this way, our sample size is above the minimum for reliable parameter identification (240) and larger than 200 – the archetypal sample size in DCE (Rose and Bliemer, 2009). The recruitment of couples is more complex than that of single individuals since the two partners are required to participate. Therefore, experiments with couples rarely result in large sample sizes. In this vein, our sample size is in-between the ones of related experimental studies about couple's preferences and decision-making[103].

[102] We conducted four sessions in Oviedo, four in Gijón, one in Avilés and one in Bilbao.

[103] Ordered from the lowest to the highest, the following sample sizes have been used: 20 couples from Hamburg (Germany) in Görges (2015), 22 couples from Jena (Germany) in de Palma et al. (2011), 31 couples from London in Munro and Popov (2013), 45 couples from Tobago in Beharry-Borg et al. (2009), 64 couples from Toulouse in Cochard et al. (2018), 76 couples from Norwich (UK) in Bateman and Munro (2005), 80 couples from Torino and Vincenza (Italy) in Rungie et al. (2014), 81 couples from Paris (France) in Couprie et al. (2020), 87 couples from Mannheim (Germany) and 69 couples from Toulouse (France) in Beblo et al. (2015), 94 couples from Utah (USA) in Gnagey et al. (2018), 95 couples from Mannheim

Before moving to the data analysis, an important issue deserves mention. Many studies recruit participants for experiments from pre-recruited panels of respondents. In this sense, most experimental studies published in top journals in economics make use of ORSEE recruitment system (e.g. Maniadis et al. 2014; Kneeland, 2015)[104]. This practise has been criticized as these 'professional respondents' tend to participate in experimental settings on a regular basis, rushing through the questionnaire and the tasks without paying the required attention to the information provided (see Dennis (2001) for a discussion). Additionally, participants from ORSEE tend to be double-selected: first, engagement into the system; second, selection into a given experiment from the recruited subject pool. In the context of studying risk preferences, Heinrich and Mayrhofer (2018) indicate their sample is composed only of males because there are more males in ORSEE subject pool. We, instead, opted for recruiting participants directly from the population. Hence, although there is an inherent potential self-selection problem[105], we avoid some of the limitations of ORSEE. In our sample, most respondents (90%) had never taken part in an experiment before. This eliminates potential confounding behaviour due to previous experience in similar experiments.

Table 3.3 presents summary statistics of the sample along with the description of the variables.

(Germany) in Beblo and Beninger (2017), 100 couples (students) from Guiyang (China) in He et al. (2012), 100 couples from Toulouse in Cochard et al. (2016), 101 couples from several villages of Guizhou (China) in Carlsson et al. (2012), 110 (monogamous) couples from Nigeria in Barr et al. (2019), 117 couples from Guizhou (China) in Carlsson et al. (2013), 119 couples from Bilbao (Spain) in Mariel et al. (2018), 130 couples from Paris (France) in Abdellaoui et al. (2013), 142 couples from 3 towns in Kenya in Robinson (2012), 142 couples from East Anglia (UK) in Bateman and Munro (2009), 146 couples from the Philippines in Ashraf (2009), 164 couples from 13 villages in Gansu (China) in Yang and Carlsson (2016), 169 couples from the USA in Boldt and Arora (2017) and 183 couples from two regions in India in Castilla (2019).

[104] See Greiner (2015) for an in-depth discussion of the software and procedures.

[105] For the case of couples participating in experiments, samples might be biased in favour of those with healthy and stable relationships, for whom being the object of an experiment does not matter.

Variables	Description	Mean	SD	Min	Max
female	=1 if female	0.508	0.501	0	1
age	age (in years)	32.40	13.99	18	89
educ1	=1 if primary education	0.076	0.266	0	1
educ2	=1 if secondary education	0.313	0.465	0	1
educ3	=1 if high education	0.611	0.489	0	1
working	=1 if employed	0.538	0.499	0	1
unempl	=1 if unemployed	0.061	0.240	0	1
inactive	=1 if inactive (housewife or retired)	0.076	0.266	0	1
student	=1 if student	0.324	0.469	0	1
whours	=0 if respondent does not work; =1 if works<15 h per week; =2 if works between 15-30 h per week; =3 if works between 30-40 h per week; =4 if works >40 h per week	1.672	1.561	0	4
income0	=1 if net monthly income=0	0.279	0.449	0	1
income1	=1 if net monthly income <€500	0.134	0.341	0	1
income2	=1 if net monthly income between €500 and €1,500	0.302	0.460	0	1
income3	=1 if net monthly income between €1,500 and €2,500	0.221	0.416	0	1
income4	=1 if net monthly income >€2,500	0.053	0.225	0	1
income	=0 if net monthly income=0; =1 if net monthly income <€500; =2 if net monthly income between €500 and €1,500; =3 if net monthly income between €1,500 and €2,500; =4 if net monthly income >€2,500	1.614	1.259	0	4
married	=1 if married	0.282	0.451	0	1
children	=1 if has children	0.248	0.433	0	1
numchildren	Number of children	0.420	0.773	0	3
spanish	=1 if Spanish	0.977	0.150	0	1
rel_less5	=1 if relationship <5 years	0.538	0.499	0	1
rel_5_15	=1 if relationship between 5 and 15 years	0.248	0.433	0	1
rel_15_25	=1 if relationship between 15-25 years	0.068	0.253	0	1
rel_more25	=1 if relationship >25 years	0.145	0.353	0	1
likeshol	=1 if likes going on holidays	0.966	0.182	0	1
travelled	=1 if travelled for leisure purposes in last 12 months (at least one overnight stay)	0.870	0.337	0	1
domtrips	=1 if prefers to travel domestically (vs abroad)	0.240	0.428	0	1
preftravcou	=1 if prefers to travel exclusively with partner when going on holidays (as opposed to alone or with friends/relatives)	0.523	0.500	0	1
nevertrav	=1 if never or hardly ever goes on holidays	0.115	0.319	0	1
onceytrav	=1 if goes on holidays once a year	0.294	0.456	0	1
twiceytrav	=1 if goes on holidays twice a year	0.359	0.481	0	1
morethree	=1 if goes on holidays three times a year or more	0.233	0.423	0	1
prefcoastal	=1 if in general prefers coastal destinations	0.328	0.470	0	1
prefurban	=1 if in general prefers urban destinations	0.279	0.449	0	1
prefnat	=1 if in general prefers nature-based destinations	0.103	0.305	0	1
noclearpref	=1 if respondent does not have a clear preference for any type of destination	0.275	0.447	0	1
partbefore	=1 if participated in a similar study before	0.103	0.305	0	1

Table 3.3.- Summary statistics of the sample (N=262)

Starting with sociodemographic features, the average age of respondents is 32.4 years, ranging from 18 to 89. About 60% have university education, being half of the sample (53.8%) currently employed. In terms of after-tax monthly income, the sample is quite balanced, with 27% having no income, 13% earning less than €500 per month, 30%

receiving between €500 and €1,500, 22% earning between €1,500 and €2,500 and only 5% earning more than €2,500. Around 28% of respondents are married and 25% have children. The vast majority are Spanish (97%). Regarding the length of the relationship with their current partners, more than half of participants (53%) have been together for less than 5 years, while a non-negligible 15% are in a relationship for more than 25 years ago.

Although our sample is not perfectly representative of the population in the four cities considered as a whole, their characteristics are reasonably well-aligned with the subpopulation of interest in our study context: those who participate in tourism activities. To examine this, we obtained microdata from the Residents Travel Survey conducted on a monthly basis by the Spanish National Statistics Institute. This dataset provides information on trips undertaken by Spanish residents, either within Spain or abroad[106]. We specifically chose the period June-September 2018 (29,516 respondents) for the comparison. We kept those who live in Asturias (1,025 respondents)[107]. Compared to these data, the educated and young people are slightly overrepresented and married people underrepresented in our sample[108].

Concerning trip preferences, most participants declare to like going on holidays (96%), and 87% went on a leisure trip with their partner at least once in the last 12 months. Only 24% of the sample state they prefer travelling domestically than abroad, whereas 52% declare to prefer travelling with their partner in comparison to alone or with friends/relatives. As for travel frequency, 11% state they never or hardly ever go on holidays, 29% say they go on holidays once a year, 36% report they travel for leisure twice a year and 23% indicate they usually take a trip three times a year or more. Finally, 32% prefer coastal locations. Urban destinations are preferred by 28%, while nature-based destinations are the main option for only 10% of respondents. The remaining 27% do not exhibit a clear preference for either of the three.

These data on trip preferences and travel frequency serve as preliminary evidence on the validity of our experimental design. Note that i) both preferences for types of destinations and travel frequency are quite balanced, ii) most people like to travel and do it regularly, and iii) their partner is the preferred travel companion. Since we aim to explore the marginal utilities for the attributes, if a great share of the sample exhibited a strong preference for one type of destination or were no interested in travelling, that would limit the trade-offs we intend to capture.

Each respondent answered six different choice cards with four alternatives each. Therefore, the total number of observations is $6 \, x \, 262 = 1{,}572$. Table 3.4 presents the number of individuals in each block and how many respondents were allocated to each of the five orders. As shown, the design is strongly balanced.

[106] The sample is obtained from the Continuous Household Survey also conducted by INE, which uses a stratified two-stage sampling procedure from the 2011 Population Census. The sample is representative of the Spanish population.

[107] Average age is 51 years old, 98% are Spanish and about 52% are married. In terms of education, 24% of this subsample has primary studies, 25% has secondary education and 50% has completed higher education. Concerning labour status, 57% of respondents are employed and 7% are unemployed.

[108] In Bimonte et al. (2016), the sample is also relatively young (mean age=41 years old) and highly educated (68% are graduates). Similarly, in De Valck et al. (2017) about half of respondents have high education.

Choice card order	Block 1	Block 2	Block 3	Total
123456	96 (6.1%)	108 (6.8%)	84 (5.3%)	288 (18.3%)
654321	108 (6.8%)	96 (6.1%)	156 (9.9%)	264 (16.8%)
321654	72 (4.6%)	84 (5.3%)	108 (6.8%)	348 (22.1%)
561243	108 (6.8%)	108 (6.8%)	96 (6.1%)	312 (19.8%)
436521	108 (6.8%)	144 (9.1%)	96 (6.1%)	360 (22.9%)
	492 (31.3%)	540 (34.3%)	540 (34.3%)	1,572 (100%)

Table 3.4.- Individuals assigned to each block and choice card order

Figure 3.2 plots a bar graph of the vacation choices made by respondents considering the 6 choice situations faced by each individual (N=1,572). This plot is merely descriptive, since we do not consider the attributes that describe each alternative. Coastal is chosen 35% of the choice tasks, closely followed by urban tourism (32%). Nature-based destinations are selected in 20% of the cases while 'none of them' is chosen in about 11% of the choice cards.

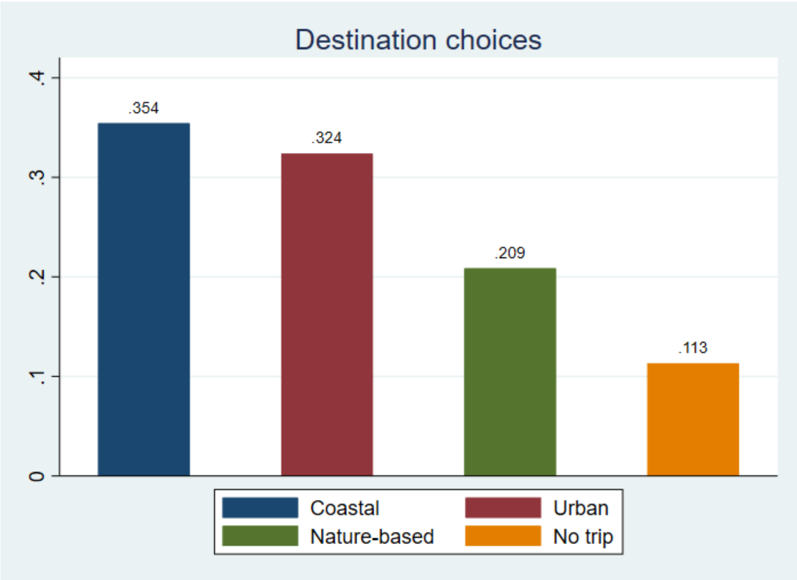


Figure 3.2.- Distribution of destination choices

Prior to formal modelling, we conducted some checks to our data. Specifically, we first examined the share of 'non-participants' in our sample. In environmental studies, these individuals are normally labelled as 'protesters'. Out of the 262 participants, only six individuals (2.2%) opted for 'none of them' in their sequence of choices. Accordingly, non-participation is not a relevant issue in our data[109].

Second, we studied the *unconditional* relationship between attribute levels and choices. The third column of Table 3.5 presents the percentage of choice alternatives (262 individuals times four choice alternatives times six choice situations=6,288 choice alternatives) in which each attribute level was present. The fourth column further reports the proportion of chosen alternatives characterized by each attribute level while the fifth

[109] Based on the questionnaire data, we document that these individuals exhibit a reduced interest in travelling, thereby being more likely to be non-participants than protesters.

column shows the correlation between the presence of the attribute level and whether the alternative was chosen. The percentage of choice alternatives with a given attribute sum 75% because the remaining 25% correspond to the 'none-of-them' option, which has no attributes.

Attribute	Level	% of choice alternatives with level k (out of 6,288)	% chosen alternatives with attribute level k (out of 1,572)	Corr(atr.level, choice)
Travel time	*shortTT*	20.7	24.6	0.056
	medTT	29.3	31.4	0.027
	longTT	25.0	32.5	0.100
Mode of transport	*car*	25.0	23.4	-0.021
	bustrain	25.0	24.7	-0.003
	plane	25.0	40.5	0.207
Length of stay	*3days*	27.6	18.2	-0.121
	7days	19.5	26.7	0.104
	10days	27.8	43.7	0.204
Accommodation	*apartment*	24.8	33.3	0.044
	2starhotel	25.1	26.7	0.021
	4starhotel	25.0	33.7	0.117
Cost	€200	4.0	7.0	0.086
	€600	32.0	43.1	0.172
	€1,000	34.7	30.4	-0.052
	€1,400	4.1	5.1	0.028

Table 3.5.- Distribution of choices and attribute levels

Respondents seem to prefer those alternatives in which the trip involves travelling by plane, staying 10 days and with a reduced cost. From the fourth column, note that the alternative with the highest cost is chosen only 5.1% of the times. This share lies into 5-10% interval, proposed as a 'rule of thumb' by Mørkbak et al. (2010) to detect suspicious behaviour[110]. This, together with the low share of non-participation appears to indicate that the cost vector is properly defined. Preferences for travel time are quite balanced, with long travel times (more than five hours) being relatively more frequently chosen than shorter ones. Nevertheless, this analysis is only descriptive since it does not condition on the remaining attribute levels in each alternative. This is what we do in the following sections.

[110] If a larger share of individuals choose the most expensive option, that could mean that the cost vector is not properly defined (possibly too low levels).

4. ECONOMETRIC MODELLING

4.1. Random Utility Maximization

The theoretical basis for Discrete Choice Modelling can be found in Lancaster's Theory of Value (Lancaster, 1966). According to this framework, individuals derive utility from the different attributes of goods, so that different characteristics lead to different levels of utility. That is, individuals demand *characteristics* rather than goods themselves[111].

The traditional way to model discrete choices follows Random Utility Maximization (hereafter RUM), originally proposed by Thurnstone (1927) and later further characterized by McFadden (1974). In line with microeconomic theory, given a choice set with a number of alternatives, agents choose the option that maximizes their utility conditional on the existing constraints (mainly the budget constraint). A key assumption here is that the quantities of the good are fixed to the unity so that individuals choose for consuming only one unit of the good at a time (Hanemann, 1984). The different alternatives are *mutually exclusive* and regarded as close substitutes for the commodity.

The true underlying decision process is unknown, but utility maximization is likely to be a valid approximation. Consistent with Lancaster's characteristics model, the econometrician links observations of actual choices to the attributes that characterize the available alternatives. However, individuals do not only have different tastes over the attributes, but they might also exhibit random behavior (i.e. they do not necessarily behave in the same way faced with the same choice situation). Put another way, part of the underlying utility is unobserved. For this reason, RUM postulates that the underlying utility function is the sum of a systematic (V_{ijt}) and a random component (ε_{ijt}) so that:

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} \tag{3.1}$$

The random component ε_{ijt} is assumed to be independent and identically distributed across individuals (i), alternatives (j) and choice situations (t). The systematic component (V_{ijt}) is usually assumed to be a linear-in-parameters function of the K attributes that describe each alternative j in each choice situation t (X_{kijt}), the cost (price) of the choice alternative ($Cost_{ijt}$) and their associated vector of parameters (β_k and β_{cost}) to be estimated. This cost gathers the purchase price of alternative j for its bundle of characteristics (X_{kijt}). Therefore, the systematic component is given by:

$$V_{ijt} = \sum_{k=1}^{K} \beta_k X_{kijt} - \beta_{cost} Cost_{ijt} \tag{3.2}$$

This is the *conditional on tastes* indirect utility function[112]. Consistent with utility maximization, individuals choose the alternative j that produces the highest level of utility

[111] See Hausman and Wise (1978) for an earlier empirical application.

[112] The individual maximizes utility given choice characteristics subject to the budget constraint. When the choice set is restricted to a compact set, utility is defined in the *Lancasterian* sense and quantities are fixed to the unity, McFadden (1981, p. 206-210) proofs that the indirect utility function can be expressed independently of income. As highlighted by McFadden (1974, p. 114, footnote 6), income cannot be specified alone in the utility function because being invariant across choice alternatives and therefore its corresponding coefficient would not be identifiable (see Haneman (1984, p. 557)). Notwithstanding this,

so that $U_{ijt} > U_{imt} \ \forall j \neq m$. Remarkably, it is implicitly assumed that the characteristics of the non-chosen alternatives have no influence on the utility of the chosen alternative (weak complementarity).

Although RUM has been the toolkit for discrete choice modelling for more than 40 years, in the last decade it has started to be criticized. RUM infers preferences from observed choices under a utility maximizing framework, but treating the underlying decision making as a 'black box'. Following the development of behavioral economics, some scholars have argued that the assumption that individuals are able to fully evaluate the attributes of a given number of alternatives (especially when this number is large) is not realistic. RUM assumes compensatory behavior so that a low level of a given characteristic can be fully compensated by a high level of another, which might not be the case. For instance, individuals may always choose the cheapest alternative independently of the remaining attributes. This phenomenon, known as lexicography, is only consistent with utility maximization when it reflects true preferences (i.e. zero marginal utilities for the remaining attributes). However, lexicographic behavior might emerge because of choice complexity by which individuals adopt simplifying decision-making heuristics. Earlier research on this includes Heiner (1983) and de Palma et al. (1994).

Expertise and previous experience in a choice domain and time pressure are other factors that affect the likelihood of adopting lexicographic processing (Dieckmann et al., 2009). In the tourist context, Li et al. (2017) examine the potential existence of non-compensatory choice strategies for tourists' choice of destination. The way individuals evaluate alternatives in a choice task can also change over time (i.e. Value Learning). When faced with repeated choice decisions, individuals appear to 'learn' the optimal way to choose (McNair et al., 2012). This tends to translate into more random choices at the beginning of choice experiments (Czajkowski et al., 2014).

Elimination by aspects (Tversky, 1972), by which individuals eliminate alternatives that do not provide a certain minimum level for a given attribute, is another of these heuristics. This aspect seems to be particularly relevant in tourism since a large body of research shows that individuals follow a funnel-like process when choosing where to recreate (e.g. Sirakaya and Woodside, 2005). When faced with a large choice set, only a subset of options passes a first cutoff point to be evaluated in a second step. Given the intangibility of tourism products, scholars also find that tourists make their holiday decisions in an abstract way, paying less attention to objective criteria than they do for other goods and more on the expected experiences and subjective image of the potential destinations. Some efforts to consider choice set endogeneity in recreational demand include Haab and Hicks (1997) and Thiene et al. (2017).

Similarly, other heuristics like status quo bias, mental accounting, anchoring, reference dependence or loss aversion have also been argued to affect discrete choices. For tourism-related decisions, individuals plan travelling expenditures and form expectations about price levels at potential destinations, usually based on previous experiences. In line with Prospect Theory, attributes and prices for each alternative might be judged with

---

tastes (β_k) may be explicitly allowed to depend on individuals characteristics such as income. Alternatively, income could be specified in equation (3.2) interacted with either the characteristics or with the choice alternative intercept.

reference to past outcomes, perceiving the balance between attributes and cost as a gain or a loss relative to that benchmark. In this regard, Nicolau (2012) documents that tourists exhibit asymmetric response to prices, providing evidence of loss aversion heterogeneity. The study by Nguyen (2016) shows that outbound tourists with high loss aversion and present biased are more likely to overspend in their incoming tourism trips.

Because of all these potential departures from RUM theory, research in discrete choice modelling has proposed several non-RUM based alternatives. Some widely known examples are Random Regret Minimization[113] (RRM, Chorus, 2010) or the Stochastic Satisficing model (González-Valdés and Ortúzar, 2018)[114]. Within the RUM framework, other scholars have developed choice models that partially address some of the heuristics raised above. For example, Zhang et al. (2004) introduces a relative utility maximization framework, where utilities are weighted by the relative interest individual attaches to each available option in the choice set. Most widely applied, another stream of research has explicitly considered the propensity to attend each attribute in what is known as Attribute Non-Attendance (ANA). Some examples include Hole (2011), Hole et al. (2013) and Thiene et al. (2015). Furthermore, since in some cases there is some uncertainty over outcomes, scholars in DCE have proposed models that explicitly weight the utilities by the likelihood of occurrence of each alternative (e.g. Glenk and Colombo, 2013; Rolfe and Windle, 2015).

Instead of simply addressing one particular deviation from the RUM assumptions, the literature has gone beyond and proposed methodologies that allow disentangling behavioral heterogeneity in choice decision making. Swait and Adamowicz (2001) propose a latent class approach that infers the most likely decision-making heuristic that underlies choice behavior for each individual. These authors allow decision strategies to vary depending on task complexity and task order, showing that increased complexity is associated with strategy switching. A limitation of their approach is that they do not allow for taste heterogeneity, so that the identification of classes of behavioral processes might be confounded with heterogeneity in marginal sensitivities. Their approach was later improved by Hess et al. (2012), who allow for heterogeneity in both marginal utilities and behavioral processes. By applying this methodology to four different datasets, they indicate that allowing for different decision rules improves model fit and willingness-to-pay estimates.

