Universitat
deGirona

TOURISM DEMAND IN SPAIN: TRIP DURATION AND BUDGET STRUCTURE, A COMPARISON OF LOW COST AND LEGACY AIRLINES USERS

Berta Ferrer Rosell

Dipòsit legal: Gi. 1112-2014
http://hdl.handle.net/10803/145438

DOCTORAL THESIS

Tourism demand in Spain:
Trip duration and budget structure
a comparison of low cost and legacy
airline users

Berta Ferrer Rosell

2014

Universitat de Girona

University of Girona

DOCTORAL THESIS

Tourism demand in Spain:

Trip duration and budget structure – a comparison

of low cost and legacy airline users

Berta Ferrer Rosell

2014

University of Girona

DOCTORAL THESIS

Tourism demand in Spain:

Trip duration and budget structure – a comparison

of low cost and legacy airline users

Compendium of publications

Berta Ferrer Rosell

2014

JOINT DOCTORAL PROGRAMME IN TOURISM

Supervised by:

Esther Martínez Garcia

Germà Coenders Gallart

Department of Economics – University of Girona

Doctor of the University of Girona

Universitat de Girona

Dr. ESTHER MARTÍNEZ GARCIA and Dr. GERMÀ COENDERS GALLART, of the Universitat de Girona,

WE DECLARE:

That the thesis TOURISM DEMAND IN SPAIN: TRIP DURATION AND BUDGET STRUCTURE – A COMPARISON OF LOW COST AND LEGACY AIRLINE USERS, presented by Berta Ferrer Rosell to obtain a doctoral degree, has been completed under our supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, we hereby sign this document.

Signatures

Dr. Esther Martínez Garcia   Dr. Germà Coenders Gallart

Berta Ferrer Rosell

Al meu tiet Ferran Ferrer Julià (Jimmy).

Abans i sempre, present en els meus passos com a investigadora.

ACKNOWLEDGMENTS – Agraïments

El procés de realització de la tesi porta a un creixement personal que sovint és molt més important que el propi treball en la tesi. El què comporta escriure una tesi, no queda plasmat, ni de bon tros, en un document d'unes quantes pàgines, o tres articles, queda impregnat en la manera de fer, de ser i de veure el món de la persona que l'escriu. Te n'adones que avances, que creixes, que aprens, quan mires enrere, quan reflexiones sobre tot el què has fet i escrit, totes les dades que has analitzat, totes les hores, dies, setmanes i mesos que hi has dedicat i has deixat de dedicar a aquelles persones que, incondicionalment hi són sempre. Aquestes persones que t'acompanyen en tot aquest llarg procés, persones de dins i de fora de l'àmbit acadèmic, que entenen, o no, el què estàs fent, però que al veure't gaudir, o en ocasions patir, la única cosa que poden fer, és donar-te forces per continuar. Vull agrair molt sincerament, doncs, a totes les persones que, d'alguna manera o altra, han format part del meu equip i han contribuït a la culminació d'aquest projecte.

En primer lloc, vull agrair a la persona que m'ha vist néixer i fer els primers passos com a investigadora, que m'ha introduït en el món de la recerca i la docència universitària i que m'ha permès descobrir les coses bones, i no tant bones d'aquest món. GRÀCIES **Esther**, et dec molt! A l'acabar el primer curs del Graduat Superior en Gestió Turística, em vas oferir feina per l'estiu. Es tractava d'anar a fer enquestes a l'aeroport de Girona – Costa Brava per realitzar un informe sobre el perfil dels turistes que hi arribaven. Et vaig dir que no, tenia altres plans per aquell estiu, però a la primavera següent vas insistir, es feia un nou estudi. Llavors sí que vaig participar-hi, vaig fer les enquestes i em vas deixar escriure part de l'informe. Aquest fet, que ens va permetre disposar de la base de dades per realitzar la màster-tesi, així com també el nostre primer article, va ser la llavor del treball que des de llavors hem fet conjuntament, va ser el primer pas cap a un tàndem, i més tard un equip, sòlid. Encara em falta moltíssim per aprendre i per fer, però tenim equip per temps i segur que no ens avorrirem!

Qui havia de dir-nos, fa poc més de 10 anys, quan en Germà va ser professor de l'assignatura Mètodes quantitatius al Graduat Superior en Gestió Turística, a Sant Feliu de Guíxols, que acabaríem sent director i doctoranda. Des de llavors, una sèrie d'aconteixements s'han anat produint en moments concrets del temps, fins a portar-nos a poder produir aquesta tesi. Després de tot, només puc dir, GRÀCIES! No tinc més paraules per agrair el què ha representat en **Germà** per aquesta tesi, i sobretot per tot el procés de producció, en què en qualsevol estadi de la recerca, aprofitant el temps al màxim, ha tret el millor de mi, com a persona i com a investigadora. "La probable futura deixeble, s'ha engrescat i ha esdevingut present deixeble. Espero tenir èxit fins arribar a ser exdeixeble!"

També vull agrair de manera molt especial al **Dr. Marc Saez** la possibilitat que em va brindar per ser membre del grup de recerca GRECS i així poder optar a ser becària de recerca UdG. A més, també, pel seu suport en referència als errors estàndards robustos del primer article, i les regions de tolerància de les classes latents, del tercer. No m'oblido de la resta de membres del **GRECS**, gràcies pel vostre suport.

Gracias al **Instituto de Estudios Turísticos** – Ministerio de Industria, Comercio y Turismo – por la cesión de la base de microdatos EGATUR 2010, con la que se ha realizado la totalidad de la presente tesis.