Similarly, Balbotin et al. (2017) go a step forward and relate taste heterogeneity with behavioral heterogeneity to see whether the apparent heterogeneity in preferences is in fact driven by heterogeneity in process heuristics. They show that taste heterogeneity can be better understood once conditioning on process heuristics. More recently, González-Valdés and Raveau (2018) have proposed a Mixed Heuristic Model through which they identify different heuristics in terms of latent classes but giving higher flexibility to the latent class allocation function. Interestingly, both their proposed procedure and

[113] This approach considers that individuals minimize the regret of making a given choice rather than maximize utility, where the regret function depends on the differences between the alternatives' attribute levels.
[114] This follows from Simon's Satisficing theory (Simon, 1955) by which individuals choose the first acceptable alternative they find, where acceptability requires all the attributes to pass a certain quality threshold.

the traditional LC method suggest that the probability that individuals follow non-RUM heuristics is low.

Despite all this criticism and the development of several alternatives, RUM continues to be the preferred framework for discrete choice modelling. Hess et al. (2018) provide a detailed discussion on RUM and its alternatives. These authors conclude that consistency with RUM does not require individuals to behave in a RUM-style process, but only in a style that leads to observed choices that are consistent with it. They indicate that any behavioral trait can be closely approximated by RUM, independently of whether individuals follow exactly the decision-making process that RUM postulates. We share their view and consider RUM theory to be the best way to approximate the complex decision-making process of holiday destination choice. This is because although RUM theory has some weaknesses, the proposed alternatives introduced before also exhibit some limitations[115]. The Random Regret Minimization (RRM) approach, which is possibly now the most accepted alternative, lacks theoretical microeconomic foundation for the derivation of welfare measures. Furthermore, in terms of model fit, Chorus et al. (2014) review 43 empirical studies and compare the performance of RUM and RRM without finding a clear winner.

The key point is that the validity of the RUM approach depends on whether the purpose is to provide an accurate representation about the predictors of choice outcomes or about how individuals arrive at that outcomes. If the purpose is the former (as is our case), consistency only require individuals to behave *as if* they truly follow the assumptions of RUM (Hess et al., 2018)[116]. As a result, our modelling approach is based on RUM theory.

4.2. Allowing for Preference heterogeneity: Latent Class Model

Since the seminal paper by Train (1998), it has been recognized the necessity and convenience of allowing for taste heterogeneity to avoid bias in attribute coefficient estimates. The typical way to do so is by means of the Random Parameter Logit (RPL) and the Latent Class Model (LCM)[117]. A growing interest has been paid to the sources of unobserved heterogeneity by relating either the means of the random parameters or the class membership function to respondents' characteristics.

Although the RPL model is the most used to represent taste heterogeneity, here we rely on the latent class framework. Latent class models have a long tradition in the recreational demand literature and DCEs (Boxall and Adamowicz, 2002; Juutinen et al., 2011; Araña et al., 2016; Schuhmann et al., 2016; Chen et al., 2019). They assume that

[115] All the methodologies that aim to disentangle heterogeneity in preferences and heterogeneity in decision heuristics are very demanding in terms of data quality. Since both features are unobserved, these approaches are highly sensitive to random noise. Moreover, the method proposed by González-Valdés and Raveau (2018) is not able to identify the different choice heuristics when the alternative specific constants dominate the marginal utilities of the attributes.

[116] Matejka and McKay (2015) derive a discrete-choice version of the rational inattention model, showing that choices follow a multinomial logit if prior knowledge about outcomes is homogeneous. Additionally, the scale term can be understood as the cost of information.

[117] The difference between them relates to the assumption of a continuous distribution for representing taste heterogeneity in the RPL as opposed to a discrete number of classes in the LCM.

the population is composed of a discrete set of (unobserved) groups or classes of individuals with homogeneous preferences within each class, but different preferences across classes (Kamakura and Russell, 1989).

There is no consensus in the literature about whether taste heterogeneity is better represented by means of random parameters or latent classes. Sen (2009) favors the use of the latter while Keane and Wasi (2013) find that models with random heterogeneity outperform latent class modelling in terms of model fit. By contrast, other authors like Greene and Hensher (2003), Provencher and Bishop (2004), Provencher and Moore (2006), and Hynes et al. (2008) conclude that neither of them is strictly preferred to the other[118]. The latent class offers the advantage of imposing less parametric structure on heterogeneity and lower computational burden at the cost of blocking individuals into groups with homogeneous preferences. Another advantage of latent class is the possibility of estimating the size (shares) of each class of individuals. Nevertheless, the adequacy of each approach seems to depend on the dataset being analyzed.

Assume that the population is composed of several classes of individuals with different preferences. Conditional on membership to class c for $c = 1, \dots, C$, and following RUM theory, the utility each individual i obtains for each alternative j in choice situation t is expressed as:

$$U_{ijt|c}{}^{*} = ASC_{jc} + \beta_c X_{ijt} + \varepsilon_{ijt|c} \qquad (3.3)$$

where X_{ijt} is a vector of observed attributes that vary over alternatives (including total cost), individuals (depending on the block and order being assigned) and choice situations, β_c is a vector of parameters to be estimated for each class, ASC_{jc} is a set of alternative-specific constants for each class that gather residual utility not captured in X_{ijt} (one of them is normalized to zero for identification), and $\varepsilon_{ijt|c}$ is the random error term. The utility ($U_{ijt|c}{}^{*}$) is latent and we *infer* preferences (β_c) from observable indicators of choices ($y_{ijt} = 1$ if alternative j is chosen by individual i in choice situation t, 0 otherwise).

A major point in our analysis is that we assume the so-called consumer sovereignty property, by which individual preferences are predetermined in any choice situation. In the words of McFadden, *desirability precedes availability* (McFadden, 2000). Without prejudice that preferences can change over time, preferences for the attributes are supposed to be exogenously determined at the time of performing the choice task. This is an important assumption in RUM theory, since it implies that the distribution of the error term does not depend on X_{ijt}. Accordingly, we allow for inter- but not for intra-consumer heterogeneity[119].

[118] Greene and Hensher (2013) develop a mixture of the two approaches by allowing for random heterogeneity within each class in what they labelled as the Latent Class Mixed Multinomial Logit. An empirical application of this in the recreational literature is Bujosa et al. (2010).

[119] Although recent attempts to model intra-consumer heterogeneity have been developed (e.g. Danaf et al., 2019), when tastes are allowed to vary over choice situations its identification becomes highly tenuous (Hess and Train, 2011). Consistent with neoclassical theory, preferences are assumed to be stable with drift and whimsy treated as nuisance factors (Ben-Akiva et al., 2019).

Since utility under the RUM framework is an additive sum of two components, it is necessary to fix the metric of the latent utility (the scale) in some way. This scale refers to the weight of the deterministic component in the overall utility and is inversely related to the standard deviation of the idiosyncratic error term. Although the scale just changes the value of the utility levels but not the corresponding choices, it matters for parameter identification. We follow the empirical literature and fix the scale to unity, which is equivalent to assume homoscedasticity in the error term. This implies that coefficient estimates cannot be directly compared across different model specifications.

The error term is assumed to be Type I Extreme Value distributed (Gumbel) so that the difference between them leads to a Logit form[120]. A normal distribution could also be assumed, although Logit has a longer tradition because of its better tractability. Given a specific value of the parameters for each class (i.e. β_c), the conditional probability of respondent i's sequence of choices is given by the product of logit probabilities:

$$P_i = Pr(y_{ijt}|c, X_{ijt}) = \prod_{t=1}^{T} \prod_{j=1}^{J} \frac{\exp\,(ASC_j + \beta_c\, X_{ijt})}{\sum_{j=1}^{J} \exp\,(ASC_j + \beta_c\, X_{ijt})} \quad (3.4)$$

being $y_{ijt|c}$ a binary indicator of whether individual i that belongs to class c chooses option j at choice situation t facing attributes X_{ijt}.

Individuals are assigned to classes probabilistically. Although class membership can be modelled semi-parametrically based on a constant term (Scarpa and Thiene, 2005), the most common way is to assign individuals to classes based on sociodemographic characteristics (Z_i) using a semiparametric multinomial logit structure (e.g. Araña et al., 2016; Chen et al., 2019). Accordingly, the class allocation function is expressed as follows:

$$\pi_{ic} = \frac{\exp\,(\mu_c + \lambda_c{}' Z_i\,)}{\sum_{c=1}^{C} \exp\,(\mu_c + \lambda_c{}' Z_i)} \quad (3.5)$$

where π_{ic} is the probability that respondent i belongs to class c, μ_c is a set of constants to be estimated for each class (normalized to one in one class for identification) and λ_c is a vector of parameters to be estimated for c-1 classes. It holds that $\sum_c \pi_{ic} = 1$ and $\pi_{ic} > 0$.

Therefore, the unconditional probability of a sequence of choices over t choice situations is given by:

$$P_i = Pr(y_{ijt}|X_{ijt}) = \sum_{c=1}^{c} \pi_{ic} \prod_{t=1}^{T} \prod_{j=1}^{J} \frac{\exp\,(ASC_j + \beta_c\, X_{ijt})}{\sum_{j=1}^{J} \exp\,(ASC_j + \beta_c\, X_{ijt})} \quad (3.6)$$

[120] As illustrated by McFadden and Train (2000), the use of the Extreme Value distribution stems from its max-stable property by which the maximum of two independent extreme value random variables with the same scale factor is an extreme value random variable with the same scale factor. Furthermore, independent identically distributed additive disturbances are consistent with RUM theory if and only if the disturbances follow an Extreme Value Type I distribution (McFadden, 2001).

Contrary to the RPL model, choice probabilities in the LCM do not require integration, so estimation is done using standard Maximum Likelihood (ML). This has the advantage of being less sensitive to simulation error (Czajkowski and Budzinski, 2019).

4.3. Marginal Rates of Substitution and Willingness to Pay

Although the model parameters reflect the marginal utility of each attribute up to the scale of the error term, possibly the most interesting issue in our context is the calculation of the marginal rates of substitution (henceforth MRS) between one characteristic and another. That is, at what rate are individuals willing to change, for instance, a longer stay by lodging at a low-quality accommodation, *ceteris paribus*?

The MRS between two attributes is given by the ratio of the corresponding marginal utilities with a minus sign. This follows directly from the linear-in-parameters specification of the systematic component of utility[121]. To illustrate this, and following the previous example, suppose we aim to compute the rate at which an individual is willing to change a longer stay (from 3 days to 10 days) by staying in an apartment (instead of a 4-star hotel), everything else being equal (among other factors, the cost), while keeping constant her utility. For this illustration, assume taste homogeneity:

$$V_{jt} = ASC_j + \beta_1\ 10days_{jt} +\ \beta_2\ 4starhotel_{jt} +\ \ddot{\beta}X_{ijt} \qquad (3.7)$$

where $\ddot{\beta}X_{ijt}$ refers to the product of the remaining attributes in the utility function other than a long stay (*10days*) and high-quality accommodation (*4starhotel*) and their marginal utilities.

The increase in utility if we change from *3days* to *10days* is given by:

$$\Delta V_{jt} = \beta_1 \qquad (3.8)$$

The change in utility if we change from *4starhotel* to *apartment* is:

$$\Delta V_{jt} = -\ \beta_2 \qquad (3.9)$$

Holding utility constant, the marginal rate of substitution (MRS) of *4starhotel* for *10days* is:

$$\Delta V_{jt} = 0\ \rightarrow\ MRS_{10days,4starhotel} = \frac{\beta_1}{-\ \beta_2} \qquad (3.10)$$

[121] Assuming Lancaster's *product characteristics* approach, the deterministic component of utility for a representative agent can be represented by a CES utility function. If hedonic attributes are assumed to be perfect substitutes, the utility collapses to a linear-in-parameters function (plus an additive random term). See Anderson et al. (1987). The linear specification is used for econometric tractability, but any monotonic transformation would describe choice behaviour equally well (Varian, 2010). Furthermore, McFadden indicates that "any continuous indirect utility function can be approximated on a compact set to any desired degree of accuracy by a linear-in-parameters specification" (McFadden, 1981, p. 220).

For the case of dummy coded attributes, the MRS depends on the base category. Importantly, the discrete nature of the indicator does not affect the computation of the slope of the indifference curve as long as the nature of the attributes that define the utility are discrete. A good property of the MRS is that it is independent of any monotonic transformation of the utility function[122].

To facilitate interpretation, a better way to understand MRS is to convert it into monetary values (Willingness to Pay, WTP). Given a marginal disutility of total cost, the MRS between a given level *k* in attribute *a* and the cost measures how much money the individual is willing to pay to change from the base category to level *k* (Varian, 2010). Therefore, the WTP is given by:

$$WTP_{ak} = \frac{\beta_{ak}}{-\beta_{cost}} \tag{3.11}$$

This WTP can be seen as “money-metric” utility in the sense of Samuelson (1950) in which the preferences for the characteristics are evaluated at given prices (total cost). They key assumption here is that the marginal utility of income equals the disutility of total cost (i.e. the opportunity cost of getting level *k* for attribute *a*). As such, assuming constant marginal utility of income, the WTP in (3.11) measures how much money a consumer is willing to pay to get level *k* in attribute *a* keeping constant her utility.

5. EMPIRICAL ANALYSIS

5.1. Model specification

Consistent with RUM theory, conditional on belonging to class c, the latent indirect utility of alternative j for individual i in choice situation t ($U_{ijt|c}{}^{*}$) is given by:

$$\begin{aligned} U_{ijt|c}{}^{*} = ASC_{jc} \; + \; \beta_{1c}\, medTT_{ijt} \; + \; \beta_{2c}\, longTT_{ijt} \; + \; \beta_{3c}\, bustrain_{ijt} + \beta_{4c}\, plane_{ijt} \\ + \beta_{5c}\, 7days_{ijt} \; + \; \beta_{6c}\, 10days_{ijt} \; + \; \beta_{7c}\, 2starhotel_{ijt} + \beta_{8c}\, 4starhotel_{ijt} \\ + \beta_{9c}\, Cost_{ijt} \; + \; \varepsilon_{ijt|c} \end{aligned} \tag{3.12}$$

where $\varepsilon_{ijt|c}$ is an iid Type I Extreme Value distributed error term, ASC_{jc} are alternative-specific constants to be estimated for each class (gathering residual preference for coastal, urban and nature-based destinations relative to the none option), $medTT$, $longTT$, $bustrain$, $plane$, $7days$, $10days$, $2starhotel$, $4starhotel$ and $Cost$ are the attributes defined above, and β are parameters to be estimated for each class.

The attribute levels (except *Cost*) are dummy coded taking ASC_4, $shortTT$, car, $3days$ and $apartment$ as the reference categories. In this regard, there is an ongoing discussion in the literature about whether category levels should be specified as dummy coded or effects coded (e.g. Bech and Gryd-Hansen, 2005). The distinction relates to

[122] Whereas the marginal utilities are affected by any monotonic transformation of the utility function, the MRS is independent of the way adopted to represent preferences.

whether the base categories are collapsed in the alternative specific constant, in which case the utility contribution for the omitted category is set to zero (dummy coding), or recoded in such a way that the omitted category equals the negative sum of all the estimated levels (effects coding). The latter normalization has been argued to be better since it avoids the usual confounding between the base levels and the alternative specific constants. However, Daly et al. (2016) prove that both procedures lead to the same results, since only differences in utilities across categories matter. They argue that effects coding does not provide the advantages claimed by some practitioners and for proper implementation it requires a weighted normalization. Additionally, the derivation of the MRS in the form of WTP for categorical variables under effects coding becomes more cumbersome[123]. Due to these reasons, in our analysis we adopt classical dummy coding. Nevertheless, the corresponding estimates for effects coding using the formulas outlined in Daly et al. (2016) are available from the authors upon request.

For the purpose of modelling class membership, we consider the following sociodemographic characteristics in the allocation function (3.5): gender, age, high education level (vs primary and secondary education) and monthly individual net income:

$$Z = (female, age, higheduc, income) \tag{3.13}$$

Of course, other individual characteristics could be also used to assign individuals to classes. However, we tried to keep the model formulation parsimonious to avoid highly correlated covariates[124]. Other variables like generic preference for types of destinations collected in the questionnaire have the problem that because being self-reported share unobservables with the error term (ε_{ijt}), which could lead to endogeneity concerns[125].

Before model estimation, the number of classes needs to be determined. Table 3.6 presents the (consistent) Akaike Information Criteria (cAIC and AIC), Bayesian Information Criteria (BIC) and log likelihood values (log L) for two, three and four classes. The AIC indicates four classes and cAIC and BIC indicate two classes. On the one hand, some authors suggest that AIC over-estimates the number of classes and, on the other hand, others document that BIC tends to favour small number of classes, especially in small sample sizes (McLachlan and Peel, 2000). Furthermore, Scarpa and Thiene (2005) and Hynes et al. (2008) indicate that, in choosing the number of classes, the statistical criteria and the significance of the parameter estimates needs to be tempered by the researcher's own judgement of the suitability of the model. Since we aim to capture groups of individuals with different preferences for the attributes, we consider taste heterogeneity to be better represented in our data by three segments.

[123] The WTP for *10days* relative to *3days* under effects coding would be:

$$WTP_{10days} = \frac{\beta_{10days} - (-\beta_{10days} - \beta_{7days})}{-\beta_{cost}} = \frac{2 * \beta_{10days} + \beta_{7days}}{-\beta_{cost}}$$

[124] For instance, there is high correlation between labour market participation and income (0.67); student status and age (-0.55); or being married and age (0.81). Collinearity is particularly problematic in LCM models, especially when working with dummy variables since they contain less variation.

[125] This could be addressed by means of hybrid choice models, which are becoming increasingly popular in the discrete choice literature for considering latent attitudes (Hoyos et al., 2015; Hess et al., 2018; Liebe et al., 2019). However, these models are quite demanding computationally in terms of the number of parameters to be estimated, which could lead to problems of weak identification, thereby requiring larger sample sizes. A formal characterization can be found in Walker and Ben-Akiva (2002).

	2 classes	3 classes	4 classes
Log L	-1,695.4	-1,637.0	-1,600.9
K	29	46	63
AIC	3,448.8	3,366	3,327.8
cAIC	3,633.2	3,658.6	3,728.5
BIC	3,604.2	3,612.6	3,665.5

Table 3.6.- Information criteria statistics

5.2. Results

Table 3.7 presents the parameter estimates for a baseline Multinomial Logit model (MNL, also known as Conditional Logit) together with the three class LCM[126]. Standard errors are Huber-White heteroskedasticity-consistent[127]. We first briefly discuss the estimates for the MNL model, and we then turn to the LCM. The estimation has been done in R both using *Apollo* (Hess and Palma, 2019) and the *gmnl* package (Sarrias and Daziano, 2017)[128]. For computational reasons, in both models the Cost attribute is rescaled by 1/100 (in hundreds of euros).

The MNL is usually used as a benchmark model in discrete choice studies. Consistent with descriptive statistics, respondents prefer the coastal alternative (ASC_1), *ceteris paribus*. Since the DCE was framed for a summer trip, this is an expected result. The urban destination (ASC_2) is the second preferred option, followed by the nature-based alternative (ASC_3). Nevertheless, since the ASC gather the mean of the error term for the utility of each alternative, these estimates need to be interpreted with caution. Strikingly, travel time is not significant for explaining destination choices and there are no differences between travelling by car (base category) or by bus or train. However, respondents prefer to travel by plane. Consistent with expectations, tourists prefer longer stays and trip cost exerts a negative effect on utility. Concerning accommodation type, 4-star hotels are preferred over apartments (base category) and 2-star hotels.

Turning to the LCM parameter estimates, individuals are assigned to the different segments probabilistically. Class 2 appears to be the largest (50%), followed by Class 3 (40%) and Class 1 (10%). For identification, the class allocation parameters were set to zero for Class 1 so that the corresponding ones for Classes 2 and 3 are interpreted relative to that class. Class 1 is more likely to be composed of relatively elderly females

[126] McFadden (2001) argues that the labelling of *Conditional Logit* is because it is a Multinomial Logit of the conditional demand given the feasible set of choice alternatives.

[127] To address potential cross-sectional dependence, we clustered standard errors at the couple level. Although the standard errors slightly differ, the magnitude of the coefficient estimates and the statistical significance remains unchanged. Results are available upon request.

[128] These two modules use Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which belongs to quasi Newton optimization methods. Alternatively, the Expectation-Maximization (EM) algorithm has been shown to be a valuable alternative, since in some instances maximizing the log likelihood using gradient-based optimization routines can produce convergence problems (e.g. Train, 2008). Indeed, Vij and Krueger (2017) show that the EM algorithm produces greater stability. To check this, we also estimated the LCM model in Stata 16 using the user-written *lclogit* module that implements the EM algorithm (Pacífico and Yoo, 2013). Parameter estimates are roughly the same up to the third decimal.

with university education and low income. By contrast, Class 2 mainly comprises young males with middle/high income and non-university education. Class 3 is similar to Class 2. However, Class 3 appears to be more balanced in terms of age and gender but to be composed in a greater proportion of individuals with high income and non-university studies[129].

For respondents in Class 1, the length of the stay and trip costs are the only attributes that drive their vacation choices. These individuals prefer longer stays and are deterred by cost. None of the remaining attributes are statistically significant. This suggests that this group of individuals might exhibit lexicographic preferences by which they only attend some attributes in their decision-making. Nevertheless, this group represents a small share of the sample (10%).

By contrast, individuals in Class 2 attach great importance to the type of destination, everything else being equal. Coastal destinations are strongly preferred, followed by urban and nature-based tourism. The plane is preferred to other modes of transport, *ceteris paribus*, while travel time is not statistically significant. Also for this group, longer vacation duration constitutes a desirable feature. Indeed, the marginal utility for this attribute is larger than for Class 1. However, the type of accommodation dwelling does not significantly affect their utilities. Put another way, respondents in Class 2 appear to be indifferent with respect to the accommodation type. Regarding cost, this variable exhibits a negative effect on utility as economic theory dictates. Price sensitivity is relatively larger for this segment than for those in Class 1.

[129] We estimated the LCM model without specifying any variable in the class allocation function apart from a constant term. The coefficient estimates and its significance remain largely unchanged.