Aquesta tesi ha estat finançada mitjançant la beca de recerca pre-doctoral (BR 18/2011) de la **Universitat de Girona**, així com també mitjançant els ajuts a la recerca i a estades del **Departament d'Economia** i els ajuts puntuals de l'**INSETUR** i de la **Facultat de Turisme**.

Gràcies **CODA family** (Martin, Vera, Marina, Carles, Marc, Pepus, Santi i Glòria) per donar-me l'oportunitat, un cop fet el curs, de presentar la meva recerca en els vostres seminaris, pels

vostres comentaris i suggeriments, molt valuosos per millorar els articles, però també per continuar introduint-me en el món de les dades composicionals.

My thesis has allowed me to undertake two research stays, in which I learned a great deal and had great personal experiences. Zahvaliti se hočem **Ljubici Knežević Cvelbar** in **Tanji Mihalič** za pobabilo na ekonomsko fakulteto univerze v Ljubljani, Danijeli Voljč, ter Adrijani Lazić za logistično in birokratsko pomoč. Hvala lepa.

I would also like to thank **Dr. Philip Long** – Associate Dean of the School of Tourism – University of Bournemouth – for inviting me and for organizing my two month stay there. Furthermore, I would like to give my warmest thanks to **Neelu Seetaram** and **Adam Blake** for the opportunities they provided me with, for introducing me to so many people, for their day to day work and collaboration, for both encouraging and challenging me to continue growing, but overall, for their hospitality. And last, but by no means least, I would like to thank all the **PGR's** I met there; Jessika, Carmen, Phil, Vicky, Yuvraj, Greg, Mirela, and Barbara, who made me feel part of the group – even before I left I started to miss you all!

Gràcies **Juan Luis Nicolau** per donar-me bons consells i facilitar-me articles clau i molt útils per la recerca.

Thanks also to **Barney Griffiths** and **Mary Jane Pratt** for your revisions of what I wrote in English.

Gràcies als membres del **Grup de Doctorands en Turisme**, Dani, Judit, Xavi, Neus, Ariadna per compartir impressions, experiències i converses, per tenir un projecte comú i per la implicació. També a tots els companys i companyes de Turisme i del Departament d'Economia, amb qui he compartit classes, despatx, recerca i altres activitats acadèmiques. Als **amics i amigues** de la Facultat de Turisme i de la Facultat d'Econòmiques amb qui m'he creuat, he esmorzat, dinat, berenat, i m'han escoltat i ajudat en dies bons i no tant bons.

Loi, tu també ets membre de l'equip, i tot el què et voldria dia per agrair-te el teu suport incondicional en aquest projecte, i en d'altres que m'embolico, no hi cap en una sola frase, però hi cap en 7 lletres: GRÀCIES! Ah! Tingues per segur que continuarem veient món junts!

Núria i Alfons, penso que el fet que ens hagueu inculcat valors com l'esforç, la constància, el saber fer i saber estar, l'autoaprenentatge i l'autoexigència, la valentia, l'excel·lència, la generositat, entre molts d'altres, ha fet possible que arribés a escriure aquesta tesi. Crec que sense aquests components, no hagués arribat ni a la meitat.

Finalment, a la resta de la família, especialment a la **Martins,** el **Sergi,** la **Judit,** el **Manel** i la **Remei**, per les preguntes: "Com portes la tesi? Ja has acabat?", "Aquell article que vas enviar a no sé quina revista, te l'han acceptat?", "Ah! Però la revista que t'han acceptat l'article és internacional?", "A on dius que tens el congrés?", "Ja és segur que marxes a Anglaterra?", "Quants articles has de tenir per entregar la tesi?" "Quan presentes?", entre moltes d'altres.

A tots vosaltres, GRÀCIES!

To all of you, THANKS!

Vsem, HVALA

List of publications resulting from the thesis

Martínez-Garcia, E., Ferrer-Rosell, B. and Coenders, G. (2012). Profile of business and leisure travelers on low cost carriers in Europe. *Journal of Air Transport Management*, 20, 12-14.

Ferrer-Rosell, B., Martínez-Garcia, E. and Coenders, G. (2012). Does travelling with low cost affect trip duration? In: Papers of the First Congress of Leisure and Tourism. Escola Universitària del Maresme, Mataró: 129-152. ISBN 978-84-695-3343-7.

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

Ferrer-Rosell, B., Coenders, G. and Martínez-Garcia, E. (accepted). Determinants in tourist expenditure composition – the role of airline type. *Tourism Economics.*

Ferrer-Rosell, B., Coenders, G. and Martínez-Garcia, .E. (under review). Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes. *Current Issues in Tourism.*

List of abbreviations

ALR: Additive log-ratio
AIDS: Almost Ideal Demand System
CAPI: Computer Assisted Personal Interview
CODA: Compositional Data Analysis
DMO: Destination Management Office
EGATUR: Encuesta de GAsto TURístico
IET: Instituto de Estudios Turísticos
ILR: Isometric log-ratio
LCA: Low Cost Airline
MANOVA: Multivariate ANalysis Of VAriance
OLS: Ordinary Least Squares
WTO: World Tourism Organisation

TABLE OF CONTENTS

Index of tables

ABSTRACT

This thesis analyses the two main variables of tourism demand that both economists and tourist policy makers use: length of stay and trip budget. These two variables have been extensively studied for some time now with either micro, macro, cross sectional or longitudinal data. This dissertation carries out a micro cross-sectional analysis of length of stay and tourist expenditure composition among air travellers and aims to observe differences between types of air travellers. In other words, to compare how low cost airline (LCA) and legacy airline users behave in terms of trip duration and distribution of their trip budget, as well as to segment LCA users according to trip budget composition. We do so by using data analysis techniques which we would argue are more adequate than those previously used. This thesis makes contributions to both the variables, some of which are included for the first time as explanatory in the models, and the statistical methods used.