	MNL			LCM								
				Class 1			Class 2			Class 3		
Variables	Coef.		Rob.SE	Coef.		Rob.SE	Coef.		Rob.SE	Coef.		Rob.SE
ASC1	1.554	***	0.176	-0.736		1.176	3.339	***	0.507	1.943	***	0.430
ASC2	1.383	***	0.182	-0.646		1.548	2.108	***	0.491	2.712	***	0.449
ASC3	0.877	***	0.183	-1.572		1.232	1.403	***	0.463	2.365	***	0.491
medTT	0.029		0.090	0.007		0.622	-0.129		0.238	0.139		0.183
longTT	-0.031		0.081	-0.490		0.478	0.018		0.197	-0.139		0.203
bustrain	0.018		0.078	-0.488		0.584	0.086		0.174	0.097		0.128
plane	0.278	***	0.068	0.348		0.471	0.468	**	0.188	0.294	**	0.115
7days	1.298	***	0.106	1.731	***	0.401	2.270	***	0.294	0.904	***	0.206
10days	1.419	***	0.109	1.808	***	0.432	2.621	***	0.343	0.883	***	0.261
2starhotel	-0.170	*	0.093	-0.419		0.618	0.004		0.171	-0.289	*	0.161
4starhotel	0.235	***	0.085	0.006		0.575	0.175		0.273	0.391	***	0.145
Cost	-0.186	***	0.015	-0.197	**	0.099	-0.272	***	0.040	-0.177	***	0.040
Class membership												
const							4.580	***	0.411	3.273	***	0.427
female							-0.739	***	0.205	-0.509	**	0.205
age							-0.090	***	0.008	-0.045	***	0.007
higheduc							-0.587	**	0.252	-1.093	***	0.248
income							0.258	***	0.091	0.300	***	0.088
LC probabilities				0.108			0.504			0.405		
N	262						262					
Observations	1,572						1,572					
Log Likelihood	-1807.7						-1637					

Table 3.7.- MNL and LCM parameter estimates
*** p<0.01, ** p<0.05, * p<0.1

For individuals assigned to Class 3, the destination labelling also matters in their choices. However, these individuals value more urban and nature-based destinations than coastal ones. As those in Class 2, respondents in Class 3 prefer travelling by plane than by car, bus or train, although here the magnitude of the marginal utility for air travelling is lower. Longer stays are also preferred but note here that the marginal utilities associated with this attribute are the lowest compared with the other classes. Contrary to the other two segments, individuals in this class attach positive value to 4-star hotels relative to apartments. Cost negatively impacts utility, although this group of respondents appears to be the least price sensitive.

Overall, we find that all respondents attach positive utility to longer stays, in line with related studies (Van Cranenburgh et al., 2014; Grigolon et al., 2012; Oppewal et al., 2015). Consistent with microeconomic theory, the higher the total cost of an alternative, the lower the likelihood of that alternative being chosen. Nevertheless, price sensitivities seem to differ across classes. This is consistent with Masiero and Nicolau (2012). For some individuals, plane is significantly preferred over car, as reported in Van Cranenburgh et al. (2014) and Grigolon et al. (2012). Concerning the type of accommodation, respondents in Classes 1 and 2 do not show a significant preference for hotels relative to apartments. This mimics the findings by Schuhmann et al. (2016) in the context of beach recreation. However, individuals in Class 3 significantly prefer 4-star hotels. This is also found in Dellaert et al. (1997) and Oppewal et al. (2015).

Possibly the most intriguing result is the non-significance of travel time, neither in the MNL nor in any of the classes in the LCM. In principle, we would expect travel time to negatively affect utility due to the opportunity cost of time. However, our estimates appear to indicate that this attribute is not affecting vacation choice in our data[130]. This is also reported in Dellaert et al. (1997) and Huybers (2003). This result is in line with the low correlation between choices and travel time levels presented in Table 3.5. It could be the case that, although after-tax monthly income (as a proxy of the opportunity cost of time) enters the class allocation function, our model is not properly controlling for the increasing disutility of travel time one would expect as income increases. To explore this, we have estimated a MNL model in which $medTT$ and $longTT$ are interacted with income[131]. None of the interactions are significant at 95% significance level, which indicates that the non-significance of travel time does not appear to be due to hidden heterogeneity in the opportunity cost of time. Furthermore, following empirical evidence in the transportation literature (Guevara, 2017), the (dis)utility of travel time could differ by mode of transport. To rule out the possibility that our results are driven by omitted interaction effects, we have estimated a MNL model with interactions between the levels of mode of transport ($bustrain$ and $plane$) and travel time ($medTT$ and $longTT$). As before, none of the interaction terms are statistically significant. This is in line with Dellaert et al. (1997).

---

[130] Instead of using dummy coding for travel time, we estimated both the MNL and the LCM models assuming this attribute to be continuous. Results are consistent with those reported in Table 7: travel time is never statistically significant for explaining choices. The estimates are available upon request.

[131] There is a discussion in the recreational demand literature about the common practise of defining the opportunity cost of time as 1/3 of the wage rate. Recent evidence by Czajkowski et al. (2019) shows this practise is unfounded. Accordingly, the use of net monthly income to explore differences in the value of time is merely approximate.

In spite of the fact that the focus groups we conducted pointed to travel time as a determinant of destination choice, its non-significance requires to be interpreted. In the context of recreation, there is some evidence about individuals attaching positive utility to transit time, which can be perceived as a 'commodity value' (Chavas et al., 1989; Mokhtarian, 2005; Larson and Lew, 2005). Indeed, Van Cranenburgh et al. (2014) document that travel time impacts utility positively. As such, it could happen that both effects cancel each other. However, this should imply statistically significant positive and negative marginal utilities across classes. Alternatively, a more plausible explanation could be that respondents do not pay enough attention to this attribute in their decision process. In this regard, Hess et al. (2007) find that, in the context of airline choice, vacation travellers are not sensitive to late arrivals and time delays. This relates to our discussion in subsection 4.1 about Attribute-Non-Attendance (ANA)[132]. Although some complex econometric models to *infer* ANA have been developed, the distinction between ignoring an attribute and attaching reduced importance to it is quite fuzzy (see Hess et al. (2013) for a discussion on the problems of identification)[133].

Several remarks are in order. First, since that the dummy-coded attributes are relative to the base category, one might be interested in exploring whether, for instance, the effect of $7days$ is statistically different from the one of $10days$. To examine this, we changed the reference category by setting to zero the middle level in the four dummy-coded attributes. We document that medium and long vacation trips (7 and 10 days) are preferred over short trips (3 days), *ceteris paribus*, for the three classes. However, there are no statistical difference between medium and long trips. This highlights the importance of allowing for non-linear effects in the marginal utilities.

Second, the model is parameterized in the so-called 'preference space', by which coefficients directly measure the marginal utilities for the attributes. When researchers introduce continuous heterogeneity in the form of random parameters, an alternative way to proceed is to define the model in the 'willingness to pay space', by which the utility is directly specified as the weighted sum of the ratio of the attributes to the cost (Train and Weeks, 2005; Scarpa et al., 2008). This is done to avoid deriving non-defined WTP ratios if the denominator is close to zero[134]. WTP estimates derived from the two approaches differ in the case of the RPL model[135]. However, as shown by Oviedo and Yoo (2017), this is not an issue in the LCM model.

---

[132] We considered the possibility of implementing a follow-up question asking respondents to report whether they attend each attribute in their choices (stated ANA). Apart from making the experiment longer, more demanding, and possibly leading to endogeneity concerns, this procedure, although informative, has been shown not to be appropriate for this purpose (Hess and Hensher, 2010).

[133] Scarpa et al. (2009) and Hole (2011) developed the inferred ANA approach in which the attribute processing strategies are modelled by a constrained LCM without considering taste heterogeneity. Recent developments have gone beyond and handle inattention and preference heterogeneity in a Latent Class Random Parameter Logit model (Hole et al., 2013; Thiene et al., 2019). However, the latter requires larger datasets for proper parameter identification.

[134] This is problematic in models that allow the covariance matrix not to be diagonal and multivariate normal.

[135] We have estimated a RPL model both in the preference and the willingness to pay spaces using *mixlogit* and *mixlogitwtp* modules in Stata 16 (Hole, 2007a). We have assumed that all the attributes except the ASCs are normally distributed using 1,000 Halton draws. The negative of the cost was specified as log-normally distributed to be consistent with economic theory following usual practice. Both in the case of correlated and uncorrelated random coefficients, the derived WTP measures notably differ between the two approaches.

Third, one might wonder whether travel experience plays a role in the elicitation of vacation preferences. This is in line with the previous discussion about the role of familiarity with the product being analysed on hypothetical bias. Those who are more used to travelling might put more interest in the choice tasks. This could be related with the scale of the error term so that experience makes choices more deterministic from the researcher perspective, and therefore reduce the variance of the error term (Czajkowski et al., 2014; 2016). That is to say that the error terms are heteroskedastic in travel experience. Other scholars refer to it as scale heterogeneity or 'ability to choose' (Christie and Gibbons, 2011). This issue is related to the rational inattention framework by which choices depend not only on preferences but also on information (Matveenko, 2020). To explore this, we estimated a Scaled Multinomial Logit Model (SMNL) in which we parametrize the scale of the Gumbel error term as a function of reported travel frequency obtained from the questionnaire (*travfreq*). Results are reported in Annex 4. We find that travel frequency is not statistically significant. Accordingly, travel experience does not make a difference in the randomness of respondents' choices.

Fourth, and more importantly, the paper aims to estimate the *individual* preferences for destination attributes using stated preference data from real life couples. However, as behavioural economists note, choice is social and tends to be influenced by others. Since the choice task in framed in the context of a joint trip with their partner, in their utilities respondents might consider the utility that each alternative produces to their partner. That is, the systematic part of the utility might include an altruistic component in the form of social preferences (i.e. *caring* preferences in the *Beckerian* sense or *deferential* preferences à la Pollak)[136].

To see this, let $k = 1{,}2$ denote each member within a given couple. Assume individual 1 has an expectation of the utility that alternative j in choice situation t produces to individual 2, given the attributes (i.e. $E_1(U_{2jt}\,|X_{2jt})$). Assume also that individual 1 attaches positive utility to the expectation of the utility each alternative produces to individual 2, so that the utility function can be reformulated as follows:

$$U_{1jt|c} = \; asc_j + \; b_c\, X_{1jt} + \; \gamma\, \mathrm{E}_1(U_{2jt}|X_{2jt}) + e_{1jt} \qquad (3.14)$$

Assume that the utility individual 1 expects for 2 is:

$$\mathrm{E}_1\big(U_{2jt}\big|X_{2jt}\big) = \widetilde{asc}_j + \; \tilde{b}X_{2jt} + \; e_{\ddot{1}jt} \qquad (3.15)$$

Since partners cannot communicate at the time of choosing, $e_{\ddot{1}jt}$ is an error term that measures the noise in the signal by which individual 1 perceives the utility of individual 2. Since by the experimental design $X_{1jt} = X_{2jt}$, equation (3.14) can be rewritten as follows:

This is consistent with previous evidence reported by Hole and Kolstad (2012). Results are available upon request.

[136] See Pollak (2003) for a discussion on their difference.

$$U_{1jt|c} = \underbrace{(asc_j + \gamma \widetilde{asc_j})}_{ASC_j} + \underbrace{(b + \gamma \tilde{b})}_{\beta} X_{1jt} + \underbrace{e_{1jt} + \gamma \ddot{e}_{1jt}}_{\varepsilon_{1jt}} \tag{3.16}$$

Accordingly, our parameter estimates measure the marginal utilities (*behavioral preferences*) that each individual attaches to each alternative. However, as shown by equation (3.16), these estimates mix up the pure individual self-regarding preferences (*core preferences*) and the altruistic ones. Only if $\gamma = 0$, revealed behavioral preferences equal the core preferences[137]. Consistent with Fershtman and Segal (2018), it is worth noting that here the source of the influence of partner's preferences is not the result of the aggregation of preferences but a kind of altruism from 1 towards 2 (and vice versa)[138]. Note also that this case departs from the peer effects documented in the economic literature (e.g. Brock and Durlauf, 2001) because individual 1 *does not observe* the preferences of the partner at the time of choosing but simply holds *a belief* about them.

Remarkably, the incorporation of the expectation of the partner's utility for each alternative in the utility index under the RUM framework makes behavioral preferences to be empirically indistinguishable from core ones. Unless further information is considered, self-regarding and social preferences rationalize choice decisions in the same way (Manski, 2004)[139].

Together with their preferences, in the DCE respondents are asked to indicate their prediction about the choice of their partner facing the same choice card. Based on this, we could think of estimating the expectation about partner's preferences and introduce them into the model as an additional regressor. If so, we would be able to control for the part of the systematic utility that is derived from altruism, and therefore the estimated parameters would measure the pure self-regarding preferences. This emerges as a direct consequence of the separable representation of the *core* and the *altruistic* preferences. However, since expectations share the same unobservables with the idiosyncratic error term, then $\mathrm{E}_1(\widehat{U_{2jt}|X_{2jt}})$ is endogenous. There are different procedures to accommodate endogeneity in discrete choice modelling, which is more difficult to deal with than in linear regression because the corrections alter the scale of the model[140]. However, all of them require valid instruments that satisfy the conditions of relevance and conditional exogeneity to be properly implemented. Importantly, instruments need to vary per alternative. Unfortunately, we lack such information.

Alternatively, rather than the estimated expectation of the partner's utility, we could proxy it by the choice probabilities derived from the model in (3.15). However, that would make us incur in the well-known *forbidden regression* due to plugging non-linear fitted values into a non-linear model (Angrist and Pishke, 2008). The only way would be to rely on

[137] The labelling of *behavioral* and *core* preferences to refer to overall and purely self-satisficing preferences is borrowed from Fershtman and Segal (2018).
[138] This case is different from a *collective* perspective because individual 1 only holds a noisily expectation of the preferences of individual 2.
[139] This is closely related to Chiapori's point that selfish preferences lead to the same testable implications than caring ones under Pareto efficiency in a collective framework (Chiappori, 1992).
[140] The control function approach is the most applied (Petrin and Train, 2010), although recently scholars have proposed other alternatives like the Multiple Indicator solution (MIS) (Fernández-Antolín et al., 2016) or the Integrated Latent Variable Model (ILVM). We refer the reader to Guevara (2015) for a review.

joint Maximum Likelihood estimation by which equations (3.14) and (3.15) were estimated simultaneously (e.g. Park and Gupta, 2012). However, this procedure would become quite cumbersome due to the course of dimensionality, especially for a small sample size.

Because of this, our data does not allow us to separate the *core* preferences from the *altruistic* ones. Nevertheless, the parameter estimates are consistently estimated and gather the compounded individual marginal utilities, as shown in (3.16).

5.3. Willingness to pay estimates

Based on the parameter estimates in Table 3.7, we derive the WTP for each attribute in each class. Since class membership is probabilistic, it seems necessary to derive a measure of the *unconditional* WTP for the attributes based on the class membership likelihood. A weighted average of individual specific WTP can be computed as the average of the WTP estimates for each class weighted by the predicted class membership probabilities (e.g. Hoyos et al., 2015) as follows:

$$\widehat{WTP}_{ik} = \sum_{c=1}^{C} \widehat{\pi_{ic}}\ WTP_{k,c} \tag{3.17}$$

Table 3.8 presents descriptive statistics of $\widehat{WTP}_{ik}$ except for $medTT$, $longTT$, $bustrain$ and $2starhotel$. Since these attributes are not statistically significant in any of the three classes, their associated WTP relative to the base category is not computed. This follows common practice (e.g. Hole, 2007b; Greene and Hensher, 2013). ASCs are also excluded since the residual preference for each alternative relative to the none-of-them is confounded with the base categories. As such, the interpretation of the WTP for the ASCs would not be straightforward.

	Min	1st Q	Median	Mean	3rd Q	Max
WTP_{plane} (€)	168.4	169.8	170.4	170.3	170.9	173.4
WTP_{7days} (€)	607.6	690.2	711.3	706.1	729.4	773.6
WTP_{10days} (€)	625.1	738.7	768.6	764.3	798.4	845.0
$WTP_{4starhotel}$ (€)	69.24	107.4	119.1	120.8	129.25	169.6

Table 3.8.- Unconditional WTP estimates

Respondents exhibit a large WTP for increasing the length of stay from 3 days (base category) to 7 and 10 days (€706 and €764, respectively). Although the different trip durations increase by about the same length (3 days), it is interesting that the WTP for 10 days stay is not the double of WTP for 7 days stay but of about the same magnitude. This suggests the existence of non-linearities in the marginal utility for length of stay. Regarding the mode of transport, respondents are on average willing to pay €170 for travelling by plane relative to by car, *ceteris paribus*. Finally, individuals are willing to pay about €120 for lodging at a 4-star hotel relative to an apartment[141].

[141] Recall that the cost attribute reflects the total cost of the trip for the couple.

A standard assumption in the derivation of the WTP is that the marginal utility of income is constant. However, it is plausible that the disutility of cost changes with income. Even though we use income in the class allocation function, one might wonder whether income should also be included in the utility function, interacted with total cost. To analyse this, we have estimated a MNL model with an interaction of income and Cost. The coefficient estimate is non-significant and therefore the assumption of constant marginal utility of income seems to be valid in our data.

Since rather than a single value per class we have derived the distribution of WTP values based on class membership probabilities, it seems relevant to examine how this distribution relates to observable sources of heterogeneity (i.e. the sociodemographic characteristics that determine class membership). To this end, Figures 3.3-3.6 present boxplots for WTP estimates, disaggregated by age (below 22, between 22-40 and more than 40), income (low income, middle income, high income), gender (male vs. female) and education level (primary and secondary studies vs. university studies)[142]. To facilitate comparison in the width of the boxplots, the horizontal axis in all the Figures consider a €350 interval.

From these figures, we document that WTP for the plane as opposed to the car does not vary with age, income, and gender or education level. The WTP for lodging at a 4-star hotel relative to an apartment slightly increases with age but it does not differ by income levels. Interestingly, the preference for a high-quality dwelling is higher among males and those with primary or secondary studies. As for the length of the stay, the preference for longer trips decreases with age, with those under 22 being the ones who are willing to pay more for a 7- or 10-day trip relative to a 3-days trip. We also see that the spread of the boxplot for length of stay is wider for those over 40. Additionally, females and those with high education exhibit larger WTP for length of stay. Conversely, these WTP do not differ by income.

[142] 22 years old (40) corresponds to the first (third) quartile of the age distribution. *Lowinc* gathers no monthly individual income, *midinc* takes value 1 if monthly individual income is between 500 and 1,500 € and *highinc* takes value 1 for monthly income over 1,500 €.

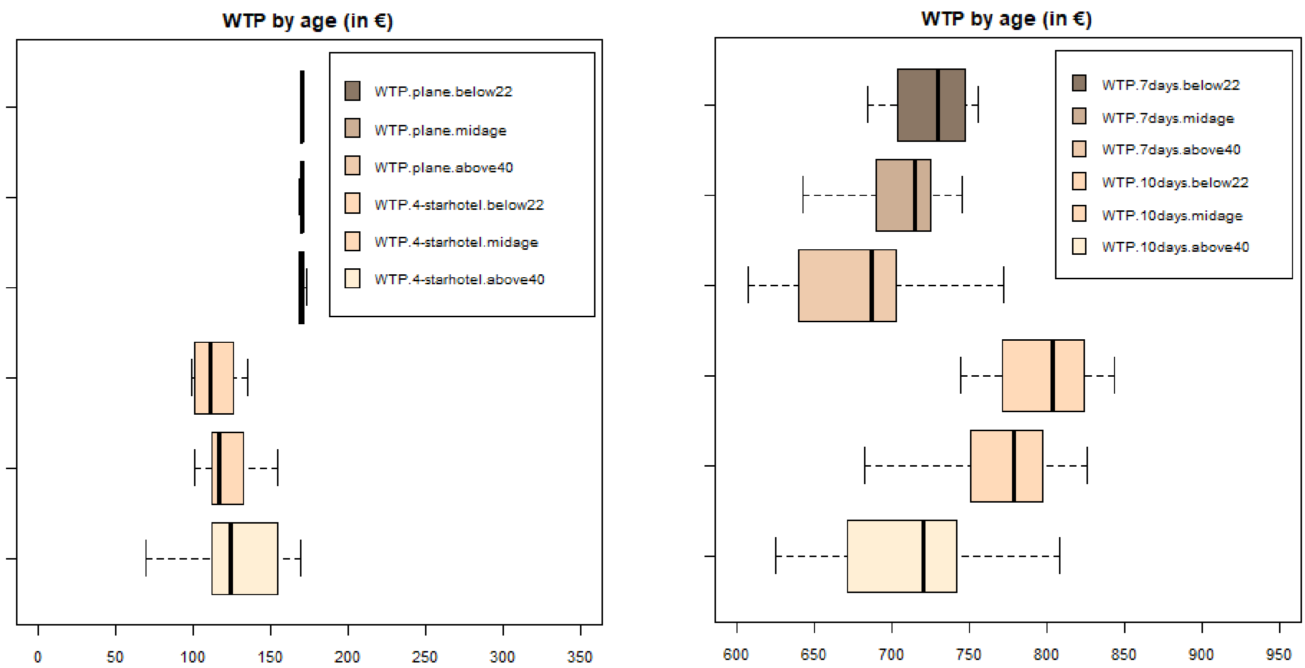


Figure 3.3.- Boxplot of WTP estimates for plane, 4starhotel, 7days and 10 days by age (in €)

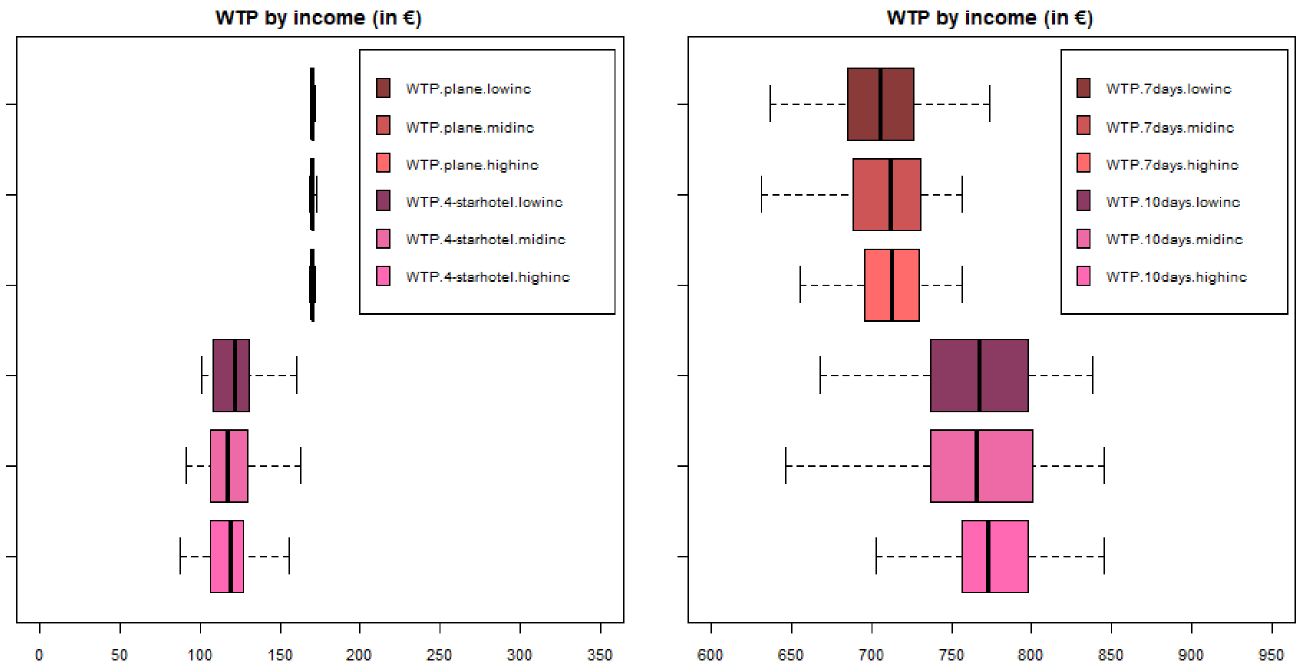


Figure 3.4.- Boxplot of WTP estimates for plane, 4starhotel, 7days and 10 days by income (in €)

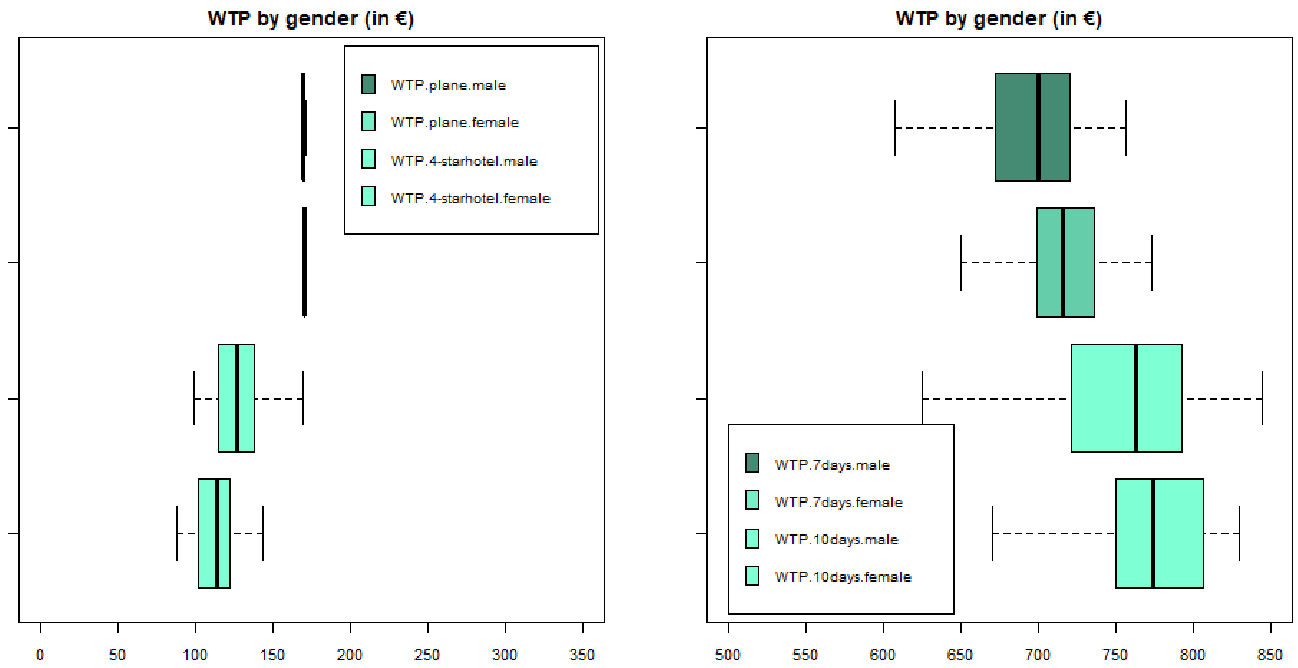


Figure 3.5.- Boxplot of WTP estimates for plane, 4starhotel, 7days and 10 days by gender (in €)

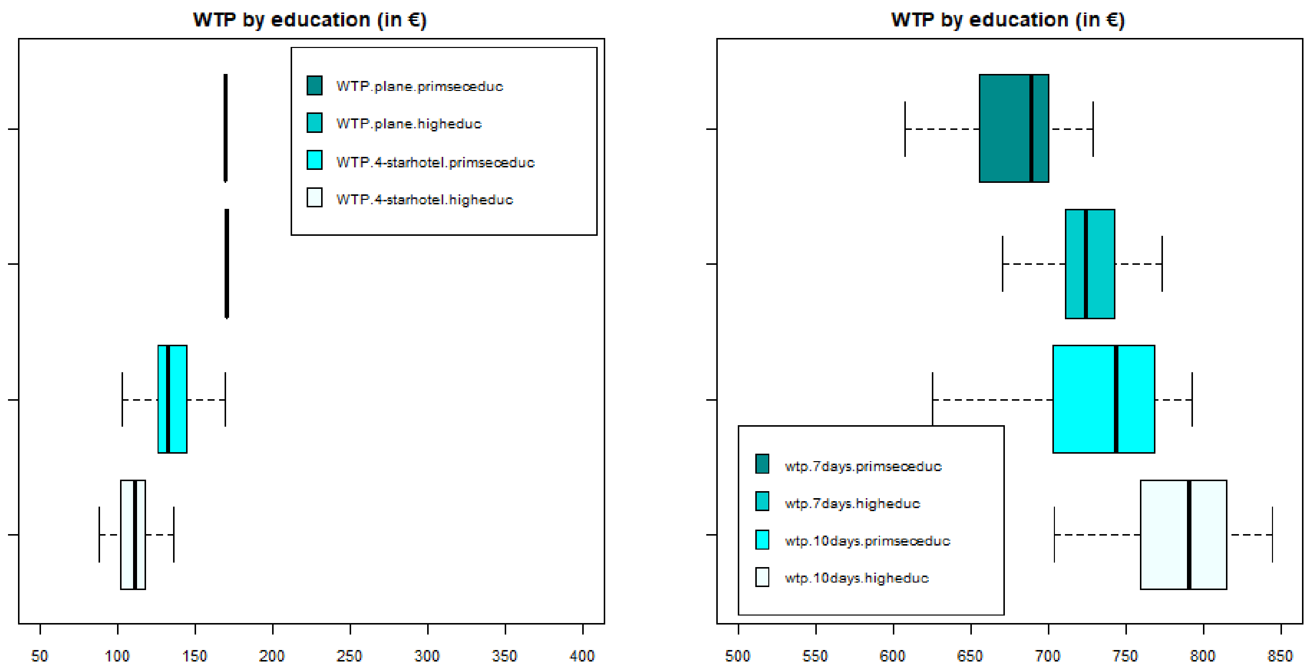


Figure 3.6.- Boxplot of WTP estimates for plane, 4starhotel, 7days and 10 days by gender (in €)

The use of the weighted WTP estimates as shown in (3.16) is the traditional way to study the MRS in monetary terms. The variability in the WTP stems from the individual-specific class allocation probabilities. However, the class allocation probabilities are time invariant in the sense that the information on the sequence of choices made is not considered. Most importantly, the weighted WTP introduced before are point estimates that do not inform on the confidence intervals. Because of this, we alternatively derive a conditional estimator of the individual specific WTP that considers the sequence of choices made by each respondent. We refer the reader to Annex 5 for the technical details and formulas.

Table 3.9 presents descriptive statistics of the means of the conditional WTP estimates together with a confidence interval (95% significance level). The confidence interval has been constructed as the mean of the conditional mean of the WTP for each attribute plus and minus 2 times the mean of the distribution of the conditional standard deviations for that attribute. This is similar to Craig et al. (2005), although they use 2.5 deviations instead of 2.

	Min	1st Q	Median	Mean	3rd Q	Max	95 % confidence Interval for the mean
WTP_{plane} (€)	166.3	166.9	171.2	170.3	171.9	176.8	(167.3; 173.3)
WTP_{7days} (€)	511.4	539.9	774.6	706.1	827.2	878.5	(558.7; 853.4)
WTP_{10days} (€)	499.8	540.3	859.1	764.2	938.6	963.2	(560.8; 967.6)
$WTP_{4starhotel}$ (€)	2.8	66.8	92.8	120.8	206.8	221.0	(45.6; 196)

Table 3.9.- Descriptive statistics for the distribution of the conditional means of the WTP values

The mean estimates of the *conditional* distribution are in line with those reported in Table 3.8 (the *unconditional* ones). However, the distribution of the estimates in Table 3.9 is wider. This may be due to both reduced sample size and reduced number of choice situations per respondents, which makes the computation of the conditional individual specific WTP estimates to be quite noisily. Consistent with Figures 3.4-3.7, we are precise at estimating the WTP for travelling by plane with respect to the use of car. Nevertheless, the distributions of the WTP for the length of the stay and for lodging at 4-star hotels (relative to an apartment) are wider. This also happens in the unconditional WTP. Correspondingly, aside from the potential noise in the derivation of the conditional WTP, it seems that the WTP are more heterogeneous for length of stay and high-quality accommodation that for travelling by plane.

5.4. Welfare analysis

There is some evidence that tourism imposes negative externalities to residents in the form of congestion, crime, noise or waste (e.g. Biagi and Delotto, 2014; Meleddu, 2014). In this respect, there is an ongoing discussion in the literature about different taxing schemes aimed at reducing the negative impact of tourism congestion on host communities (e.g. Gago et al., 2015).

Suppose that policy makers were considering the possibility of introducing a tourism tax of €1 per person and day only in one of the types of destinations considered in our analysis (e.g. coastal destination). How would individuals' welfare be affected by this? A theory-consistent way to explore this is to calculate the *Hicksian* compensating variation (i.e. the hypothetical transfer of money required to keep individuals at the same indifference curve they were before the tax setting).

The introduction of a daily tax per person in one of the alternatives is a price change that produces both a substitution (through altering choice probabilities) and an income effect. This leads to a change in consumer surplus that is obtained by integrating the uncompensated probabilistic demand for each alternative with respect to the price change, and then transforming it from *utils* to money. Consistent with Small and Rosen (1981), the compensating variation is given by:

$$CV = \frac{1}{\alpha}\left[\ln \sum_{j=1}^{J} exp^{V1} - \ln \sum_{j=1}^{J} exp^{V0}\right] \tag{3.18}$$

where $V1$ and $V0$ are the deterministic part of the utility function for each alternative after and before the tax, α is the marginal utility of money that converts the change in welfare into monetary values and $\ln \sum_{j=1}^{J} exp^{V}$ is the *logsum* or inclusive value that corresponds with the expected maximum utility level (Ben-Akiva and Lerman, 1985). As such, "the expected maximum utility is the sum of the utility of being in several states of the world weighted by the probability that each state occurs" (Lancsar and Savage, 2004, p. 903)[143]. Therefore, through the systematic component of utility (V_{ijt}), the compensating variation is a function of the attributes in the choice set (X_{ijt}) and the distribution of preferences in the population (β_c). In applied work, the marginal utility of income is assumed to equal the negative of the marginal disutility of cost (i.e. $\alpha = -\beta_{cost}$).

Since preferences are assumed to be heterogeneous, this price change would produce a distribution of compensating variations rather than a single value. Considering the probabilistic nature of class membership, the CV for three latent classes is given by:

$$CV_i = \sum_{c=1}^{C} \pi_i \frac{1}{-\beta_{cost|c}}\left[\ln \sum_{j=1}^{J} exp^{V1|c} - \ln \sum_{j=1}^{J} exp^{V0|c}\right] \tag{3.19}$$

Let us consider two possible scenarios. In Scenario 1 (SC1), think of a trip for 7 days with a total cost of €600, setting the rest of attributes to the base category. We introduce a daily tax per person in alternative j that translates into an additional cost of €7. In Scenario 2 (SC2), assume we have a 10-day trip with a cost of €1,000 and we introduce the same type of tax, also setting the rest of attributes to the base levels. Therefore, this would translate into an additional cost of €10 in the latter case.

We calculate the compensating variation in each scenario for each individual in the sample assuming the price change occurs: i) only for the coastal alternative, ii) only for the urban alternative, and iii) only for the nature-based alternative. Table 3.10 presents the CV monetary estimates (in euros) for each class and the overall CV weighted one for the three alternatives and the two scenarios following equation (3.19). The negative

[143] The derivative of the *logsum* with respect to the utility of any of the alternatives is that alternative's choice probability.

sign here indicates a reduction in welfare. A similar derivation of weighted CV for a latent class model can be found in Hynes et al. (2008) and Zhang and Sohngen (2018).

		Scenario 1 LOS=7 days $Cost_0$=€600 $Cost_1$= €607	Scenario 2 LOS=10 days $Cost_0$=€1,000 $Cost_1$= €1,010
Price change in coastal destination	CV Class 1	-1.86	-1.99
	CV Class 2	-4.79	-6.74
	CV Class 3	-1.43	-1.97
	Mean of Weighted CV	-2.99	-4.10
Price change in urban destination	CV Class 1	-2.04	-2.02
	CV Class 2	-1.39	-1.95
	CV Class 3	-3.10	-4.26
	Mean of Weighted CV	-2.19	-2.93
Price change in nature-based destination	CV Class 1	-0.80	-0.86
	CV Class 2	-0.68	-0.96
	CV Class 3	-2.19	-3.01
	Mean of Weighted CV	-1.33	-1.80

Table 3.10.- Compensating Variation estimates under the two scenarios

Individuals in Class 2 are the most affected by a tourism tax in coastal destinations in both scenarios. However, for the case of a tax setting either in the urban or in the nature-based alternative, individuals in Class 3 are who experience the largest welfare loss. Nevertheless, the magnitude of the loss in consumer surplus differs across the two scenarios. Focusing on the mean of the weighted CV, it can be seen that Scenario 2 corresponds to the highest amount of money needed to give respondents back to their original utility levels. Interestingly, the compensating variation is larger for a price change in the coastal alternative, followed by the urban and the nature-based ones. This is consistent both with descriptive statistics and the results from the MNL model in Table 3.7 showing a higher preference for this alternative, *ceteris paribus*. Additionally, the estimated compensating variations largely depart from the price change in each scenario. If choice probabilities were insensitive to price, then the CV would equal the value of the tourism tax.

Panels a) and b) in Figure 3.7 depict kernel density plots of the distribution of CV for Scenario 1 and 2, respectively. The horizontal axis is set to lie between -6 and -0.5 to facilitate comparison. Whereas the shape of the CV for price changes in urban and nature-based alternatives is similar in the two scenarios, the distribution of the CV for a tourism tax in a coastal destination is wider.

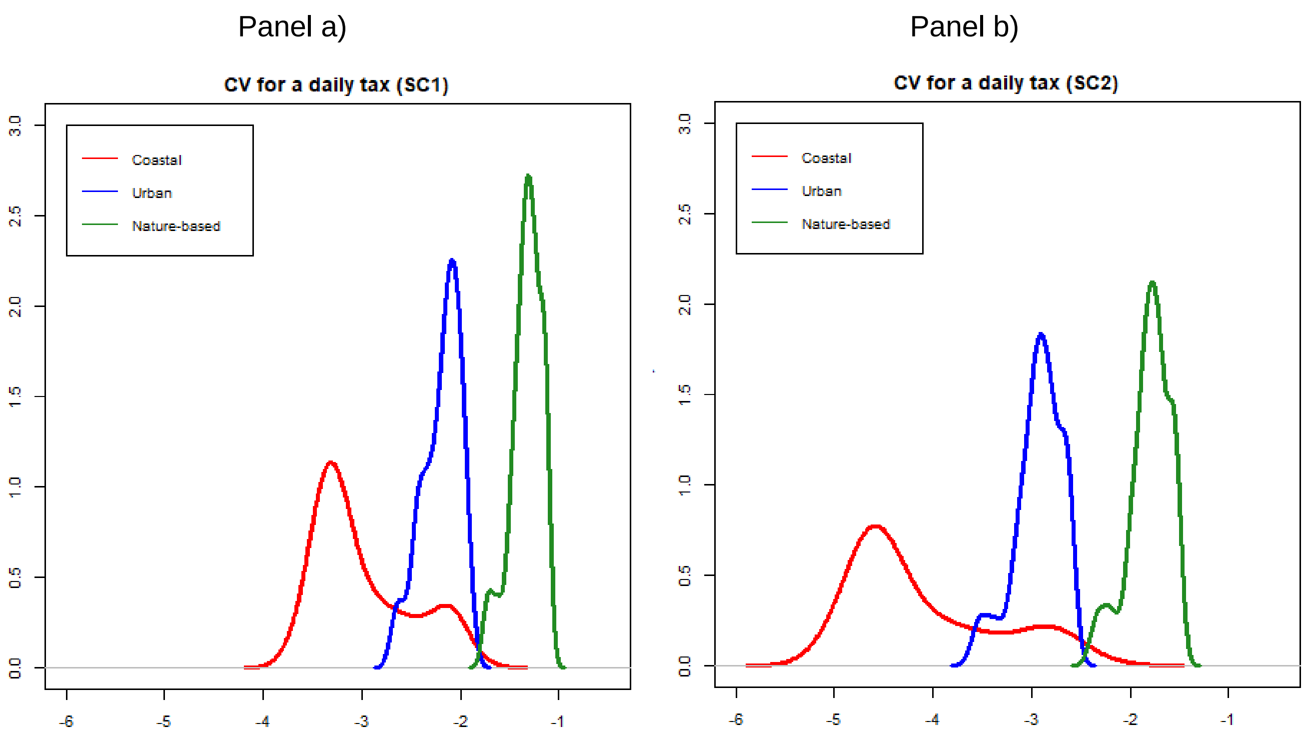


Figure 3.7.- Kernel density plots for CV for a daily tax in the two scenarios for the three destinations

Before ending this subsection, two important remarks deserve mention. First, the closed-form formula of the CV relies on the following assumptions: i) disturbances enter additively in the utility function and are independent of the explanatory variables, ii) the policy change does not affect the individual's draw for the iid Extreme Value distribution, and ii) the marginal utility of income is constant[144]. Second, we cannot ascertain which choice alternative would be chosen by a given respondent after the tax setting. Instead, we only assess the change in choice probabilities (probabilistic demand for a fixed quantity equal to one) for each alternative after the introduction of the tax only in one of them, *ceteris paribus*. More importantly, the analysis explicitly considers the change in the choice probabilities for the non-choice option. Hence, the estimated compensating variations take into account potential corner solutions after the price change.

6. DISCUSSION AND CONCLUSIONS

In this paper, we have conducted a DCE for studying the relative importance of vacation attributes in destination choice. Specifically, our aim has been to estimate the marginal rates of substitution among vacation attributes expressed in monetary terms (i.e. willingness to pay). To this end, we have recruited a sample of real-life couples and asked them to choose individually and separately in a lab their preferred option for a joint trip. This has been repeated in a series of six different choice tasks. Each choice card scenario is characterized by three alternatives, each one described by five attributes with

[144] See Morey and Rossman (2008) for a discussion on the first assumption and Herriges and Kling (1999) for a characterization of the CV under non-linear income effects.

different levels. The choice of the attributes and levels has been based on previous studies, comments received from focus groups and a pilot study.

We have estimated a Latent Class Model assuming discrete classes of preference groups in the sample. Our results identified three classes. About half of the sample is classified into one class, mainly composed of young males with moderate income and non-university education. These individuals assign a great value to the length of the stay, to travelling by plane with respect to the use of car and are highly sensitive to cost. The second class is composed of about 40% of the sample, also with relatively more males and young people. However, respondents in this class have non-university studies and higher income with greater likelihood. Individuals in this latter class place lower importance to travelling by plane, length of stay and cost relative to those in the former group, but derive positive utility from staying at a 4-star hotel. Finally, the remaining 10% belong to a third class comprising elderly females with high education and low income. This group does only pay attention to the length of the stay and to trip cost. Interestingly, none of the classes are estimated to give importance to travel time. Overall, participants attach greater importance to length of stay relative than to transportation or accommodation features. Therefore, we have shown that in the context of a summer trip individuals mainly demand time for recreation.

From the model estimates, we have derived a weighted average of the willingness to pay across classes, and a conditional estimator of the willingness to pay that considers the sequence of choices made by each respondent. Both procedures lead to similar results. We have found that respondents are willing to pay about €170 for plane travelling with respect to car, €120 for staying at 4-star hotel relative to an apartment, and €760 for a 10-day trip relative to a 3-day one. We have also explored the differences in willingness to pay based on sociodemographic characteristics. Consistent with the model estimates, the preference for high-quality accommodation is higher among males and those without university studies. The willingness to pay estimates for longer trip duration decreases with age but is higher among females and highly educated individuals.

As a final empirical exercise, we have computed the compensating variation for the hypothetical setting of a daily tax of €1 per person in each of the three destinations under two different scenarios. We have derived the distribution of compensating variations needed to give respondents back to their original utility levels. We have shown that, in line with the heterogeneity in preferences across classes, a daily tax in the coastal destination would lead to the largest loss in welfare. Nevertheless, the estimated compensating variations are always lower than the corresponding rise in prices, suggesting relevant substitution patterns in choice probabilities across destinations.