According to WTO rankings, Spain is the fourth most visited tourist destination in the world, and the majority of visitors arrive by air; with LCAs having recently increased their popularity. Thus, in this thesis we analyse the inbound tourism demand of those arriving in Spain by air. To do that, we use data from a 2010 official statistics survey of air travellers to Spain called EGATUR (survey of tourist expenditure) provided by the *Instituto de Estudios Turísticos* (IET – Spanish Institute for Tourism Studies).

While the explanatory variables used in the statistical models include the usual socio-demographic and trip characteristics, this dissertation has two new features. The first is the inclusion of activities undertaken at destination. These activities are strongly related to tourist expenditure at destination, and thus have relevant policy implications for destination management. The second innovative feature we include, in order to compare LCA and legacy airline users, is the variable of how a trip is booked (package trip or self-booking) along with the type of company used (LCA and legacy airline) as moderating effects.

To accommodate the observed multimodality in the variable, length of stay is studied using an ordered logit model rather than the widely used survival models. On the other hand, trip budget composition, (share of transportation expenses, and of basic and

discretionary expenses at destination) is analysed with Compositional Data Analysis (hereafter CODA) methodology and then fitted to a MANOVA to study its determinants, which is in itself new to tourism budget allocation research. Finally, regarding segmentation of LCA users according to budget composition, we propose a new method by hand-picking the more suitable techniques of the CODA methodology and latent class models. This new approach seems to be methodologically sound, and not only can it be tailored to the research questions of its user and be easily used in applied research; it is also completely new to budget research in any field.

The small differences between users of both types of airline are a very relevant and recurrent finding in this thesis. Notwithstanding, package tourists differ considerably from independent travellers for both airline types. Assuming that destination management offices (DMOs) are interested in lengthening the stay of tourists and in increasing their expenditure at destination, they should be just as interested in attracting LCA flights as they are in attracting legacy airline flights.

The results of this thesis show that, Spanish DMOs should focus their marketing efforts on those tourists who travel with LCAs and undertake activities (in particular nautical sports, golf, hiking, visiting a spa, cultural visits, nightlife or visiting friends and relatives). In addition, DMOs should concentrate on those inbound tourists from European countries other than the UK and Ireland, let us say Scandinavians, Germans (especially those using LCAs), Italians and visitors from the Benelux countries (above all those flying with legacy airlines). DMOs should also direct their attention to tourists coming for urban-cultural tourism purposes and especially if travelling with LCAs. By virtue of LCA market segmentation according to trip budget composition, DMOs can benefit from six segments identified as spending differently at destination (more on basic expenditure such as accommodation and food or more on discretionary expenditure such as activities and shopping).

This dissertation makes several key, and above all original, contributions. The first is comparing air travellers by introducing airline type and booking method as moderators in the statistical models. The second contribution is including tourist activities in the models. The third involves the first segmentation by travel budget share, while the fourth refers to the first segmentation of LCA user budgets. Finally, the fifth

contribution is the innovative methodology. We challenge some of the methods previously used to analyse length of stay and trip budget, as well as propose a new method to segment tourists based on their expenditure composition.

Key words: low cost tourism, low cost airlines, ordered logit model, length of stay, Compositional Data Analysis Methodology (CODA), market segmentation, tourist expenditure, latent class models.

RESUM

La present tesi analitza les dues variables més importants de la demanda turística, tant per als economistes com per als responsables de les polítiques turístiques: durada de l'estada i pressupost del viatge. Aquestes dues variables han sigut extensament estudiades des de temps enrere, tant amb dades microeconòmiques, com macroeconòmiques, com de seccions creuades i longitudinals. Aquesta tesi fa una anàlisi microeconòmica i de secció creuada de la durada de l'estada i la composició de la despesa del turista entre els turistes que viatgen en avió amb l'objectiu d'observar diferències entre aquests; dit d'una altra manera, comparar com es comporten els usuaris de les companyies de baix cost i de les companyies tradicionals en termes de dies d'estada i distribució del pressupost del viatge, així com també segmentar els usuaris de baix cost segons la composició del pressupost del viatge. Això es fa utilitzant tècniques d'anàlisis de dades no utilitzades prèviament i que sostenim que són més adequades. La present tesi fa contribucions tant pel que fa a les variables incloses com a explicatives en els models, algunes de les quals es tenen en compte per primera vegada, com pels mètodes estadístics utilitzats.

En aquesta tesi, s'analitza la demanda turística receptora que arriba per via aèria a Espanya – quarta destinació turística segons el rànquing de la OMT – on la majoria dels visitants utilitzen l'avió, i les companyies de baix cost han augmentat recentment la seva popularitat. Per a això, s'han utilitzat dades oficials de l'enquesta EGATUR (enquesta de despesa turística) resposta per turistes que van arribar a Espanya per via aèria l'any 2010, i facilitades per l'Instituto de Estudios Turísticos – IET –.