The study has some features that distinguish it from related studies. Contrary to data obtained from surveys about trips already made, the experimental procedure has allowed us to have information on the non-chosen options and to control for the environment and context in which choices are made. In contrast to other studies that use students or pre-recruited panellists, our sample comprises real-life married and non-married couples from the general population of four cities in Northern Spain. Also different from related studies, individuals are required to make their individual choices among three of the most common types of destinations (coastal, cultural and nature-based) rather than limiting the analysis to one of them.

Our main contribution is the econometric modelling of tourists' preferences. Different from similar choice experiments, we have derived both unconditional and conditional estimators of the willingness to pay estimates for the attribute levels that are individual-specific based on class membership probabilities. Therefore, we have explored the distribution of the monetary marginal rates of substitution in the sample and related them with the sociodemographic characteristics. Most similar studies lack this type of analysis and concentrate on the point estimates of the WTP within classes without exploring how it correlates with observable characteristics. This could be a valuable way for deepening into the sources of heterogeneity in preferences. Different from prior literature, based on the estimated probabilistic demands we have performed a simulation analysis of the welfare loss in case of a daily tourism tax. We have taken advantage of the experimental setting to create a market and study the impacts of a policy intervention.

Furthermore, prior to the main analysis, we have implemented a Monte Carlo simulation exercise to examine the sample size needed for reliable parameter identification. We recommend practitioners to conduct similar simulation studies to inspect the validity of the experimental design.

Our study possesses some limitations. First, young and educated people are slightly overrepresented and married couples underrepresented. We acknowledge this sampling bias as a potential limitation of this research. Second, our analysis of preferences is conditional on the alternatives and the attributes presented in the choice task. In real-life situations, couples have the possibility to find any suitable combination of vacation hedonic attributes (i.e. the choice set is larger). Nevertheless, it is precisely the impossibility of having information on the non-chosen alternatives what hinders the analysis of preferences using revealed preference data. We instead, encompass the study of marginal rates of substitution by exogenously restricting the choice set and identifying preferences conditional on that.

Although we collect information on choices made by the two members of the couple, we have devoted our attention to the estimation of individual preferences separately. In this respect, our results are robust to potential cross-sectional dependence at the couple level. However, our data does not allow us to separately identify pure individual preferences from altruistic or deferential ones. Further investigation on this and the modelling of joint choices is part of our future research agenda.

REFERENCES

Abdellaoui, M., l'Haridon, O. and Parashiv, C. (2013). "Individual vs. couple behavior: an experimental investigation of risk preferences". *Theory and Decision*, 75, 175-191.

Anderson, S.P., de Palma, A. and Thisse, J.F. (1987). "The CES is a discrete choice model?" *Economics Letters*, 24(2), 139-140.

Albadalejo-Pina, I.P. and Díaz-Delfa, M.T. (2009). "Tourist preferences for rural house stays: evidence from discrete choice modelling in Spain". *Tourism Management,* 30, 805-811.

Andersen, S., Harrison, G.W., Hole, A.R., Lau, M. and Rutström, E.E. (2012). "Non-linear mixed logit". *Theory and Decision*, 73, 77-96.

Angrist, J.D. and Pischke, J.S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Araña, J.E. and León, C.J. (2008). "The impact of terrorism on tourism demand". *Annals of Tourism Research*, 35(2), 299–315.

Araña, J.E., León, C.J., Carballo, M.M. and Moreno-Gil, S. (2016). "Designing tourist information offices: the Role of Human Factor". *Journal of Travel Research*, 55(6), 764-763.

Ashraf, N. (2009). "Spousal control and Intra-household decision making: an experimental study in the Philippines". *American Economic Review*, 99(4), 1245-1277.

Balbotin, C., Hensher, D.A. and Collins, A.T. (2017). "Is there a systematic relationship between random parameters and process heuristics?" *Transportation Research Part E*, 106, 160-177.

Barr, A., Dekker, M., Janssens, W., Kebede, B. and Kramer, B. (2019). "Cooperation in polygynous households". *American Economic Journal: Applied Economics*, 11(2), 266-283.

Bateman, I. and Munro, A. (2005). "An experiment on risky choice amongst households". *The Economic Journal*, 115, C176-C189.

Bateman, I.J. and Munro, A. (2009). "Household versus individual valuation: what's the difference?" *Environmental and Resource Economics*, 43, 119-135.

Beblo, M. and Beninger, D. (2017). "Do husbands and wives pool their incomes? A couple experiment". *Review of Economics of the Household*, 15, 779-805.

Beblo, M., Beninger, D., Cochard, F., Couprie, H. and Hopfensitz, A. (2015). "Equality-efficiency trade-off within French and German Couples - A comparative experimental study". *Annals of Economics and Statistics*, 117/118, 233-252.

Bech, M. and Gyrd-Hansen, D. (2005). "Effects coding in discrete choice experiments". *Health Economics*, 14, 1079-1083.

Beck, M.J., Rose, J.M. and Hensher, D. (2013). "Consistently inconsistent: the role of certainty, acceptability and scale in choice". *Transportation Research Part E,* 56, 81-93.

Beck, M.J., Fifer, S. and Rose, J.M. (2016). "Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty". *Transportation Research Part B*, 89, 149-167.

Beharry-Borg, N., Hensher, D.A. and Scarpa, R. (2009). "An analytical framework for joint vs separate decisions by couples in choice experiments: the case of coastal water quality in Tobago". *Environmental Resource Economics*, 43, 95-117.

Ben-Akiva, M., McFadden, D. and Train, K. (2019). *Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis*. In: Foundations and Trends in Econometrics: Vol. 10, No. 1-2, pp 1–144.

Ben-Akiva, M.E. and Lerman, S.R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT press.

Biagi, B. and Delotto, C. (2014). "Crime as tourism externality". *Regional Studies*, 48(4), 693-709.

Bimonte, S. Ferrini, S. and Grilli, G. (2016). "Transport infrastructures, environmental impacts and tourists' welfare: a choice experiment to elicit tourists' preferences in Siena-Italy". *Journal of Environmental Planning and Management*, 59(5), 891-910.

Bliemer, M.C.J., Rose, J.M. and Hess, S. (2008). "Approximation of Bayesian efficiency in experimental choice designs". *Journal of Choice Modelling*, 1(1), 98-126.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N. and Freeman, P. (2008). "Eliciting willingness to pay without bias: evidence from a field experiment". *Economic Journal*, 118(525), 114–137.

Boldt, L. and Arora, N. (2017). "Dyadic compromise effect". *Marketing Science*, 36(3), 436-452.

Börger, T. (2016). "Are fast responses more random? Testing the effect of response time on scale in an online choice experiment". *Environmental and Resource Economics*, 65, 389-413.

Boxall, P. and Adamowicz, W.L. (2002). "Understanding heterogeneous preferences in random utility models: a latent class approach". *Environmental and Resource Economics*, 23, 421-446.

Brock, W. and Durlauf, S. (2001). "Discrete choice with social interactions". *Review of Economic Studies*, 68, 235–260.

Bujosa, A., Riera, A. and Hicks, R.L. (2010). "Combining discrete and continuous representations of preference heterogeneity: a latent class approach". *Environmental and Resource Economics*, 47, 477-493.

Bujosa, A., Torres, C. and Riera, A. (2018). "Framing decisions in uncertain scenarios: an analysis of tourist preferences in the face of global warming". *Ecological Economics*, 148, 36-42.

Carballo, M.M., Araña, J.E., León, C.E. and Moreno-Gil, S. (2015). "Economic valuation of tourism destination image". *Tourism Economics*, 21(4), 741-759.

Carlsson, F. and Martinsson, P. (2003). "Design techniques for stated preference methods". *Health Economics*, 12, 281-294.

Carlsson, F., He, H., Martinsson, P., Qin, P. and Sutter, M. (2012). "Household decision making in rural China: using experiments to estimate the influences of spouses". *Journal of Economic Behavior and Organization*, 84(2), 525–536.

Carlsson, F., Martinsson, P., Qin, P. and Sutter, M. (2013). "The influence of spouses on household decision making under risk: an experiment in rural China". *Experimental Economics*, 16(3), 383–401.

Carson, R.T., Louviere, J.J., Anderson, D.A., Arabie, P., Bunch, D.S., Hensher, D.A., Johnson, R.M., Kuhfeld, W.F., Steinberg, D., Swait, J., Timmermans, H. and Wiley, J.B. (1994). "Experimental analysis of choice". *Marketing Letters*, 5(4), 351-368.

Castilla, C. (2019). "What's yours is mine, and what's mine is mine: Field experiment on income concealing between spouses in India". *Journal of Development Economics*, 137, 125-140.

Caussade, S., Ortuzar, J. de D., Rizzi, L. I., and Hensher, D. A. (2005). "Assessing the influence of design dimensions on stated choice estimates". *Transportation Research Part B*, 39, 621–640.

Chaminuka, P., Groeneveld, R.A., Selomane, A.O. and Van Ierland, E.C. (2012). "Tourist preferences for ecotourism in rural communities adjacent to kruger national park: a choice experiment approach". *Tourism Management*, 33(1), 168–176.

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.

Chen, N., Masiero, L. and Hsu, C.H.C. (2019). "Chinese outbound tourist preferences for all-inclusive group package tours: a latent class choice model". *Journal of Travel Research*, 58(6), 916-931.

Chiappori, P.A. (1992). "Collective labor supply and welfare". *Journal of Political Economy*, 100(3), 437-463.

Choi, A.S., Ritchie, B.W., Papandrea, F. and Bennett, J. (2010). "Economic valuation of cultural heritage sites: a choice modeling approach". *Tourism Management*, 31(2), 213–220.

ChoiceMetrics (2012). *Ngene 1.1.1*, User Manual and Reference Guide.

Chorus, C.G. (2010). "A new model of random regret minimization". *European Journal of Transport and Infrastructure Research*, 10 (2), 181–196.

Chorus, C., Van Cranenburgh, S. and Dekker, T. (2014). "Random regret minimization for consumer choice modelling: assessment of empirical evidence". *Journal of Business Research*, 67(11), 2428-2436.

Chung, C., Boyer, T. and Han, S. (2011). "How many choice sets and alternatives are optimal? Consistency in Choice Experiments". *Agribusiness*, 27(1), 114-125.

Christie, M., and Gibbons, J. (2011). "The effect of individual 'ability to choose' (scale heterogeneity) on the valuation of environmental goods". *Ecological Economics*, 70, 2250–2257.

Cochard, F., Couprie, H. and Hopfensitz, A. (2016). "Do spouses cooperate? An experimental investigation". *Review of Economics of the Household*, 14(1), 1–26.

Cochard, F., Couprie, H. and Hopfensitz, A. (2018). "What if women earned more than their spouses? An experimental investigation of work-division in couples". *Experimental Economics*, 21, 50-71.

Concu, N. and Atzeni, G. (2012). "Conflicting preferences among tourists and residents". *Tourism Management*, 33, 1293-1300.

Couprie, H., Cudeville, E. and Sofer, C. (2020). "Efficiency versus gender roles and stereotypes: an experiment in domestic production". *Experimental Economics*, 23, 181-211.

Craig, C.S., Greene, W.H. and Douglas, S.P. (2005). "Culture matters: consumer acceptance of US films in foreign markets". *Journal of International Marketing*, 13(4), 80–103.

Crouch, G.I., Devinney, T.M., Louviere, J.J. and Islam, T. (2009). "Modelling consumer choice behavior in space tourism". *Tourism Management*, 30, 441-454.

Crouch, G.I., Del Chiappa, G. and Perdue, R.R. (2019). "International convention tourism: a choice modelling experiment of host city competition". *Tourism Management*, 71, 530-542.

Czajkowski, M. and Budzinski, W. (2019). "Simulation error in maximum likelihood estimation of discrete choice models". *Journal of Choice Modelling*, 31, 73-85.
Czajkowski, M. and Hanley, N. (2009). "Using labels to investigate scope effects in stated preference methods". *Environmental and Resource Economics*, 44, 521-535.
Czajkowski, M., Hanley, N. and LaRiviere, J. (2014). "The effects of experience on preferences: theory and empirics for environmental public goods". *American Journal of Agricultural Economics*, 97(1), 333-351.
Czajkowski, M., Hanley, N. and LaRiviere, J. (2016). "Controlling for the effects of information in a public goods discrete choice model". *Environmental and Resource Economics*, 63, 523-544.
Czajkowski, M., Glerglczny, M., Kronenverg, 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.
Czajkowski, M., Giergiczny, M. and Greene, W.H. (2014). "Learning and fatigue effects revisited: investigating the effects of accounting for unobservable preference and scale heterogeneity". *Land Economics*, 90(2), 324-351.
Daly, A., Dekker, T. and Hess, S. (2016). "Dummy coding vs effects coding for categorical variables: clarifications and extensions". *Journal of Choice Modelling*, 21, 36-41.
Danaf, M., Atasoy, B. and Ben-Akiva, M. (2019). "Logit mixture with inter and intra-consumer heterogeneity and flexible mixing distributions". *Journal of Choice Modelling*, https://doi.org/10.1016/j.jocm.2019.100188
De Palma, A., Myers, G. and Papageorgiou, Y. (1994). "Rational choice under an imperfect ability to choose". *American Economic Review*, 84, 419-440.
De Palma, A., Picard, N. and Ziegelmeyer, A. (2011). “Individual and couple decision behavior under risk: evidence on the dynamics of power balance”. *Theory and Decision*, 70(1), 45-64.
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.
Dellaert, B., Borgers, A. and H. Timmermans. (1995). “A day in the city: using conjoint choice experiments to model urban tourists’ choice of activity packages”. *Tourism Management*, 16(5), 347-53.
Dellaert, B.G.C., Borgers, A.W.J and Timmermans, H.J.P. (1997). "Conjoint models of tourist portfolio choice: theory and illustration". *Leisure Sciences*, 19, 31-58.
Dennis, J.M. (2001). “Are internet panels creating professional respondents?” *Marketing Research*, 13, 34–38.
DeShazo, J.R. and Fermo, G. (2002). “Designing choice sets for stated preference methods: the effects of complexity on choice consistency”. *Journal of Environmental Economics and Management*, 44, 123–143.
Dieckmann, A., Dippold, K. and Dietrich, H. (2009). "Compensatory versus non-compensatory models for predicting consumer preferences". *Judgement and Decision Making*, 4(3), 200-213.
Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2010). "Are risk aversion and impatience related to cognitive ability?" *American Economic Review*, 100, 1238-1260.
Eppink, F.V., Hanley, N. and Tucker, S. (2019). “How best to present complex ecosystem information in stated preference studies?” *Ecological Economics*, 158, 20-25.
Fernández-Antolín, A., Guevara, C.A., Lapparent, M. and Bierlaire, M. (2016). "Correcting for endogeneity due to omitted attitudes: empirical assessment of a modified MIS method using RP mode choice data". *Journal of Choice Modelling*, 20, 1-15.
Fershtman, C. and Segal, U. (2018). "Preferences and social influence". *American Economic Journal: Microeconomics*, 10(3), 124-142.
Figini, P. and Vici, L. (2012). "Off-season tourists and the cultural offer of a mass-tourism destination: the case of Rimini". *Tourism Management*, 33, 825-839.
Fleischer, A., Tchetchik, A. and Toledo, T. (2012). “The Impact of fear of flying on travelers’ flight choice: choice model with latent variables”. *Journal of Travel Research*, 51(5), 653–663.
Gago, A., Labandeira, X., Picos, F. and Rodríguez, M. (2009). "Specific and general taxation of tourism activities. Evidence from Spain". *Tourism Management*, 30, 381-392.
Glenk, K. and Colombo, S. (2013). "Modelling outcome-related risk in choice experiments". *Australian Journal of Agricultural and Resource Economics*, 57, 559-578.
Glenk, K., Meyerhoff, J., Akaichi, F. and Martin-Ortega, J. (2019). "Revisiting cost vector effects in discrete choice experiments". *Resource and Energy Economics*, 57, 135-155.
Gnagey, M.K., Grijalva, T.C. and Rong, R. (2018). "Spousal dictator game: household decisions and other-regarding preferences". *Games*, 9(3), 1-11.

González-Valdes, F. and Raveau, S. (2018). "Identifying the presence of heterogeneous discrete choice heuristics at an individual level". *Journal of Choice Modelling*, 28, 28-40.
González-Valdés, F. and Ortúzar, J. (2018). “The stochastic satisficing model: a bounded rationality discrete choice model”. *Journal of Choice Modelling*, 27, 74-87.
Görges, L. (2015). "The power of love: A subtle driving force for unegalitarian labor division?" *Review of Economics of the Household*, 13, 163–192.
Greene, W.H. (2012). *Econometric Analysis, 7th Edition*. Chapter 15: Simulation-based estimation and inference and random parameter models. Prentice Hall.
Greene, W. H., and Hensher, D. A. (2003). "A latent class model for discrete choice analysis: contrasts with mixed logit". *Transportation Research Part B*, 37, 681–698.
Greene, W.H. and Hensher, D.A. (2013). "Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model". *Applied Economics*, 45, 1897-1902.
Greene, W., Harris, M.N. and Spencer, C. (2014). "Estimating the Standard Errors of Individual-specific Parameters in Random Parameters Models". Bankwest Curtin Economics Centre. Working paper.
Greiner, B. (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". *Journal of the Economic Science Association*, 1(1), 114-125.
Grigolon, A.B., Kemperman, A.D.A.M. and Timmermans, H.J.P. (2012). "The influence of low-fare airlines on vacation choices of students: results of a stated portfolio choice experiments". *Tourism Management*, 33, 1174-1184.
Guevara, C.A. (2015). "Critical assessment of five methods to correct for endogeneity in discrete-choice models". *Transportation Research Part A*, 82, 240-254.
Guevara, C.A. (2017). "Mode-valued differences of in-vehicle travel time savings". *Transportation*, 44(5), 977-997.
Haab, T.C. and Hicks, R.L. (1997). “Accounting for choice set endogeneity in random utility models for recreation demand”. *Journal of Environmental Economics and Management*, 34, 127-147.
Haener, M.K., Boxall, P.C. and Adamowicz, W.L. (2001). "Modelling recreation site choice: do hypothetical choices reflect actual behavior?". *American Journal of Agricultural Economics*, 83(3), 629-642.
Hanemann, W.M. (1984). "Discrete/continuous models of consumer demand". *Econometrica*, 52(3), 541-561.
Hausman, J.A. and Wise, D.A. (1978). "A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences". *Econometrica*, 46(2), 403-426.
He, H., Martinsson, P. and Sutter, M. (2012). "Group decision making under risk: an experiment with student couples". *Economics Letters*, 117, 691-693.
Hearne, R.R. and Salinas, Z.M. (2002). “The use of choice experiments in the analysis of tourist preferences for ecotourism development in Costa Rica”. *Journal of Environmental Management*, 65, 153-163.
Heiner, R. (1983). "The origin of predictable behavior". *American Economic Review*, 73, 560-595.
Heinrich, T. and Mayrhofer, T. (2018). "Higher-order risk preferences in social settings". *Experimental Economics*, 21(2), 434-456.
Hensher, D. A. (2010). “Hypothetical bias, choice experiments and willingness to pay”. *Transportation Research Part B*, 44, 735–752.
Hensher, D. A. and Barnard, P. O. (1990). “The orthogonality issue in stated choice designs, in: M. Fischer, P. Nijkamp and Y. Papageorgiou (Eds) Spatial Choices and Processes, pp. 265–278 (Amsterdam: North-Holland).
Hensher, D.A., Greene, W.H. and Li, Z. (2011). "Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings". *Transportation Research Part B*, 45, 954-972.
Hensher, D.A. and Rose, J.M. (2007). "Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: a case study". *Transportation Research Part A*, 41, 428-443.
Herriges, J.A. and Kling, C.L. (1999). "Nonlinear income effects in random utility models". *The Review of Economics and Statistics*, 81(1), 62-72.
Hess, S., Adler, T. and Polak, J.W. (2007). "Modelling airport and airline choice behavior with the use of stated preference survey data". *Transportation Research Part E*, 43, 221-233.
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., Stathopoulos, A. and Daly, A. (2012). “Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies”. *Transportation*, 39(3), 565–591.
Hess, S., Daly, A. and Batley, R. (2018). "Revisiting consistency with random utility maximization: theory and implications for practical work". *Theory and Decision*, 84, 181-204.

Hess, S., Spitz, G., Bradley, M. and Coogan, M. (2018). "Analysis of mode choice for intercity travel: application of a hybrid choice model to two distinct US corridors". *Transportation Research Part A*, 116, 547-467.
Hess, S. and Train, K.E. (2011). "Recovery of inter- and intra-personal heterogeneity using mixed logit models". *Transportation Research Part B*, 45(7), 973-990.
Hess, S. and Palma, D. (2019). "Apollo: a flexible, powerful and customizable freeware package for choice model estimation and application". *Journal of Choice Modelling*, https://doi.org/10.1016/j.jocm.2019.100170
Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V. and Caussade, S. (2013). "It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity". *Transportation*, 40, 583-607.
Hole, A.R. (2007a). "Fitting mixed logit models by using maximum simulated likelihood". *The Stata Journal*, 7(3), 388-401.
Hole, A.R. (2007b). "A comparison of approaches to estimating confidence intervals for willingness to pay measures". *Health Economics*, 16, 827-840.
Hole, A.R. (2011). "A discrete choice model with endogenous attribute attendance". *Economic Letters*, 110, 203-205.
Hole, A.R. and Kolstad, J.R. (2012). "Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment". *Empirical Economics*, 42, 445-469.
Hole, A.R., Kolstad, J.R. and Gyrd-Hansen, D. (2013). "Inferred vs. stated attribute non-attendance in choice experiments: a study of doctors' prescription behavior". *Journal of Economic Behavior and Organization*, 96, 21-31.
Hoyos, D. (2010). "The state of the art of environmental valuation with discrete choice experiments". *Ecological Economics*, 69(8), 1595-1603.
Hoyos, D., Mariel, P. and Hess, S. (2015). "Incorporating environmental attitudes in discrete choice models: an exploration of the utility of the awareness of consequences scale". *Science of the Total Environment*, 505, 1100-1111.
Huber, J. and Train, K. (2001). "On the similarity of classical and bayesian estimates of individual mean partworths". *Marketing Letters*, 12(3), 259-269.
Huybers, T. (2003). "Modelling short-break holiday destination choices". *Tourism Economics*, 9(4), 389-405.
Hynes, S., Hanley, N. and Scarpa, R. (2008). "Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models". *American Journal of Agricultural Economics*, 90(4), 1011-1027.
Johnston, R.J., Boyle, K.J, Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., and Vossler, C.A. (2017). "Contemporary guidance for stated preference studies". *Journal of the Association of Environmental and Resource Economists*, 4(2), 319-405.
Juutinen, A., Mitani, Y., Mäntymaa, E., Shoji, Y., Siikamäki, P. and Svento, R. (2011). "Combining ecological and recreational aspects in national park management: a choice experiment application". *Ecological Economics*, 70, 1231-1239.
Kamakura, W. and Russell, G. (1989). "A probabilistic choice model for market segmentation and elasticity structure". *Journal of Marketing Research*, 26, 379-390.
Karl, M., Reintinger, C., and Schmude, J. (2015). “Reject or select: mapping destination choice”. *Annals of Tourism Research*, 54, 48–64.
Keane, M. and Wasi, N. (2013). "Comparing alternative models of heterogeneity in consumer choice behaviour". *Journal of Applied Econometrics*, 28, 1018-1045.
Kelly, J., Haider, W., Williams, P. W., and Englund, K. (2007). "Stated preferences of tourists for eco-efficient destination planning options". *Tourism Management*, 28(2), 377-390.
Kim, D. and Park, B.R. (2017). "The moderating role of context in the effects of choice attributes on hotel choice: a discrete choice experiment". *Tourism Management*, 63, 439-451.
Kneeland, T. (2015). "Identifying higher-order rationality". *Econometrica*, 83(5), 2065-2079.
Koo, T.T.R., Caponecchia, C. and Williamson, A. (2018). "How important is safety in making flight choices? Evidence from simple choice experiments". *Transportation*, 45, 159-175.
Kragt, M.E. (2013). "The effects of changing cost vectors on choices and scale heterogeneity". *Environmental and Resource Economics*, 54, 201-221.
Kragt, M.E. and Bennett, J.W. (2012). "Attribute framing in choice experiments: how do attribute level descriptions affect value estimates?" *Environmental and Resource Economics*, 51, 43-59.
Lancaster, K.J. (1966). “A new approach to consumer theory”. *Journal of Political Economy*, 74, 132–157.

Lancsar, E. and Louviere, J. (2008). "Conducting discrete choice experiments to inform healthcare decision making. A users' guide". *Pharmacoeconomics*, 26, 661-677.
Lancsar, E. and Savage, E. (2004). "Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory". *Health Economics*, 13, 901-907.
Lancsar, E., Fiebig, D.G. and Hole, A.R. (2017). "Discrete choice experiments: a guide to model specification, estimation and software". *Pharmacoeconomics*, 35, 697-716.
Landry, C.E. and List, J.A. (2007). "Using ex ante approaches to obtain credible signals for value in contingent markets: evidence from the field". *American Journal of Agricultural Economics*, 89, 420–429.
Larson, D.M. and Lew, D.K. (2005). "Measuring the utility of ancillary travel: revealed preferences in recreation site demand and trips taken". *Transportation Research Part A*, 39(2-3), 237-255.
León, C.J., de León, J., Araña, J.E. and González, M.M. (2015). "Tourists' preferences for congestion, residents' welfare and the ecosystems in a national park". *Ecological Economics*, 118, 21-29.
Li, C., McCabe, S. and Song, H. (2017). "Tourist choice processing: evaluating decision rules and methods of their measurement". *Journal of Travel Research*, 56 (6), 699-711.
Liebe, U., Mariel, P., Beyer, H. and Meyerhoff, J. (2019). "Uncovering the nexus between attitudes, preferences, and behavior in sociological applications of stated choice experiments". *Sociological Methods and Research*, https://doi.org/10.1177/0049124118782536
List, J.A. and Gallet, C.A. (2001). "What experimental protocol influences disparities between actual and hypothetical stated values?". *Environmental and Resource Economics*, 20, 241-254.
Logar, I. and Brouwer, R. (2018). "Substitution effects and spatial preference heterogeneity in single- and multiple-site choice experiments". *Land Economics*, 94(2), 302-322.
Louviere, J.J. and Hensher, D.A. (1982). "Design and analysis of simulated choice or allocation experiments in travel choice modeling". *Transportation Research Record*, 890, 11–17.
Louviere, J.J., Woodworth, G. (1983). "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data". *Journal of Marketing Research*, 20, 350–357.
Louviere, J.J., Islam, T., Wasi, N., Street, D. and Burgess, L. (2008). "Designing discrete choice experiments: do optimal designs come at a price?" *Journal of Consumer Research*, 35, 360-375.
Maniadis, Z., Tufano, F. and List, J.A. (2014). "One swallow doesn't make a summer: new evidence on anchoring effects". *American Economic Review*, 104(1), 277-290.
Manski, C.F. (2004). "Measuring expectations". *Econometrica*, 72(5), 1329-1376.
Mariel, P., Scarpa, R. and Vega-Bayo, A. (2018). "Joint parental school choice: exploring the influence of individual preferences of husbands and wives". *Regional Science and Urban Economics*, 68, 23-35.
Masiero, L., and Nicolau, J.L. (2012). "Tourism market segmentation based on price sensitivity finding similar price preferences on tourism activities". *Journal of Travel Research*, 51(4), 426–435.
Masiero, L., Yang, Y. and Qiu, R.T.R. (2019). "Understanding hotel location preference of consumers: comparing random utility and random regret decision rules". *Tourism Management*, 73, 83-93.
Matejka, F. and McKay, A. (2015). "Rational inattention to discrete choices: a new foundation for the multinomial logit model". *American Economic Review*, 105(1), 272-298.
Matthews, Y., Scarpa, R. and Marsh, D. (2017). "Using virtual environments to improve the realism of choice experiments: a case study about coastal erosion management". *Journal of Environmental Economics and Management*, 81, 193-208.
Matveenko, A. (2020). "Logit, CES, and rational inattention". *Economics Letters*, https://doi.org/10.1016/j.econlet.2019.108537
McFadden, D. (1974) *Conditional logit analysis of qualitative choice behavior.* In Frontiers in Econometrics (pp. 105–142). New York: Academic Press.
McFadden, D. (1981). *Econometric Models of Probabilistic Choice*. In: Structural Analysis of Discrete Data (pp. 198–272). Ed. by C. Manski and D. McFadden. Cambridge: MIT Press.
McFadden, D. (2000). "Disaggregate behavioral travel demand's RUM side—a 30 years retrospective". In D.A. Hensher (Ed.), *Travel behaviour research* (pp. 17–63). Amsterdam: Elsevier.
McFadden, D. (2001). "Economic choices". *American Economic Review*, 91(3), 351-378.
McFadden, D. and Train, K. (2000). "Mixed MNL models for discrete response". *Journal of Applied Econometrics*, 15, 447-470.
McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. New York: Jon Wiley and Sons.
McNair, B.J., Hensher, D.A. and Bennett, J. (2012). "Modelling heterogeneity in response behavior towards a sequence of discrete choice questions: a probabilistic decision process model". *Environmental and Resource Economics*, 51, 599–616.

Meleddu, M. (2014). "Tourism, residents' welfare and economic choice: a literature review". *Journal of Economic Surveys*, 28(2), 376-399.
Mokhtarian, P.L. (2005). "Travel as a desired end, not just a means". *Transportation Research Part A*, 39(2), 93–96.
Mørkbak, R.M., Christensen, T. and Gyrd-Hansen, D. (2010). "Choke price bias in choice experiments". *Environmental and Resource Economics*, 45, 537–551.
Morey, E. and Rossmann, K.G. (2008). "Calculating, with income effects, the compensating variation for a state change". *Environmental and Resource Economics*, 39, 83-90.
Morley, C. L. (1994). "Discrete choice analysis of the impact of tourism prices". *Journal of Travel Research*, 33, 8-14.
Munro, A. and Popov, D. (2013). "A portmanteau experiment on the relevance of individual decision anomalies for households". *Experimental Economics*, 16(3), 1–14.
Nickerson, N.P. and Jurowski, C. (2001). "The influence of children on vacation travel patterns". *Journal of Vacation Marketing*, 7(1), 19-30.
Nicolau, J.L. (2012). "Asymmetric tourist response to price: loss aversion segmentation". *Journal of Travel Research*, 51(5), 568-576.
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.
Nguyen, Q. (2016). "Linking loss aversion and present bias with overspending behavior of tourists: insights from a lab-in-the field experiment". *Tourism Management*, 54, 152-159.
Oehlmann, M., Meyerhoff, J., Mariel, P. and Weller, P. (2017). "Uncovering context-induced status quo effects in choice experiments". *Journal of Environmental Economics and Management*, 81, 59-73.
Oppewal, H., Huybers, T. and Crouch, G.I. (2015). "Tourist destination and experience choice: a choice experimental analysis of decision sequence effects". *Tourism Management*, 48, 467-476.
Oviedo, J.L. and Yoo, H.I. (2017). "A latent class nested logit model for rank-ordered data with application to cork oak reforestation". *Environmental and Resource Economics*, 68(4), 1021-1051.
Pacífico, D. and Yoo, H. (2013). "lclogit: a Stata command for fitting latent-class conditional logit models via the EM algorithm". *The Stata Journal*, 13(3), 625-639.
Park, J. Y., and Jang, S. S. (2013). "Confused by too many choices? Choice overload in tourism". *Tourism Management*, 35, 1–12.
Park, S. and Gupta, S. (2012). "Handling endogenous regressors by joint estimation using copulas". *Marketing Science*, 31(4), 567-586.
Petrin, A. and Train, K. (2010). "A control function approach to endogeneity in consumer choice models". *Journal of Marketing Research*, 47(1), 3-13.
Pollak, R.A. (2003). "Gary Becker's contributions to family and household economics". *Review of Economics of the Household*, 1, 111-141.
Provencher, B. and Bishop, R.C. (2004). "Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study". *Journal of Environmental Economics and Management*, 48, 793-810.
Provencher, B. and Moore, R. (2006). "A discussion of ''using angler characteristics and attitudinal data to identify environmental preferences classes: a latent class model". *Environmental and Resource Economics*, 34(1), 117–124.
Robinson, J. (2012). "Limited insurance within the household: evidence from a field experiment in Kenya". *American Economic Journal: Applied Economics*, 4(4), 140-164.
Rolfe, J., and Bennet, J. (2009). "The impact of offering two versus three alternatives in choice modelling experiments". *Ecological Economics*, 68, 1140-1148.
Rolfe, J. and Windle, J. (2015). "Do respondents adjust their expected utility in the presence of an outcome certainty attribute in a choice experiment?" *Environmental and Resource Economics*, 60, 125-142.
Rose, J. and Bliemer, M.C.J. (2009). "Constructing efficient stated choice experimental designs". *Transport Reviews*, 29(5), 587-617.
Rose, J.M., Bliemer, M.C.J., Hensher, D. and Collins, A.T. (2008). "Designing efficient stated choice experiments in the presence of reference alternatives". *Transportation Research Part B*, 42, 395-406.
Rungie, C., Scarpa, R. and Thiene, M. (2014). "The influence of individuals in forming collective household preferences for water quality". *Journal of Environmental Economics and Management*, 68, 161-174.
Samuelson, P. (1950). "The problem of integrability in utility theory". *Economica*, 17, 355–385.

Sándor, Z. and Wedel, M. (2001). “Designing conjoint choice experiments using managers' prior beliefs”. *Journal of Marketing Research*, 38(4), 430-444.
Sándorf, E.D., Sourd, R.C. and Mahieu, P.A. (2018). "The effect of attribute-alternative matrix displays on preferences and processing strategies". *Journal of Choice Modelling*, 29, 113-132.
Sarrias, M. and Daziano, R.A. (2017). "Multinomial logit models with continuous and discrete individual heterogeneity in R: the gmnl package". *Journal of Statistical Software*, 79(2), 1-46.
Sarrias, M. and Daziano, R.A. (2018). "Individual-specific point and interval conditional estimates of latent class logit parameters". *Journal of Choice Modelling*, 27, 50-61.
Scarpa, R. and Thiene, M. (2005). “Destination choice models for rock climbing in the northeastern alps: a latent-class approach based on intensity of preferences”. *Land Economics*, 81(3), 426-444.
Scarpa, R. and Rose, J.M. (2008). "Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why". *The Australian Journal of Agricultural and Resource Economics*, 52, 253-282.
Scarpa, R., Gilbride, T. J., Campbell, D. and Hensher, D. A. (2009). "Modelling attribute non-attendance in choice experiments for rural landscape valuation". *European Review of Agricultural Economics*, 36(2), 151–174.
Scarpa, R., Thiene, M. and Train, K. (2008). "Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps". *American Journal of Agricultural Economics*, 90(4), 994-1010.
Schuhmann, P.W., Bass, B.E., Casey, J.F. and Gill, D.A. (2016). "Visitor preferences and willingness to pay for coastal attributes in Barbados". *Ocean and Coastal Management*, 134, 240-250.
Sever, I. and Verbic, M. (2018). "Providing information to respondents in complex choice studies: a survey on recreational trail preferences in an urban nature park". *Landscape and Urban Planning*, 169, 160-177.
Shahvali, M., Kerstetter, D.L. and Townsend, J.N. (2020). "The contribution of vacationing together to couple functioning". *Journal of Travel Research*, https://doi.org/10.1177/0047287519892340
Shen, J. (2009). "Latent class model or mixed logit model? A comparison by transport mode choice data". *Applied Economics*, 41(22), 2915–2924.
Shi, S., Wedel, M. and Pieters, F. (2013). “Information acquisition during online decision making: a model-based exploration using eye-tracking data”. *Management Science*, 59 (5), 1009–1026.
Shoji, Y. and Tsuge, T. (2015). "Heterogeneous preferences for winter nature-based tours in sub-frigid climate zones: a latent class approach". *Tourism Economics*, 21(2), 387-407.
Simon, H. (1955). *A Behavioral Model of Rational Choice*. Oxford University Press, pp. 99–118, 69(1).
Sirakaya, E. and Woodside, A. G. (2005). “Building and testing theories of decision making by travelers”. *Tourism Management*, 26, 815-832.
Small, K.A and Rosen, H.S. (1981). "Applied welfare economics with discrete choice models". *Econometrica*, 49(1), 105-130.
Swait, J. and Adamowicz, W. (2001). "The influence of task complexity on consumer choice: a latent class model of decision strategy switching". *Journal of Consumer Research*, 28, 135-148.
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.
Talpur, M.A., Koetse, M.J. and Brouwer, R. (2018). "Accounting for implicit and explicit payment vehicles in a discrete choice experiment". *Journal of Environmental Economics and Policy*, 7(4), 363-385.
Thai, N.T. and Yuksel, U. (2017). "Too many destinations to visit: tourists' dilemma?" *Annals of Tourism Research*, 62, 38-53.
Thiene, M., Scarpa, R. and Louviere, J.J. (2015). "Addressing preference heterogeneity, multiple scales and attribute attendance with a correlated finite mixing model of tap water choice". *Environmental and Resource Economics*, 62, 637-656.
Thiene, M., Swait, J. and Scarpa, R. (2017). "Choice set formation for outdoor destinations: the role of motivations and preference discrimination in site selection for the management of public expenditures on protected areas". *Journal of Environmental Economics and Management*, 81, 152-173.
Thiene, M., Franceschinis, C. and Scarpa, R. (2019). "Congestion management in protected areas: accounting for respondents' inattention and preference heterogeneity in stated choice data". *European Review of Agricultural Economics*, 46(5), 834-861.
Thornton, P., Shaw, G. and Williams, A.M. (1997). "Tourist group holiday decision-making and behavior: the influence of children". *Tourism Management*, 18(5). 287-297.
Thurstone, L.L. (1927). “A law of comparative judgment”. *Psychological Review*, 34, 273–286.
Train, K. (1998). "Recreation demand models with taste variation". *Land Economics*, 74, 230–239.

Train, K. (2008). "EM algorithms for nonparametric estimation of mixing distributions". *Journal of Choice Modelling*, 1(1), 40-69.
Train, K. (2009). *Discrete choice models with simulation*. Second Edition. Cambridge University Press.
Train, K. and Weeks, M. (2005). *Discrete choice models in preference space and willingness-to-pay space*. In: Alberini, A. and Scarpa, R. (eds), Applications of Simulation Methods in Environmental and Resource Economics. Springer, Dordrecht.
Tversky, A. (1972). “Elimination by aspects: A theory of choice”. *Psychological Review*, 79, 281–299.
Van Cranenburgh, S. (2018). "Short-term responses of Dutch vacationers to a sharp increase in transport costs". *Current Issues in Tourism*, 21(14), 1601-1620.
Van Cranenburgh, S. Chorus, C.G. and Van Wee, B. (2014). "Vacation behavior under high travel cost conditions - A stated preference of revealed preference approach". *Tourism Management*, 43, 105-118.
Van Cranenburgh, S. and Collins, A.T. (2019). "New software tools for creating stated choice experimental designs efficient for regret minimisation and utility maximisation decision rules". *Journal of Choice Modelling*, 31, 104-123.
Van Cranenburgh, S., Rose, J.M. and Chorus, C.G. (2018). "On the robustness of efficient experimental design towards the underlying decision rule". *Transportation Research Part A*, 109, 50-64.
Van Haefen, R.H., Massey, D.M., Adamowicz, W.I. (2005). “Serial nonparticipation in repeated discrete choice models”. *American Journal of Agricultural Economics*, 87(4), 1061–1076.
Varian, H.R. (2010). *Intermediate Microeconomics. A Modern Approach*. Chapter 3: Preferences. Chapter 4: Utility. Eight Edition. W.W. Norton & Company, New York.
Viglia, G. and Dolnicar, S. (2020). "A review of experiments in tourism and hospitality". *Annals of Tourism Research*, 80, 102858.
Vij, A. and Krueger, R. (2017). "Random taste heterogeneity in discrete choice models: flexible nonparametric finite mixture distributions". *Transportation Research Part B*, 106, 76-101.
Vossler, C.A., Doyon, M. and Rondeau, D. (2012). "Truth in consequentiality: theory and field evidence on discrete choice experiments". *American Economic Journal: Microeconomics*, 4, 145–171.
Walker, J.L. and Ben-Akiva, M. (2002). "Generalized random utility model". *Mathematical Social Sciences*, 43, 303-343.
Windle, J. and Rolfe, J. (2014). “Assessing the trade-offs of increased mining activity in the Surat Basin, Queensland: preferences of Brisbane residents using nonmarket valuation techniques”. *Australian Journal of Agriculture and Resource Economics*, 58 (1), 111–129.
Wong, I.A., Law, R. and Zhao, X. (2017). "When and where to travel? A longitudinal multilevel investigation on destination choice and demand". *Journal of Travel Research*, 56(7), 868-880.
Wuepper, D. (2017). "What is the value of world heritage status for a German national park? A choice experiment from Jasmund, I year after inscription". *Tourism Economics*, 23(5), 1114-1123.
Yang, X. and Carlsson, F. (2016). "Influence and choice shifts in households: an experimental investigation". *Journal of Economic Psychology*, 53, 54-66.
Zhang, W. and Sohngen, B. (2018). "Do U.S. Anglers care about harmful algal blooms? A Discrete Choice Experiment of Lake Erie Recreational Anglers". *American Journal of Agricultural Economics*, 100(3), 868-888.
Zhang, J., Timmermans, H., Borgers, A. and Wang, D. (2004). "Modelling traveler choice behavior using the concepts of relative utility and relative interest". *Transportation Research Part B*, 38, 215-234.

SUPPLEMENTARY MATERIAL

Understanding Marginal Rates of Substitution Among Holiday Destination Attributes: A Discrete Choice Experiment

ANNEX 1.- Instructions for individual DCE

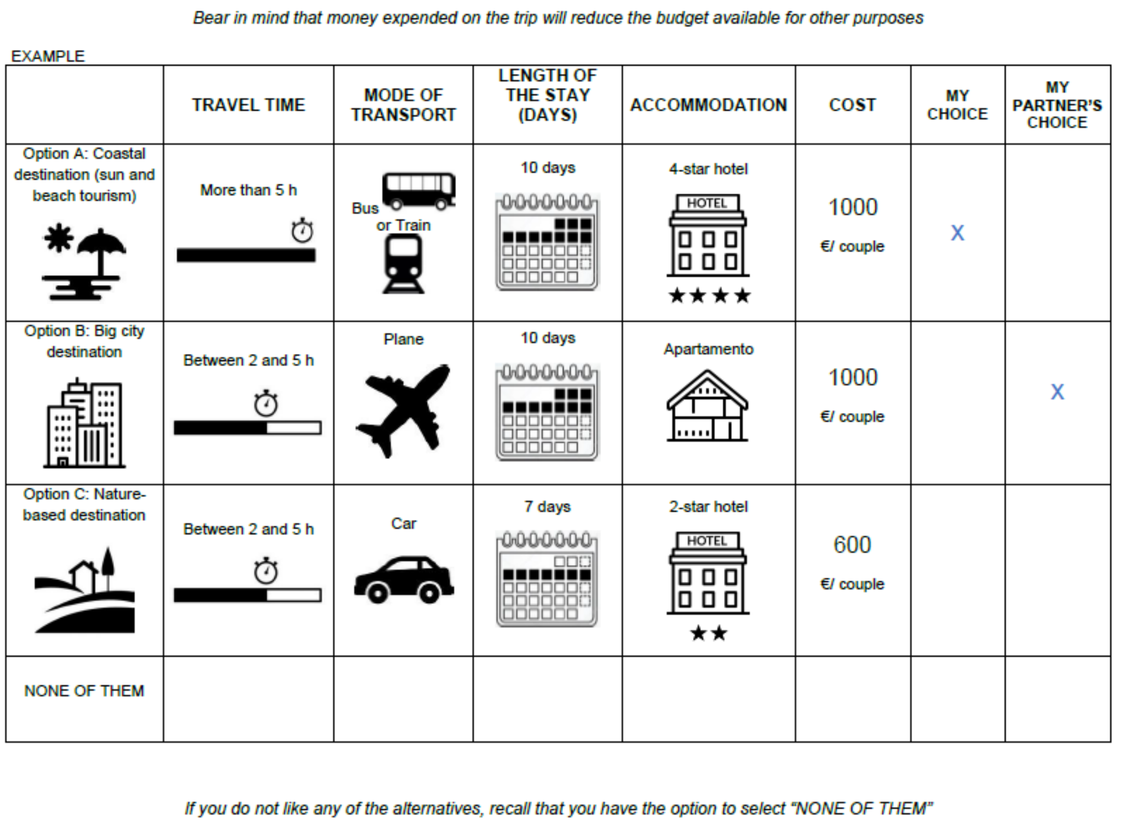
*Imagine that both you and your partner have at least **15 days of holidays** between June and September. Now we are going to show you different alternatives for a **holiday trip together** (only you and your partner, without children, relatives or friends) within your country.*

It is important to bear in mind the following:

- *<u>The cost includes the accommodation and transportation expenses (both).</u>*
- *The accommodation only includes breakfast*
- *The commuting time refers to the transit time between you depart from your living place and you arrive at the accommodation.*

We want you to choose which of the three options in each block you would choose for a trip with your partner supposing you have the full power to decide. In addition, we also ask you to indicate which you think would be your partner's choice.