Les variables explicatives utilitzades en els models estadístics inclouen les habituals característiques sociodemogràfiques i del viatge. A més, aquesta tesi té dos nous trets distintius. El primer és la inclusió de les activitats realitzades a la destinació, que estan estretament relacionades amb la despesa turística realitzada i, per tant, tenen implicacions polítiques pertinents en relació amb la gestió de destinacions. En segon lloc, per tal de comparar els usuaris de companyies de baix cost i companyies tradicionals, s'inclouen la variable com es va reservar el viatge (amb un paquet tancat o individualment), i el tipus de companyia aèria (baix cost o tradicional) com a efectes moderadors.

Pel que fa a la metodologia, la durada de l'estada s'estudia amb un model logit ordenat, en lloc dels models de supervivència utilitzats prèviament, per tal de donar cabuda a la multimodalitat observada en la variable. D'altra banda, la composició del pressupost del viatge (part de les despeses de transport i de les despeses bàsiques i discrecionals en destinació) s'analitza amb la metodologia de dades composicionals (Compositional Data Analysis – CODA) i després s'ajusta en un model MANOVA (Anàlisi de la Variància Multivariant) per estudiar els seus determinants, fet que representa una novetat en la recerca en pressupost del viatge. Finalment, pel que fa a la segmentació dels usuaris de companyies de baix cost segons la composició del pressupost, en aquesta tesi es proposa un nou mètode que selecciona les tècniques més adequades de la metodologia CODA i dels models de classes latents. Aquest nou enfocament sembla metodològicament adequat, es pot adaptar a les preguntes de recerca de l'investigador, pot ser utilitzat fàcilment pels investigadors aplicats i també és nou en el camp d'investigació pressupostària.

Un resultat rellevant i recurrent d'aquesta tesi fa referència a les petites diferències trobades entre els usuaris dels dos tipus de companyies. Per contra, els turistes que viatgen en paquet difereixen d'ambdós. Tenint en compte que als gestors de destinacions els interessa allargar l'estada dels turistes i que incrementin les seves despeses a la destinació, no tenen per què estar menys interessats en atraure vols de companyies de baix cost que de companyies tradicionals.

Els resultats d'aquesta tesi, demostren que els gestors de destinacions a Espanya haurien d'enfocar els seus esforços de màrqueting en aquells turistes que realitzen alguna activitat a la destinació (especialment els que fan esports nàutics, golf, visites culturals, senderisme, spa, entreteniment o visites a amics i familiars) i viatgen en companyies de baix cost; en aquells turistes d'altres països europeus que no siguin Regne Unit i Irlanda, com són els dels països nòrdics, alemanys (sobretot els que utilitzen companyies de baix cost), italians i belgues (sobretot els que volen en companyies tradicionals): i també els turistes que venen per motivació de turisme urbà-cultural, i viatgen majoritàriament en companyies de baix cost. Pel que fa a la segmentació en base a la distribució del pressupost del viatge, els gestors de destinacions es poden beneficiar de la identificació de sis segments que gasten el pressupost del viatge de

manera diferent (més en despeses bàsiques com l'allotjament i el menjar, o més en despesa discrecional, com activitats i compres).

Aquesta tesi fa diverses contribucions originals. La primera contribució és en termes de comparar els turistes que viatgen en avió mitjançant la introducció de les variables tipus de companyia i mode de reserva com a efectes moderadors en els models estadístics. La segona contribució fa referència a la inclusió de les variables d'activitats. La tercera fa referència a la segmentació segons composició del pressupost del viatge. La quarta és la segmentació dels usuaris de companyies de baix cost. Finalment, la cinquena fa referència a la metodologia. Qüestionem alguns dels mètodes utilitzats anteriorment per analitzar la durada de l'estada i el pressupost del viatge, així com també proposem un nou mètode per segmentar turistes en base a la composició de la despesa.

Paraules clau: Turisme de baix cost, companyies de baix cost, model logit ordenat, durada de l'estada, anàlisi de dades composicionals (CODA), segmentació de mercats, despesa del turista, models de classes latents

RESUMEN

La presente tesis analiza las dos variables más importantes de la demanda turística, tanto para los economistas como para los responsables de las políticas turísticas: duración de la estancia y presupuesto del viaje. Estas dos variables han sido extensamente estudiadas desde tiempo atrás, tanto con datos microeconómicos, como macroeconómicas, como de secciones cruzadas o longitudinales. Esta tesis hace un análisis microeconómico y de sección cruzada de la duración de la estancia y la composición del gasto del turista entre los turistas que viajan en avión con el objetivo de observar diferencias entre ellos; dicho de otro modo, comparar cómo se comportan los usuarios de las compañías de bajo coste y de las compañías tradicionales en términos de días de estancia y distribución del presupuesto del viaje, así como también segmentar los usuarios de bajo coste según la composición del presupuesto del viaje. Esto se hace utilizando técnicas de análisis de datos no utilizados previamente y de las que sostenemos que son más adecuadas. La presente tesis hace contribuciones tanto a las variables incluidas como explicativas en los modelos, algunas de las cuales se tienen en cuenta por primera vez, como por los métodos estadísticos utilizados.

En esta tesis, se analiza la demanda turística receptora que llega a España por vía aérea – cuarto destino turístico según el ranking de la OMT – donde la mayoría de los visitantes utilizan el avión, y las compañías de bajo coste han aumentado recientemente su popularidad. Para ello, se han utilizado datos oficiales de la encuesta EGATUR (Encuesta de Gasto Turístico) respuesta por turistas que llegaron a España en avión el año 2010, y facilitados por el Instituto de Estudios Turísticos (IET).

Las variables explicativas utilizadas en los modelos estadísticos incluyen las habituales características sociodemográficas y del viaje. Además, esta tesis tiene dos nuevos rasgos distintivos. El primero, es la inclusión de las actividades realizadas en el destino, que están estrechamente relacionadas con el gasto turístico realizado, y por lo tanto tienen implicaciones políticas pertinentes en relación con la gestión de destinos. En segundo lugar, a fin de comparar los usuarios de compañías de bajo coste y compañías tradicionales, se incluyen la variable de cómo se reservó el viaje (con un paquete cerrado o individualmente), y el tipo de compañía aérea (bajo coste o tradicional) como efectos moderadores.

En cuanto a la metodología, la duración de la estancia se estudia con un modelo logit ordinal en vez de los modelos de supervivencia utilizados previamente, para dar cabida a la multimodalidad observada en la variable. Por otra parte, la composición del presupuesto del viaje (parte de los gastos de transporte y de los gastos básicos y discrecionales en destino) se analiza con la metodología de datos composicionales (Compositional Data Analysis - CODA) para posteriormente ajustarla a un modelo MANOVA (Análisis de la Varianza Multivariante) para estudiar sus determinantes, lo que representa una novedad en la investigación en presupuestos del viaje. Finalmente, en cuanto a la segmentación de los usuarios de compañías de bajo coste según la composición del presupuesto, en esta tesis se propone un nuevo método que selecciona las técnicas más adecuadas de la metodología CODA y los modelos de clases latentes. Este nuevo enfoque parece metodológicamente adecuado, se puede adaptar a las preguntas de investigación del investigador, puede ser utilizado fácilmente por los investigadores aplicados y también es nuevo en el campo de investigación de presupuestos.

Un resultado relevante y recurrente de esta tesis hace referencia a las pequeñas diferencias encontradas entre los usuarios de los dos tipos de compañías. Por el contrario, los turistas que viajan en paquete difieren de los usuarios de ambos tipos de compañías. Teniendo en cuenta que a los gestores de destinos les interesa alargar la estancia de los turistas y que éstos incrementen sus gastos en el destino, no tienen por qué estar menos interesados en atraer vuelos de compañías que bajo coste que de compañías tradicionales.

Los resultados de esta tesis, demuestran que los gestores de destinos en España deberían enfocar sus esfuerzos de marketing en aquellos turistas que realizan alguna actividad en el destino (especialmente los deportes náuticos, golf, visitas culturales, senderismo, spa, entretenimiento, visitas a amigos y familiares) y viajan en compañías de bajo coste, en aquellos turistas de otros países europeos que no sean el Reino Unido e Irlanda, como son los de los Países Nórdicos, alemanes (sobre todo los que utilizan compañías de bajo coste), italianos y belgas (sobre todo los que vuelan en compañías tradicionales); y también los turistas que vienen por motivación de turismo urbano-cultural, y viajan mayoritariamente en compañías de bajo coste. En cuanto a la segmentación en base a la distribución del presupuesto del viaje, los gestores de destinos se pueden beneficiar de

la identificación de seis segmentos que gastan el presupuesto del viaje de manera diferente (más en gastos básicos, como el alojamiento y la comida, o más en gasto discrecional, como actividades y compras).

Esta tesis hace varias contribuciones originales. La primera contribución es en términos de comparar los turistas que viajan en avión mediante la introducción de las variables tipo de compañía y modo de reserva como efectos moderadores en los modelos estadísticos. La segunda contribución se refiere a la inclusión de las variables de actividades. La tercera se refiere a la primera segmentación según composición del presupuesto del viaje. La cuarta, es la segmentación de los usuarios de compañías de bajo coste. Finalmente, la quinta hace referencia a la metodología. Cuestionamos algunos de los métodos utilizados anteriormente para analizar la duración de la estancia y el presupuesto del viaje, así como también proponemos un nuevo método para segmentar turistas en base a la composición del gasto.

Palabras clave: turismo de bajo coste, compañías de bajo coste, modelo logit ordinal, duración de la estancia, metodología de datos composicionales (CODA), segmentación de mercados, gasto del turista, modelos de clases latentes.

1   GENERAL INTRODUCTION

According to the neoclassical theory of consumer demand, money and time that an individual can use to consume goods are the main constraints on tourism demand. Money is a major constraint for most tourists because tourist services are a major deduction from their budget, while at the same time the possession of a stock of disposable time constrains the ability of the tourist to do all the things desired at destination (Bull, 1995). Like income, disposable time will vary between market segments, countries, and individuals, thereby determining tourism consumption.

From an empirical point of view, demand may be studied with data from a specific origin to multiple destinations (i.e. ex ante, see Divisekera, 2009; Li, Song and Witt, 2004) or from multiple origins to a specific destination; which is the case in this thesis (i.e. ex post, see Divisekera, 2009). In ex ante studies it is possible to conceptualize tourism consumption as a multi-stage process. In the first stage, tourists initially determine how much of their time and budget they can allocate to tourism consumption. In the second stage, the tourist decides how much time and budget to designate to each trip, and in the third stage the tourist decides how much of the trip budget to destine towards goods and services, including transportation, accommodation, and so forth. Ex post studies, as in this thesis, only consider the final stage.

According to Lim (1997) and Song and Li (2008), empirical studies of demand are mostly based on ex post data and define demand as the number of arrivals at the destinations. The authors stipulate that tourist expenditure, after tourist arrivals, is the second most frequent branch of research in tourism demand models (See Sainaghi, 2012, for an overview about tourist expenditure research). On the other hand, length of stay is another of the most commonly used explained variables in tourism demand models. Furthermore, both length of stay and expenditure are used to calculate the economic impact in the host destinations (Archer and Fletcher, 1996; Dwyer, Forsyth and Dwyer 2010; Mihalič, 2002; Sinclair, 1998). However, this thesis focuses on the analysis of determinants of tourist expenditure and length of stay, rather than on the economic impact.