EXAMPLE:



Bear in mind that money expended on the trip will reduce the budget available for other purposes

EXAMPLE

	TRAVEL TIME	MODE OF TRANSPORT	LENGTH OF THE STAY (DAYS)	ACCOMMODATION	COST	MY CHOICE	MY PARTNER'S CHOICE
Option A: Coastal destination (sun and beach tourism)	More than 5 h	Bus or Train	10 days	4-star hotel	1000 €/ couple	X	
Option B: Big city destination	Between 2 and 5 h	Plane	10 days	Apartamento	1000 €/ couple		X
Option C: Nature-based destination	Between 2 and 5 h	Car	7 days	2-star hotel	600 €/ couple		
NONE OF THEM							

If you do not like any of the alternatives, recall that you have the option to select "NONE OF THEM"

Figure A1.- Example shown to respondents

***<u>Important:</u>** If you do not specially like any of the three proposed alternatives, remember that you have the option to mark "none of them".*

ANNEX 2.- A D-efficient design

Given the chosen attributes and their corresponding levels, the total number of possible combinations of the level of the attributes (full factorial design) will be 324 (3x3x3x3x4). Asking respondents to trade-off all possible combinations would impose them high cognitive burden. Because of this, researchers normally select an appropriate subset from the full factorial design.

Traditionally, scholars used to generate designs that minimize the correlations between the attribute levels. This is referred as the principle of orthogonality. In linear models, orthogonality avoids multicollinearity and minimizes the variances of the parameter estimates. However, as shown by Rose and Bliemer (2009), orthogonality is an inappropriate design criterion for most cases in discrete choice modelling[145]. Furthermore, Hensher and Bernard (1990) show that discrete outcomes, if so, requires orthogonality in the differences between attribute levels, not in their absolute levels.

Due to the monetary and time costs a DCE involves, researchers have moved to experimental designs that either provide the maximum reliability of the parameter estimates for a given sample size, or reduce the sample size needed to obtain a certain level of reliability in the estimates. In other words, researchers have developed experimental designs aimed at providing the greatest statistical efficiency. This is achieved by minimizing the asymptotic standard errors of the parameter estimates, which are obtained by taking the square root of the elements in the asymptotic variance-covariance matrix (hereafter AVC)[146]. The smaller the elements of the AVC, the more statistically efficient the experimental design is (at least asymptotically)[147]. Experimental designs whose combination of attributes yields to the lowest AVC are thus called *efficient designs*[148].

To measure the efficiency of the design some kind of error is used. The known as D-error is by far the most used (Bliemer et al., 2008) and has been shown to provide better results than traditional orthogonal designs[149]. The D-error is given by the determinant of the AVC matrix scaled by the number of parameters (K)[150] so that:

$$Derror = det\ (\Omega\ (\tilde{\beta}|X))^{\frac{1}{K}} \qquad (3.20)$$

[145] The reason is that the AVC matrix of the non-linear models employed in the DCE literature (those from the logit family) is different from that of linear regression. Accordingly, orthogonality in such models does not produce statistical efficiency (Rose and Bliemer, 2009).
[146] The AVC equals the negative inverse of the Fisher information matrix (FIM), which is given by the expected values of the second derivatives of the log-likelihood function (see Train, 2003).
[147] Fixing N, the smaller the asymptotic standard errors, the smaller the width of the confidence intervals around the parameter estimates and the higher the asymptotic t-ratios.
[148] As shown by Yao et al. (2015), efficient designs induce choice behaviour to be more consistent with the assumption of fully compensatory choice that underlies the Random Utility Framework as opposed to other criteria like orthogonal designs (OD) or orthogonal in the difference designs (ODD).
[149] Monte Carlo simulations by Carlsson and Martinsson (2003) show that D-error designs produce unbiased parameter estimates with lower mean square errors than orthogonal designs. See also Rose et al. (2008).
[150] The determinant of the AVC matrix will rise with the number of elements in β, so it is necessary to scale it by 1/K.

where $\Omega\ (\tilde{\beta}|X)$ denotes the AVC of the design, X is a vector of attribute levels and $\tilde{\beta}$ refers to the prior parameter values.

Designs that minimize the D-error statistic are called *D-efficient* designs. The minimization of the D-error requires the researcher to specify some parameter priors, which usually constitutes a major challenge (Hoyos, 2010). A common approach is to use the parameter estimates from the pilot study. This was what we did. Rose and Bliemer (2009) compare the performance of several experimental designs. They provide evidence that orthogonal designs only outperform efficient designs when all the parameters are close to zero. Such a situation is unlikely to hold as the researcher normally has arguments for believing that the effect of an attribute on choice is different from zero[151].

Efficient design theory has mainly concentrated on the Multinomial Logit Model (MNL), which usually constitutes the baseline model in any empirical application. Accordingly, our experimental design was generated for an MNL model using *Ngene* software (Choice Metrics, 2012, version 1.2)[152]. Following good practice, we let the search algorithm to do at least a few iterations to minimize the D-error. The minimization of the D-error is normally done assuming one single respondent. Since the AVC matrices of DCE are asymptotically divisible by N (McFadden, 1974), once having derived the AVC matrix for a single respondent, the corresponding AVC matrix for a sample size N is given by dividing each element in the AVC matrix by N.

Each choice tasks was checked for the possible presence of dominant alternatives. In such case, that alternative will display a probability of one of being chosen and respondents would not make the intended trade-offs between the attributes (Hensher and Rose, 2007). We also examined whether any alternative was described by an implausible combination of attributes. Implausibility needs to be addressed from the respondent's perspective (Lancsar and Louviere, 2008). To do so, we relied on comments received from the pilot. Based on that, we imposed some restrictions to the attribute combinations in the final design.

Several remarks are in order. First, a design that is efficient for the MNL model might not be efficient for another model. Second, some researchers impose the attribute level balance constraint by which each attribute level is forced to occur an equal number of times. However, imposing this constraint would further reduce statistical efficiency as the combination of attribute levels that minimize the AVC matrix is not freely chosen. Third, our experimental design is efficient under the assumption that respondents are linear-additive utility maximisers. However, as shown by Van Craneburgh et al. (2018) and Van Cranenburgh and Collins (2019), conventional RUM efficient designs could be statistically inefficient in case the underlying decision rule is random regret minimization.

[151] For further details about statistical efficiency in experimental designs see Sándor and Wedel (2001) and Scarpa and Rose (2008).

[152] Although new alternatives have emerged such as the Robust Design Generator (RDG) in MATLAB (Van Cranenburgh and Collins, 2019), Ngene is by far the most flexible and customizable tool for experimental design generation in discrete choice modelling.

Based on the parameter estimates obtained from the pilot study, we report below the NGENE code used for generating the experimental design used in the analysis:

```
?EXPERIMENTAL DESIGN

?3 alternatives + none; 4 levels for Cost; 3 levels for Lodging;

?18 choice situations in 3 blocks + imposing restrictions

?Time: <2 h (0), between 2-5 h (1), >5 h (2)

?LOS: 2-3 days (0), between 5-7 days (1), 10 days (2)

?Accommodation: apartment (0), two-star hotel (1), four-star hotel (2)

?Transport: Car (2), Bus or train(1), Plane(2)

?Cost: 200, 600, 1000, 1400

Design; alts = alt1, alt2, alt3, alt4; rows = 18; block = 3; eff = (mnl,d);

cond:

if(alt1.Cost=1.4, alt1.Accom=[2]),

if(alt1.Cost=1.4, alt1.LOS=[2]),

if(alt2.Cost=1.4, alt2.Accom=[2]),

if(alt2.Cost=1.4, alt2.LOS=[2]),

if(alt3.Cost=1.4, alt3.Accom=[2]),

if(alt3.Cost=1.4, alt3.LOS=[2]),

if(alt1.Cost=0.2, alt1.Accom=[0]),

if(alt1.Cost=0.2, alt1.LOS=[0]),

if(alt2.Cost=0.2, alt2.Accom=[0]),

if(alt2.Cost=0.2, alt2.LOS=[0]),

if(alt3.Cost=0.2, alt3.Accom=[0]),

if(alt3.Cost=0.2, alt3.LOS=[0]);

model:

U(alt1) = ASC1[0.12] + b2.dummy[0.3|-0.1] * Time[1,2,0] + b3.dummy[0.2|-0.2]  *
Transport[1,2,0] +  b4.dummy[1.05|1.25] * los[1,2,0] + b5.dummy[-0.8|-0.2] *
Accom[1,2,0] + b6[-1.4] * Cost[0.2,0.6,1.0,1.4] /

U(alt2) = ASC2[0.09] + b2* Time + b3 * Transport + b4.dummy* LOS + b5.dummy*
Accom + b6 * Cost        /

U(alt3) =                  b2* Time + b3 * Transport + b4.dummy* LOS + b5.dummy*
Accom + b6 * Cost        /

U(alt4) = ASC4[-2.2]    $
```

ANNEX 3.- Individual questionnaire

Individual identifier: Couple identifier:

Now I would like you to answer the following questionnaire. Before starting, it is important to highlight that your answers will be anonymous. Neither your partner nor anyone could know the content of your answers (for us your answers will be associated with an identificatory number but not with your name). Hence, please feel free to answer. We encourage you to be sincere when answering.

If you do not want to answer certain questions, you can leave it in blank.

The questionnaire comprises four blocks:

A. SOCIODEMOGRAPHIC CHARACTERISTICS
B. RELATIONSHIP WITH YOUR PARTNER
C. TASTE FOR TRAVELLING
D. PERSONALITY

Please circle the correct answer in each case.

SOCIODEMOGRAPHIC CHARACTERISTICS

First of all, we start asking you some basic questions about your personal characteristics.

Q1. Please, what is your gender?:

a. Male b. Female

Q2. What is your age?: ________

Q3. What is your educational level?:

a. Primary education (*ESO*)
b. Secondary education (it includes *Bachillerato* and *vocational training of degree medium and higher*)
c. University education

Q4. What is your civil status?:

a. Married
b. Consensual union
c. Separated/Divorced/Widow(er)

Q5. What is your labor status?:

a. Self-employed
b. Employee
c. Housekeeper
d. Retired
e. Student
f. Unemployed

Q6. If you are working and earning a salary now, can you please indicate (approximately) the average number of hours you work per week?

a. Less than 15 hours
b. Between 15-30 hours
c. Between 30-40 hours
d. More than 40 hours

Q7a. Do you have children?

a. Yes b. No

Q7b. In case you have children, how many do you have?: __________

Q7c. In case you have children, how old are they? _______ ________ _______ ________

Q8. Could you please indicate the range of your monthly after tax INDIVIDUAL income?:

a. Zero euros (I do not earn money)
b. Less than 500 euros
c. Between 500 and 1,500 euros
d. Between 1,500 and 2,500 euros
e. More than 3,500 euros

Q9. Please choose the correct option:

a. I live in Oviedo/Gijon/Avilés
b. I live in another place

Q10. Nationality:

a. Spanish b. Other

Q11. How do you assess your health status?

a. Good b. Bad

Q12. Now I am going to enumerate you a list of three issues that might or might not be true for you. Please indicate just the NUMBER of them (0,1,2,3) that are correct in your case. Do not tell me which of them but how many:

A. I have travelled by car without the seat belt at least once
B. I have celebrated a meeting with my family or some friends in the last month
C. I have been about to break with my partner/spouse
D. I have been abducted by aliens

The number of statements that are true for me are: ________

RELATIONSHIP WITH THE PARTNER

Now we will ask you some questions about you and your partner, and about how your relationship it. Please, do not worry because we do not make you any sensitive question. This information is crucial for us in order to examine the factors that explain the way couples make joint decisions.

Remember that this information would remain strictly private and your personal data will not be related to your name. Your partner will NOT know the content of your answers. Hence, please feel free to answer.

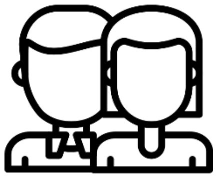
Q13. Please indicate the type of relationship you have with your partner in the experiment:

a. We are married
b. We are in a consensual union but unmarried

Q14. In case you are married, how long have you been married? ________________

Q15. Now please indicate how long you have been in this relationship (since you started)

a. Less than 5 years
b. Between 5-15 years
c. Between 15-25 years
d. More than 25 years



Q16a. Please indicate your birthdate:

____

Q16b. Now please indicate the birthdate of your partner:

____

Q17. In general terms, who do you consider has more power for making decisions within the couple?

a. Mainly me
b. Mainly my partner
c. We make most decisions jointly.

Q18. In general terms, how satisfied are you with the WAY decisions are taken in your couple?

To answer use a scale from 0 to 10 where 0 indicates that you are "NOT SATISFIED AT ALL" and 10 that you are "FULLY SATISFIED".

____

Q19. Please, indicate on a 0-10 scale where 0 means "NOTHING AT ALL" and 10 means "A GREAT DEAL" the degree you consider:

a. That you know your partner (his/her preferences) _______________
b. That you worry about your partner (about his/her preferences) _____________
c. That your partner will support and help you in case of necessity or unexpected bad situation (example: health problem) ___________

Q20. Taking everything into account, how satisfied are you with your live?

To answer this question, use a 0-10 scale where 0 means "NOT SATISFIED AT ALL" and 10 means "FULLY SATISFIED"

____

Q21. Please indicate your agreement/disagreement with the following statements on a 1-7 scale where 1 means "I STRONGLY DISAGREE" and 7 means "I STRONGLY AGREE":

a.	*It is inevitable that during a relationship some conflicts and arguments arise, independently of whether one tries to do things as better as possible*

1 2 3 4 5 6 7

b.	*When my partner and I got angry with each other for any reason, I have the ability to become reconciled easily*

1 2 3 4 5 6 7

c.	*When my partner and I got angry with each other for any reason, I am the person who moves first to solve the conflict*

1 2 3 4 5 6 7

d.	*When my partner and I got angry with each other for any reason, we simply make time pass until we forgot about the argument*

1 2 3 4 5 6 7

e.	*Success in a relationship merely depends on the effort each one puts into it. If both strive, the relationship works*

1 2 3 4 5 6 7

f.	*My partner's moods are a true mystery for me. In most cases, I do not know what happens to him/her*

1 2 3 4 5 6 7

g.	*Most of our arguments and misunderstanding arise due to stupid and circumstantial things*

1 2 3 4 5 6 7

h.	*I believe my partner and I would remain together and happy even under bad and extreme situations*

1 2 3 4 5 6 7

i.	*I consider that the success of a relationship depends on clear communication. Being sincere and talking with your partner about problems and throughs are crucial factors for a successful relationship*

1 2 3 4 5 6 7

TRAVELLING TASTE

Now we will make you some questions about a different topic: your travelling taste and the frequency with which you travel. This information is relevant to us because we aim to analyse your preferences for the different types of tourist destinations previously shown.

Q22. Do you like to go on holidays?

a. Yes b. No

Q23. If you can choose, what do you prefer?

a. Travelling domestically
b. Travelling abroad

Q24. If you can choose, what would you prefer?

a. Travelling alone
b. Travelling with my partner
c. Travelling with other people different from my partner (friends/relatives)
d. Depending on the type of trip and circumstances, sometimes I prefer to travel with my partner and in other cases I prefer to travel with other people

Q25. Have you been on holidays with your partner during the last year (at least one night away from home?

a. Yes b. No

Q26. Who is the person that normally decides where to go on holidays?

a. Mainly me
b. Mainly my partner
c. We decide it consensually

Q27. How often do you go on holidays (travel for leisure purposes)?

a. I hardly ever go on holidays
b. Once every two years
c. Once a year
d. Twice a year
e. Three times a year
f. More than three times a year

Q28. When you go on holidays with your partner, about how much money do you (the two) usually spend (in euros)?

a. Less than 250 euros
b. Between 250-500 euros
c. Between 500-800 euros
d. Between 800-1,200 euros
e. More than 1,200 euros

Q29. Please indicate which of the following statements suits you the best when going on vacation:

a. My partner and I each contribute about 50% of the total expenses
b. I pay for most of the expenses
c. My partner pays for most of the expenses

Q30. When choosing where to go, which type of the following alternatives you prefer the most?

a. Sun and beach tourism
b. Cultural tourism, big cities
c. Nature-based tourism
d. I do not have a clear preference for any of them, I like all of them

PERSONALITY

Finally, we make you some questions about your personality. This information is important for us because the way people make decisions is partially determined by the way of being of each.

Q31. In general terms, on a 0-10 scale where 0 means "NOTHING" and 10 means "VERY MUCH", how patience you consider yourself?

____

Q32. In general terms, how willing are you to take risks in general? To answer this question, use a 0-10 scale where 0 means "NOTHING AT ALL" and 10 means "A GREAT DEAL"

____

Q33. Now we present you a number of personality traits that may or may not apply to you. Please indicate, for each statement, the extent to which you agree of disagree that they apply to you.

To answer, use a scale from 1 to 7 where 1 means "DISAGREE STRONGLY" and 7 means "AGREE STRONGLY"

A. *Extraverted, enthusiastic*

1 2 3 4 5 6 7

B. *Critical, quarrelsome*

1 2 3 4 5 6 7

C. *Dependable, self-disciplined*

1 2 3 4 5 6 7

D. *Anxious, easily upset*

1 2 3 4 5 6 7

E. *Open to new experiences, complex*

1 2 3 4 5 6 7

F. *Reserved, quiet*

1 2 3 4 5 6 7

G. *Sympathetic, warm*

1 2 3 4 5 6 7

H. *Disorganized, careless*

1 2 3 4 5 6 7

I. *Calm, emotionally stable*

1 2 3 4 5 6 7

J. *Conventional, uncreative*

1 2 3 4 5 6 7

Q34. Have you participated before in any study like this?

a. Yes b. No

End of the individual questionnaire

Thank you very much for your collaboration

ANNEX 4.- Travel frequency and scale heterogeneity

To explore whether travel frequency (*travfreq)* makes some individuals to be more deterministic in their choices than others, we estimate the following Scaled Multinomial Logit Model (SMNL):

$$U_{ijt|c} = ASC_j + \beta_c X_{ijt} + \frac{\varepsilon_{ijt}}{\sigma_i} = \sigma_i (ASC_j + \beta_c X_{ijt}) + \varepsilon_{ijt} \qquad (3.21)$$

Instead of fixing the scale to one, this model allows the scale (the inverse of the variance) of the Gumbel error term to be heterogeneous in the population so that:

$$Var(\varepsilon_{ijt}) = \frac{\pi^2}{6\sigma^2{}_i}$$

$$\sigma_i = \exp(\bar{\sigma} + \tau\epsilon_i) \quad \text{where } \epsilon_i \sim N(0,1) \qquad (3.22)$$

The scale parameter is assumed to be distributed lognormal. For identification purposes, the mean of σ_i is set to one so that $\bar{\sigma} = -\frac{\tau^2}{2}$. If we introduce *travfreq* as a shifter in the scale equation, then it becomes:

$$\sigma_i = \exp(-\frac{\tau^2}{2} + \delta\, travfreq_i + \tau\epsilon_i) \qquad (3.23)$$

Descriptive statistics for *travfreq* are presented in Table A1[153]. The parameter estimates for the SMNL model based on 1,000 Halton draws are shown in Table A2. The τ parameter is about the unity whereas *travfreq* is not statistically significant. This implies that there no evidence of scale heterogeneity in our data related to travel experience.

travfreq	Description	%
0	I hardly ever go on holidays	7.6
1	Once every two years	3.8
2	Once a year	29.4
3	Twice a year	35.9
4	Three times a year	11.4
5	More than three times a year	11.8

Table A1.- Descriptive statistics for *travfreq*

[153] The wording of this question is: How often do you go on holidays/travel for leisure purposes?

	SMNL		
Variables	Coef.		SE
ASC1	1.955	***	0.416
ASC2	1.640	***	0.361
ASC3	1.262	***	0.319
medTT	0.180	*	0.100
longTT	0.113		0.087
bustrain	-0.089		0.091
plane	0.247	**	0.088
7days	1.532	***	0.317
10days	1.623	***	0.328
2-starhotel	-0.195	**	0.090
4-starhotel	0.177	*	0.094
Cost	-0.195	***	0.039
τ	1.030	***	0.131
travfreq	0.069		0.061

Table A2.- Parameter estimates for SMNL model
*** p<0.01, ** p<0.05, * p<0.1

ANNEX 5.- Conditional Willingness-to-Pay estimates

One of the attractive features of imposing parameter structure to preference heterogeneity, generally in the form of a continuous distribution (as in RPL), is the possibility of deriving conditional (posterior) individual-specific estimates of preferences for the attributes. This consists on moving from the *unconditional* to the *conditional* distribution of preferences. After the seminar paper by Revelt and Train (2000), several studies have derived conditional parameters at the individual level after RPL once conditioning on all available information about the individual. Some examples include Hess and Hensher (2010) and Olsen et al. (2020). Recently, Sarrias and Daziano (2018) have extended this procedure to the LCMNL model.