The type of data used to observe tourism demand (level of tourist expenditure, and length of stay at destination, among others) can be distinguished as micro or macro data. Crouch (1994), Li, Song and Witt (2005), Lim (1997; 2006), Song and Li (2008), Song, Dwyer, Li and Cao (2012) reviewed studies carried out on general or macro tourism demand. This dissertation focuses on micro level analysis. The data sources of micro data are typically general population surveys and visitor surveys at a particular destination. This thesis uses the latter.

The following lines are related to research done on the two main constraints, time and money available, which, in terms of microeconomic analysis, are also the two main explained variables.

Firstly looking at budget restriction, Wang and Davidson (2010), who reviewed the micro-analyses of tourist expenditures, concluded that further micro-analysis is needed, because tourist demographic characteristics and trip related variables change tourist purchasing behaviour. The authors classified the micro-economic studies into three types; individual's choice, individual's expenditure and tourism price. They also stated that the most common method in these studies is the multiple regression model. As far as dependent variables are concerned, the most common variable used is total trip expenditure, followed by expenditure per person and per day (24 out of 27 studies in their review concern the prediction of such an aggregated expenditure variable). On the other hand, the most widely used independent variables are income, age, place of residence and gender, as well as length of stay, size of party and seasonality. Another recent review of microeconomic models on determinants of tourist expenditure was undertaken by Brida and Scuderi (2013). In that case, the authors analyzed 86 articles using cross-section econometric models to observe the determinants of tourist expenditure from individual data. They concluded that future research should move forward and explore new methods, use psychographic variables and relate the results to the economic theory. According to the authors, large numbers of studies have focused mainly on analyzing tourist expenditure in absolute terms. This tourist expenditure could be total expenditure for the whole trip or expenditure per day, per person, or both together.

Furthermore, there is another interesting branch in tourist expenditure analysis, namely expenditure composition; also referred to as budget share. The analysis of expenditure composition provides valuable information for destination management in terms of type of tourist according to how they distribute their travel budget. The travel budget can be broken down into non-discretionary and discretionary components. The non-discretionary are those in accommodation, transportation and food where a minimum amount has to be spent, whereas discretionary expenditure is that which includes extra activities, shopping, and moving around at the destination, among others. This distinction between expenditure components is relevant for destinations managers as they are more interested in local spending than in spending that is totally or partially made outside the destination as, for example, those paid to tour operators or airline companies. Moreover, it is already well-known that tourists travelling to the same place in the same time period might spend their budgets in very different ways, which will have different impacts on the economy of the destination (Legohérel, 1998; Legohérel and Wong, 2006). To our knowledge, there are only two studies that predict tourist budget share from individual characteristics (Fleischer, Peleg and Rivlin, 2011 and Coenen and Van Eekeren, 2003) and there are no existing studies using budget share as a market segmentation variable.

As far as expenditure is concerned, this thesis focuses on using an appropriate methodology to analyse tourist expenditure as a composition. Compositions are expressed as percentages of a total budget and their sum can only be 100. Compositional data thus lie in a restricted space and only convey information regarding the relative size of the components to one another. By transforming the compositions by means of logarithms of ratios, components can be estimated and interpreted as a whole, taking into account the fact that the relative importance of one component can only increase if the relative importance of at least one other decreases. While this approach is frequently used in family budget studies (Fry, 2011) it has never been used in tourism budget studies.

Moving onto time, which is the second constraint of tourist demand, length of stay at destination has long been analysed (Archer and Shea, 1975; Fleischer and Pizam, 2002; Mak and Moncur, 1979; Mak, Moncur and Yonamine, 1977; Silberman, 1985; Thumberg and Crotts, 1994, among others). Since 2008, the number of studies carried

out in this field has increased considerably (see Alegre, Mateo and Pou, 2011; Alegre and Pou, 2006, 2007; Barros and Machado, 2010; Martínez-Garcia and Raya, 2008; Menezes, Moniz and Vieira, 2008; Thrane, 2012; Yang, Wong and Zhang, 2011, among others).

Most studies on length of stay include both trip and stay characteristics and socio-demographic variables as explanatory variables in the model. In the case of the former, the most common variables are travel cost, destination attributes, organization, motivation, repeat visits to the same destination, accommodation, and travelling group. Some studies have also considered distance, destination, season, time of booking, number of trips per year/experience in travelling abroad, and satisfaction. When it comes to studying the length of stay variable, this represents a statistical challenge, as reflected by the various and complex methods used in the past, including Poisson models, negative binomial models, and survival models; the latter being the most common in recent studies. However, like Alegre and Pou (2006) and Alegre et al. (2011), we encountered multimodality in the length of stay variable, and it made no sense to us to use unimodal models (which include Poisson models, negative binomial models, survival models or even Ordinary Least Squares – OLS). Thus, as in the case of the tourist expenditure composition analysis, this thesis makes a methodological contribution by using, for the first time, an ordered logit model to account for multimodality in the length of stay variable.

Another interesting branch of research presented here, along with the two main tourism demand restrictions, is the analysis of Low Cost Airline (LCA) users. Some years ago nobody could have imagined that nowadays LCAs would be playing such a key role in the European, and above all in the Spanish, tourism industry. Year after year LCAs have steadily increased their market share and attract a very heterogeneous market. Some legacy companies have adapted their business models to also offer products similar to those of the LCA, hence making the differences between LCA and legacy airlines minute in some cases.

Since the consolidation of LCAs, a large body of researchers has carried out studies into numerous aspects related to the LCA phenomenon. In fact, in recent years there has been a boom in the number of studies made related to LCAs and their demand, effects

and impacts, as well as relationships with airports and destinations, among others. One branch of research has focused mainly on the demand differences between LCA and legacy airlines (see Chiou and Chen, 2010 – behavioural intentions –; Forgas, Moliner, Sánchez and Palau, 2010 – airline passenger loyalty –; O'Connell and Williams, 2005 – airline perception –, among others). So, it would seem to us that there is room to analyze the differences between them when it comes to length of stay and expenditure. There is another line of research focused on LCA users and their heterogeneity (see Castillo-Manzano and Marchena-Gómez, 2011 – airline choice; Martínez-Garcia, Ferrer-Rosell and Coenders, 2012 – business passengers, see appendix; Kim and Lee, 2011 – airline satisfaction; Martínez-Garcia and Royo-Vela, 2010 – segments based on flight characteristics; Raya-Vilchez and Martínez-Garcia, 2011 – nationality; among others). So, in this case it would also seem to us that there is room to analyze the differences among LCA users as well.

According to the WTO (2013) Spain is ranked the 4th most popular tourism destination in the world. According to the IET (Instituto de Estudios Turísticos – the Spanish Institute for Tourism Studies), Spain received 52.7 million tourists in 2010, 77% of whom travelled by air. Of that 77%, 56% flew with an LCA and 84.3% lived in a European country; which means European countries are the main inbound markets for Spanish destinations. Since 2003, LCAs have experienced significant expansion and consolidation in Spain. In 2003, according to the IET, 23.5% of all tourists who arrived in Spain by plane used an LCA, this percentage rose to 29.2% in 2004, 29.7% in 2005, 35.1% in 2006, 39.9% in 2007, 48.4% in 2008, 54.1% in 2009, and to 56% in 2010. It is well worth noting that the total number of tourists who visited Spain in 2003 was 52.5 million and in 2010 52.7 million. These data show how relevant to this study the tourism industry in Spain is, and how appropriate it is to study the expansion of LCAs in the context of the Spanish tourism industry. However, to the best of our knowledge, there are very few studies which have analysed the tourist demand in Spain and have also related it to LCAs.

The studies which do have the whole country as their scope of analysis, have mostly dealt with forecasting, modelling demand, assessing tourist elasticities, analysing the effect of low cost airlines on tourism, and analysing the probability of tourism consumption (Alegre and Pou, 2004; Alegre, Mateo and Pou, 2009; Garcia-Ferrer and

Queralt, 1997; Garín-Muñoz, 2006; 2007; 2011; Garín-Muñoz and Perez Amaral, 2000; González and Moral, 1995; 1996; Nicolau and Más, 2005; 2006; 2009; Rey, Myro and Galera, 2011, among others).

On the other hand, there is quite a long list of studies which have analysed various aspects of tourism demand in general, as well as length of stay or tourist expenditure, in specific regions, or even islands and municipalities in Spain (Aguiló and Juaneda, 2000; Aguiló, Riera and Roselló, 2005; Alegre and Cladera, 2010; Alegre, Cladera and Sard, 2011; Alegre and Garau 2010; Alegre and Juaneda, 2006; Alegre et al., 2011; Alegre and Pou, 2006; 2007; Artal-Tur, García-Sánchez and Sánchez-Garcia, 2008; Calderón and Ruiz, 2002; Díaz-Pérez, Bethencourt-Cejas and Álvarez-González, 2005; Garín-Muñoz, 2006; Lopez-Delgado, Alarcón-Urbistondo and Martín-Armario, 2000; Martínez-Garcia and Raya, 2008; 2009; Medeiros, McAleer, Slottje, Ramos and Rey-Maquieira, 2008; Medina-Muñoz and Medina-Muñoz, 2012; Perdellas de Blas, Padín-Fabiero and Bouzada-Fernández, 2003; Pulido-Fernández and Sánchez-Rivero, 2010; Raya-Vilchez and Martínez-Garcia, 2011; Rodríguez, Dávila and Rodríguez, 2003; Roselló, Riera and Sansó, 2004; among others).

The list of research carried out into LCA tourist demand in Spain is a little shorter (Castillo-Manzano and Marchena-Gómez, 2011; Espino, Martín and Román, 2008; Forgas et al., 2010; Martínez-García et al., 2012; Martínez-García and Raya, 2008; 2009; Martínez-García and Royo-Vela, 2010; Nicolau, 2011; Raya-Vilchez and Martínez-Garcia, 2011; Rey et al., 2011).

In terms of geographical scope, this thesis takes a whole country (i.e. Spain) into account, that is, it analyses the two main aspects of tourism demand (stay and expenditure) in Spain as a whole tourist destination, unlike the majority of the literature which focuses its research on regional destinations, islands or even municipalities.

On having reviewed the available literature, we realised that there was a gap in the tourism demand literature which could be filled with the research presented in this dissertation. This dissertation aims to make a transversal analysis of LCAs by taking into account two of the most important determinants/restrictions/constraints of tourism demand; expenditure and its composition, and length of stay. The research includes the

following novel features which, to our knowledge, no other author has ever contemplated before:

- ➢ compare the length of stay determinants for LCA and legacy airline users
- ➢ compare the tourist expenditure composition between LCA and legacy airline users
- ➢ analyse the tourist expenditure composition using appropriate statistical methods
- ➢ segment tourists according to travel budget share
- ➢ segment the expenditure of LCA users

Now that the research gaps have been revealed and to close this introductory section, we would like to point out that this thesis is divided into three articles. The first article is about length of stay, the second is about the determinants of tourist expenditure composition and the third is about segmenting low cost users based on their expenditure composition.