Let the population distribution of preferences for the case of three latent classes to be given by:

$$g(\beta_i|\lambda) = \begin{cases} \beta_1 \text{ with probability } \pi_{i1} \\ \beta_2 \text{ with probability } \pi_{i2} \\ \beta_3 \text{ with probability } \pi_{i3} \end{cases} \tag{3.24}$$

being λ the parameters that characterize class membership. The unconditional probability of individual i sequence of choices is:

$$f(y_i|X_i,\theta) = \sum_{c=1}^{3} f(y_i|X_i,\beta_c)\, g(\beta_i|\lambda) \tag{3.25}$$

where $\theta = (\lambda, \beta)$ and X_i reflects the attribute level faced by individual i in his sequence of choices. Based on Bayes' theorem, the posterior distribution of the individual-specific marginal utilities given θ is expressed as:

$$f(\beta_i|y_i,X_i,\theta) = \frac{f(y_i|X_i,\beta_c)g(\beta_i|\lambda)}{f(y_i|X_i,\theta)} \tag{3.26}$$

This posterior conditional distribution of individual parameters $f(\beta_i|y_i,X_i,\theta)$ differs from the unconditional one $g(\beta_i|\lambda)$ in that it conditions on the sequence of choices y_i made when facing a design matrix of attributes X_i and on the structural parameters θ that characterize the distribution of preferences in the population (Train, 2009, pp. 259-281).

Following Sarrias and Daziano (2017, 2018), the conditional expectation of β_i (point estimates) for the LCMNL model is given by:

$$\hat{\beta}_i = E[\beta_i|y_i,X_i,\theta] = \frac{\sum_{c=1}^{3} \beta_c f(y_i|X_i,\beta_c)g(\beta_i|\lambda)}{\sum_{c=1}^{3} f(y_i|X_i,\beta_c)g(\beta_i|\lambda)} \tag{3.27}$$

As shown by Greene (2012, pp. 682-685), an estimator of the conditional variance from the point estimates in (3.21) is obtained as[154]:

[154] This way of deriving the conditional variance has the drawback that it does not consider the sampling variability around $\hat{\theta}$. As discussed in Greene et al. (2014), the estimator in (3.28) is an estimator of the

$$\widehat{Var(\widehat{\beta_i})} = E\left[{\beta_i}^2 \middle| y_i, X_i, \theta\right] - E[\beta_i | y_i, X_i, \theta]^2 \quad (3.28)$$

Following Sarrias and Daziano (2018), we derived the conditional individual specific conditional WTP estimates, since the conditional expectation can be obtained from any statistic $k(\beta_i)$[155]. Although the derivation of the posterior individual-specific marginal utilities is usually performed under a Bayesian framework, Huber and Train (2001) show that both the classical (as we do) and the Bayesian procedure lead to equivalent results.

variance of the conditional distribution of $\hat{\beta}_i | data_i$, but not an estimator of the sampling variance of the estimator of E($\hat{\beta}_i | data_i$)

[155] An earlier application of this for the LCMNL model is Scarpa and Thiene (2005).

CONCLUSIONES

La tesis estudia las preferencias de los consumidores por las actividades turísticas. Está compuesta por tres estudios empíricos que analizan las decisiones turísticas a nivel individual utilizando diferentes metodologías. En este sentido, considero que la modelización econométrica de las preferencias es el principal valor añadido de la tesis.

En el capítulo 1 se analiza la relación entre un conjunto de características sociodemográficas, de oferta, y factores relativos al viaje en la duración de la estancia de los turistas que visitan Asturias. Se hace uso de una base de datos novedosa y no explotada anteriormente que contiene información detallada de más de 19.000 individuos encuestados a lo largo de todo el año, desde 2010 hasta 2016. Al contrario que estudios previos, se modeliza la decisión de pernoctar en el destino. Así, se identifican los factores que distinguen a los turistas de los excursionistas. En línea con los resultados de la literatura, se controla por diferentes características individuales, prestando especial atención a la relación que existe entre los atractivos del destino, como el clima, su entorno natural o su tranquilidad, y la duración de la estancia. Además, se estudia cómo el conocimiento acerca del destino, en términos de experiencia previa o haber visto algún tipo de promoción, está asociado con estancias más prolongadas. Las estimaciones muestran que la búsqueda de tranquilidad, el clima de Asturias y su entorno natural están positivamente relacionadas tanto con la probabilidad de ser un turista como con la duración de la estancia. Una experiencia previa positiva y la publicidad también destacan como dos factores que aumentan la estancia. Además, en comparación con los excursionistas, los turistas parecen estar guiados por la búsqueda de novedad, la gastronomía y la recomendación de amigos y familiares. Los resultados tienen implicaciones importantes en términos de mejorar el atractivo de Asturias como destino turístico.

Además de esto, el capítulo 1 contribuye a la literatura empírica desde una perspectiva econométrica. Se propone un modelo de conteo tipo valla con truncamiento en cero que modeliza conjuntamente i) la decisión binaria de pernoctar, y ii) la duración de la estancia de aquellos que deciden pernoctar. En el capítulo se discute la validez de este método en relación con otros modelos usados en la literatura, como mínimos cuadrados ordinarios o los modelos de duración. El modelo tipo valla propuesta es mejor que los modelos con inflación de ceros porque se considera que aquellos que no pernoctan son ‘ceros verdaderos’ dada la naturaleza infrecuente de lo que se analiza (estancia en destino turístico). El modelo en dos etapas propuesto tiene también la ventaja de ser robusto a selección endógena. Asumiendo un proceso de Poisson para los recuentos positivos, se han considerado dos distribuciones para la heterogeneidad no observada, dando lugar al modelo Poisson-gamma (Binomial Negativo) y al Poisson-log normal. Los modelos son comparados usando un *test* propuesto por Santos-Silva, Tenreyro y Windmeier en 2015. El modelo Binomial Negativo ajusta los datos de mejor modo. Es importante señalar que, entre las diferentes variantes de este modelo, se estima un modelo Binomial Negativo tipo P que, en lugar de imponer una especificación de la varianza condicionada que sea lineal o cuadrática, estima el parámetro P que caracteriza su forma funcional. El parámetro P estimado dista bastante de los valores 1 y 2 comúnmente asumidos. De este modo, el estudio pone de manifiesto la necesidad de emplear especificaciones más flexibles en los estudios aplicados de variables de recuento.

En el capítulo 2 se trata de responder a una pregunta de investigación relativamente antigua acerca de qué factores atraen a los turistas a moverse entre regiones por motivos recreacionales, esto es, los factores que determinan la elección de destino turístico. Aunque existe amplia literatura que estudia esto usando flujos agregados, existen menos estudios que ofrezcan una caracterización microeconómica que combine datos individuales con características de las regiones. Para el caso de los viajes por motivo naturaleza o deporte, se estima un modelo Logit de parámetros aleatorios correlacionados con componentes de error que considera heterogeneidad no observada en la preferencia por los atributos y en el gusto por las regiones. Se han especificado variables socioeconómicas, factores temporales y características del viaje como moderadoras de las utilidades marginales de los atributos. Los resultados muestran que los turistas obtienen, en media, utilidad positiva de viajar a destinos más cálidos y a regiones con un gran número de áreas naturales protegidas, puntos de interés turístico, parques naturales y nacionales, y kilómetros disponibles para esquiar. Por el contrario, los turistas se ven desincentivados por la distancia, los precios y elevadas precipitaciones. Sin embargo, la preferencia media por regiones más cálidas y por destinos cercanos enmascara una importante heterogeneidad. Las regiones relativamente más cálidas son menos preferidas en el primer y cuarto trimestre y por motivos relacionados con la visita de áreas naturales y el montañismo. El efecto negativo de la distancia se ve moderado por la edad, la renta, y por motivos de viajes como la práctica de actividades acuáticas o deportes de aventura. Por el contrario, aquellos que viajan en grupos grandes, en fines de semana o en el cuarto trimestre del año se ven fuertemente desincentivados a viajar a regiones lejanas.

Posiblemente, el resultado más relevante del capítulo 2 sea el cálculo de las relaciones marginales de sustitución entre distancia y ganancias de temperatura. Para ello, a partir de los parámetros estimados por el modelo, se calculan en primer lugar las estimaciones condicionales de las utilidades marginales por estos dos atributos. Dada la naturaleza de sección cruzada de nuestros datos, este estimador explota la distribución de preferencias en la muestra condicionando por toda la información disponible acerca de cada individuo, una vez integrados los efectos aleatorios para los grupos de regiones en la forma de componentes del error. Para evitar problemas de singularidad de estas utilidades marginales condicionales en el vecindario de cero, se adopta el procedimiento propuesto por Hess y Hensher. Las estimaciones indican que los individuos están dispuestos, en media, a viajar 159 kilómetros para obtener una ganancia marginal de temperatura en relación con su origen. Es relevante destacar que aproximadamente un 70% de la muestra sustituye distancia por temperaturas más cálidas. Sin embargo, el restante 30% prefiere viajar a regiones más frías que su origen. Además, se calculan las elasticidades propias y cruzadas para el índice de precios y para las temperaturas relativas. Basándome en las elasticidades precio, se documenta un patrón de sustitución Norte-Sur. En lo que respecta a las elasticidades cruzadas para el cociente de temperaturas, incrementos en temperatura en las regiones del sur reducen las probabilidades de elección de las regiones del norte.

Finalmente, en el capítulo 3 se adopta una metodología diferente. A diferencia del análisis de preferencias reveladas a partir de datos de encuestas de los capítulos 1 y 2, aquí se realiza un experimento de elección discreta con el fin de examinar las preferencias por un viaje vacacional. Esto ofrece la ventaja de permitir al investigador controlar el contexto, en entorno, el marco y el conjunto de alternativas entre las que se elige. Se estiman las utilidades marginales de ciertas características del viaje, de tal manera que estos efectos son netos de factores que habitualmente pueden entremezclarse en este tipo de análisis. Al contrario que otros estudios relacionados que utilizan estudiantes universitarios o panelistas procedentes de servicios de reclutamiento profesionales, mi muestra es reclutada de la población general. De esta manera, aunque la muestra puede no ser perfectamente representativa de la población, considero que los resultados son, al menos, más generales que estudios previos. Además, la muestra está compuesta por parejas reales. En mi opinión, pedirles a los dos miembros de la pareja que, de manera individual, indiquen sus preferencias para un viaje conjunto incrementa el grado de involucración en la tarea planteada.

Se estima un modelo de clases latentes que permite identificar grupos con distintas preferencias al mismo tiempo que considera la estructura de panel de los datos. Se identifican tres grupos con diferentes sensibilidades a los atributos vacacionales en función de sus características sociodemográficas. Se calculan las relaciones marginales de sustitución en términos monetarios (disponibilidad a pagar), tanto como una media ponderada de las clases como usando un estimador condicional que explota la secuencia de elecciones. Las estimaciones de las dos aproximaciones son similares, aunque la distribución de éstas en el segundo procedimiento es un poco más ancha. En base a esto, se encuentra evidencia de que los individuos están dispuestos a pagar 170 euros por viajar en avión frente a hacerlo en coche, 120 euros por alojarse en un hotel de 4 estrellas en relación con hacerlo en un apartamento, y 760 euros por un viaje de 10 días en relación a uno de 3 días.

A partir de la demanda probabilística estimada, se lleva a cabo un ejercicio de simulación para calcular la pérdida de bienestar que se produciría ante el hipotético establecimiento de una tasa turística diaria por persona en cada una de las alternativas. Los resultados muestran que los individuos que tienen una mayor preferencia por el destino costero requerirían una mayor compensación económica para ser devueltos a sus niveles de utilidad originales.

En general, se destaca la relevancia de las condiciones climáticas y la distancia en las decisiones turísticas. En el capítulo 2, se muestra que la *desutilidad* de viajar a destinos lejanos es moderada por los propósitos del viaje en términos de las actividades que se desean realizar en el destino. En el capítulo 1, se encuentra evidencia de que los turistas que proceden de regiones lejanas tienen una mayor probabilidad de pernoctar y permanecen en el destino durante más días. En lo relativo al clima, en el capítulo 2 se documenta que la preferencia general por destinos más cálidos que el origen viene moderada por propósitos del viaje que requieran actividades de interior al aire libre, como puede ser el montañismo o la visita a áreas rurales y naturales. Esto es consistente con los resultados del capítulo 1, que muestra que los turistas que aprecian el clima moderado y húmedo de Asturias tienden a pernoctar más días.

En el capítulo 2 se calcula la relación de sustitución entre distancia y temperatura usando datos de preferencias reveladas. En el capítulo 3 se realiza un análisis similar a partir de las elecciones vacacionales declaradas en el experimento. En suma, los resultados subrayan la relevancia de la heterogeneidad en preferencias a la hora de priorizar unos atributos sobre otros. Estudios futuros sobre las preferencias individuales por los atributos vacacionales podrían extender este análisis considerando funciones de utilidad más genéricas.

La tesis modeliza dos de las decisiones turísticas más relevantes: la elección de destino y el tiempo de estancia en el destino. En el capítulo 1, la elección de destino se considera como dada, por lo que se estima una función de demanda de tiempo condicional. En el capítulo 2 se modeliza la elección de destino condicional en haber decidido viajar, sin prestar atención a la duración de la estancia en el destino seleccionado. En el capítulo 3, la duración de la estancia se considera como un atributo del destino que determina la elección de un paquete vacacional. Futuros estudios podrían combinar estas dos dimensiones utilizando sistemas de demanda discreta y continua en un marco econométrico unificado.

Un aspecto relevante que la tesis no analiza es la participación en las actividades turísticas. Como futura extensión, considero relevante modelizar la decisión individual de viajar, y si hacerlo domésticamente o al extranjero. En línea con la discusión acerca de si el turismo es un bien normal o un bien de lujo, parece necesario un análisis en profundidad sobre el papel de la renta, las restricciones de tiempo y la composición del hogar en la decisión de participación turística. El uso de datos longitudinales puede ayudar a un adecuado estudio del papel que juegan las experiencias recientes y la formación de gusto en la decisión de viajar.

Otra cuestión de interés para la caracterización microeconómica de la demanda turística es las preferencias del hogar. A la hora de viajar en parejas, los individuos generalmente tienen que encontrar un equilibrio entre las preferencias de cada uno de los miembros. La elección conjunta puede entenderse como un proceso por el cual las preferencias individuales se consideran como in *input* que se transforma en la decisión final a través de una negociación. Esto constituye uno de los aspectos más atractivos y aún sin resolver en la economía del hogar. Este tema forma parte de mi agenda de investigación futura.

CONCLUDING REMARKS

The thesis examines consumer preferences for tourism activities. It comprises three empirical studies that analyse individual decisions regarding tourism using different approaches and methodologies. I view the econometric modelling of tourism preferences as the main take-away from the thesis.

Chapter 1 studies the relationship between a large set of sociodemographic, supply-based and trip-related factors on the length of stay of tourists visiting Asturias. I use a novel and unexploited dataset containing detailed microdata for more than 19,000 individuals surveyed throughout the whole year from 2010 to 2016. Contrary to previous studies, I model the decision to stay overnight at the destination, thereby identifying the factors that distinguish tourists from same-day visitors. Following findings from the previous literature, I control for several individual characteristics, devoting attention to the linkages between destination pull factors, such as climate, natural environment or tranquillity, and tourists' length of stay. Additionally, I examine how knowledge about the destination in the form of previous experience or having seen any kind of advertising is associated with longer stays. The estimates show that the search for tranquillity, Asturias' mild climate and its natural environment are positively related with both the likelihood of becoming a tourist and with the length of the stay. Positive past experience and advertising also emerge as two factors that lengthen the stay. Additionally, as opposed to same-day visitors, tourists appear to be driven by the search for novelty, gastronomy and recommendation by friends and relatives. The findings have important implications for the purposes of enhancing the attractiveness of Asturias as a tourist destination.

Apart from this, Chapter 1 contributes to the empirical literature from an econometric point of view. I estimate a Zero-Truncated Hurdle count data model that jointly addresses i) the binary decision to stay overnight, and ii) the number of days to stay for those with positive stays. I discuss the adequacy of this method relative to alternative approaches used in the literature like OLS or duration models. The proposed Hurdle model is better suited than the Zero-Inflated counterpart because I regard those with zero stays as 'genuine zeros' due to the infrequent nature of the outcome being analysed. The two-part model has also the advantage of being robust to endogenous selection. Assuming a Poisson process for the positive counts, I consider two distributions for the unobserved heterogeneity, leading to the Poisson-gamma (Negative Binomial) and the Poisson-log Normal models. The performance of the two models in our data is assessed using a test proposed by Santos-Silva, Tenreyro and Windmeier in 2015. The Negative Binomial fits the data best. Remarkably, among the different variants of this model, I estimate the NBP model that, instead of imposing a linear or quadratic specification for the conditional variance, estimates the parameter P that characterizes its functional form. The estimated parameter P largely departs from the commonly assumed values of 1 and 2. In this way, the study points to the necessity of moving to more flexible model specifications in applied studies for counts.

In Chapter 2, I address a relatively old question about the factors that pull tourists to temporarily move across space for recreation purposes (i.e. the drivers of tourists' destination choice). Although there is extant empirical literature concerned with this using aggregate flows, there is little research that provides a microeconomic characterization combining individual microdata with place-based attributes. For the case of nature-based trips, I estimate a correlated Random Parameter Logit with Error Components model that controls for unobserved preference heterogeneity for the attributes and for the regions. Socioeconomic, temporal and trip-related features are considered as taste shifters. The results show that tourists derive, on average, positive utility from travelling to warmer regions and to areas with protected natural spaces, sightseeing spots, natural and national parks and kilometres available for skiing. Conversely, tourists are deterred by distance, prices and high rainfall. However, the mean preference for warmer regions and the distaste for distance masks relevant heterogeneity. Warmer locations are less preferred in the first and fourth quarters and for purposes of trekking and visiting natural areas. Distaste for distance is moderated by age, income and by the purpose of practising aquatic or adventure sports. Conversely, larger travel party sizes, weekend trips and trips in the fourth quarter are associated with a larger disutility for distant locations.

Possibly, the most relevant finding from Chapter 2 is the computation of the marginal rate of substitution between distance and gains in temperature. To this end, from the model estimates, I compute the conditional estimates of the marginal utilities for these two attributes. Given the cross-sectional nature of the dataset, this estimator exploits the distribution of preferences in the sample by conditioning on all available information about each individual, after having integrated out the random effects for the macro-regions in the form of error components. To avoid singularities for conditional marginal utilities in the neighbourhood of zero, I adopt Hess and Hensher's approach. Individuals are, on average, willing to travel 159 kilometres to get a marginal gain in relative temperatures. Worthy of note, about 70% of the retained sample substitute distance for warmer locations. However, a non-negligible 30% of the sample travels to cooler areas. Additionally, I compute the own- and cross-elasticities for price indexes and relative temperatures. Based on the cross-price elasticities, I document a Northwest-South substitution pattern. Regarding the cross elasticities for the ratio of temperatures, rises in temperature in the South areas reduce choice probabilities of Northern regions.

Finally, a different approach is adopted in Chapter 3. As opposed to the analysis of revealed preferences from survey data, I conduct a Discrete Choice Experiment for the purpose of assessing preferences for a holiday trip. This offers the advantage of allowing the researcher to control for the context, the environment, the framing and the choice set from which decisions are made. As a result, I estimate the marginal utilities of certain trip features in a way that is net of some of the usual confoundings encountered in applied analysis. Contrary to many related studies conducted with university students or panellists from professional recruitment services, my subject pool is recruited from the general population. In this way, although I acknowledge that the sample is possibly not perfectly representative of the population, I consider my results to be, at least, more general than previous studies. Furthermore, my sample comprises real-life couples. In my view, asking the two members of the couple to individually make a choice for a joint trip enhances the degree of involvement in the choice task.

I estimate a Latent Class Model that allows for discrete classes of preference groups while considering the panel structure of the data. I identify three groups with different sensitivities to the vacation attributes based on sociodemographic characteristics. I compute the marginal rates of substitution in monetary terms (willingness to pay estimates), both as a weighted average across classes and as a conditional estimator that exploits the sequence of choices made by respondents. The estimates under the two approaches are similar, although the ones for the latter are wider. I find that respondents are willing to pay about €170 for plane travelling with respect to the use of car, €120 for staying at 4-star hotel relative to an apartment, and €760 for a 10-day trip relative to a 3-day one.

Based on the estimated probabilistic demand, I perform a simulation exercise to assess the welfare loss produce by the hypothetical setting of a daily tourism tax per person in each of the alternatives. The results show that individuals with a greater preference for coastal destinations would require more money to give them back to their original utility levels.

Overall, the thesis highlights the relevance of climate conditions and distance in tourism decisions. In Chapter 2, I show that the disutility of travelling to distant regions is moderated by tourist's goals in the form of the activities to perform at the destination. In Chapter 1, I find that those coming from farther away regions are more likely to stay overnight and stay for longer. Concerning climate conditions, Chapter 2 documents that the general preference for warmer locations is moderated by trip purposes that involve inland outdoor activities, such as trekking or visiting natural and rural areas. This is consistent with the findings of Chapter 1, in which I show that those who appreciate Asturias' wet and mild climate tend to stay for longer.

In Chapter 2, I compute substitution rates between distance and temperature using revealed preference data. In Chapter 3, a similar exercise is done based on vacation choices elicited in the experiment. Altogether, my findings underlie the relevance of heterogeneity in preferences at the time of trading-off the attributes of the different tourism goods. Future research on individual preferences for vacation attributes might extend this analysis by considering more generic utility functions.

The thesis has modelled two of the most relevant decisions regarding tourism: destination choice and time spent on vacation (i.e. length of stay). In Chapter 1, the choice of destination is taken as given so that we estimate a conditional demand function for time. In Chapter 2, I model the choice of destination conditional on travelling, without paying attention to the length of the stay at the chosen destination. In Chapter 3, the length of the stay is considered as a trip feature that determines the choice of a holiday package. Future research might combine both dimensions using discrete-continuous demand systems in a unified econometric framework.

A relevant issue the thesis does not address is the participation in tourism activities. As a future extension, I consider the modelling of individuals' decision to travel, and whether to travel domestically or abroad to be two aspects of great interest. In line with the discussion about whether tourism is a normal or a luxury good, an in-depth examination of the role of income, time constraints and household composition on tourism participation seems necessary. Longitudinal data could allow for a proper examination of the role of recent travel experiences and habit formation in the decision to take a vacation trip.

Another relevant issue for the microeconomic characterization of tourism demand is the analysis of households' joint preferences. When travelling in couples, individuals normally must find a balance between the preferences of each member. The joint choice of a vacation destination can be understood as a process by which individual preferences are taken as an input and transformed to a decision outcome through bargaining. This constitutes one of the most appealing and unresolved matters in household economics. This topic is part of my research agenda.