This thesis is structured as follows: once we have introduced the focus, we present the general and specific objectives. Then, we include a description of the data and the methods used along with the transcription of the three thesis articles. We continue with a summary and discussion of the main results and contributions, and finally, we present the conclusions.

This dissertation makes several innovative contributions. The first is to compare air travellers by using airline type and booking method as moderators in the statistical models. The second is the inclusion of tourist activities in the models. The third refers to the first segmentation by travel budget share, while the fourth contribution refers to the first segmentation of LCA user budgets. Finally, the fifth unique contribution is in the methodology used. We not only challenge some of the methods previously used to analyse length of stay and trip budget, but we also propose a new method, based on tourist expenditure composition, with which to group tourists by.

2 OBJECTIVES OF THE THESIS AS A WHOLE

This section is dedicated to the objectives of the thesis. There are, firstly, the general objectives of this dissertation as a whole, and secondly, the specific objectives for each article.

The general objectives of the thesis are to:

> ➢ Analyse the determinants of air travellers' length of stay as an ordered categorical variable (week-end, long week-end, one week, ten days, two and three weeks, and more than three weeks).

> ➢ Study the composition of tourist expenditure, that is, the share of tourist expenditure allocated to the different categories of a travel budget (proportion of the total travel expenditure devoted to transportation, to accommodation and food, and to activities, shopping and moving around).

> ➢ Compare low cost and legacy airline travellers using moderating effects. The airline type has been used as a moderator variable, thus modifying the effect of all exogenous variables on length of stay and travel budget share.

> ➢ Explore expenditure heterogeneity amongst low cost airline travellers by market segmentation.

> ➢ Develop an appropriate method to segment tourists based on their expenditure composition, by hand-picking a combination of statistical techniques which is methodologically sound, can be tailored to the investigator's research questions and can be easily used in applied research. This has been done to take a step further into exploring heterogeneity amongst low cost airline travellers.

The three articles, and in general the whole thesis, share the scope of analysis which refers to those European tourists who arrived in Spain by plane in 2010 and stayed less than 120 nights.

The following outlines the specific objectives of each of the three articles in this thesis.

The first article, which is entitled "Package and no-frills air-carriers as moderators of length of stay", has been published in *Tourism Management* and focuses on time

restriction at destination, i.e. it studies the length of stay. The objective is to analyse the effects of how the trip is booked on length of stay; albeit as a moderator effect on the other explanatory variables. It considers three different booking methods: 1) booking a package tour where the airline is already included and the tourist cannot choose the airline type; 2) self-booking and flying with an LCA; and 3) self-booking and flying with a legacy airline. Besides the usual explanatory variables relating to individual, trip and stay characteristics, the moderating effects of how the trip is booked and the remaining explanatory variables are of key interest in order to observe how effects change according to how the trip is booked. Also, for the first time in length of stay research, the article systematically analyses activities undertaken at the destination. Yet another new feature of this article is analysing length of stay as an ordinal variable.

Moving onto the other tourist demand restriction, which is budget restriction, the objective of the second article, entitled "Determinants in tourist expenditure composition – the role of airline types" and which has been accepted for publishing in *Tourism Economics*, is to study the determinants of the composition of tourist expenditure, in other words, the share of tourism expenditure allocated to the different categories of a trip budget, by taking into account tourist heterogeneity and distinguishing between legacy and LCA market segments. Two main research questions are addressed:

➢ How does travelling with an LCA vs. a legacy airline affect the distribution of trip expenses between transportation and other costs and the distribution of the non-transportation expenses between discretionary and non-discretionary?

➢ Do passenger characteristics affect budget composition? If so, do they affect it in the same way for both types of airline? In other words, does the airline type have a moderating effect?

As previously mentioned in the introduction, non-discretionary expenses are those which are unavoidable when travelling (i.e. accommodation, food and transportation), and discretionary expenses are those elements in the trip budget which may be optional (i.e. activities, shopping, moving around, etc.).

Furthermore, in order to expand the research on low-cost-user heterogeneity (in terms of budget restriction and the composition of tourist expenditure), segmenting the LCA

users based on their expenditure structure is of key interest. Our own research on length of stay and tourist expenditure composition in the first two articles concludes that differences between LCA and legacy-airline users are not that relevant. The literature points out that LCA users are growing in number and becoming as heterogeneous as legacy airline users. There is no precedent study segmenting tourists based on expenditure share, not to mention the specific LCA case. Furthermore, all segmentation methods previously used for this objective are statistically flawed in some respect. Thus, we think that for academia, it is worthwhile proposing a new method to provide an easy, sound and useful tool to fulfil this objective.

The third article of this thesis is entitled "Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes", has been submitted and is under revision. The article focuses on developing a new method for segmenting tourists, based on travel budget share, by combining compositional data analysis and latent class models. The method properly treats the restrictions in budget share data, and allows researchers to make statistical inference on the relationship between segments and external variables. Thus, the objective of the third article is twofold. Firstly, to put forward an appropriate method to study heterogeneous market segments according to their travel budget share and secondly, to carry out the empirical analysis of expenditure segments of LCA users in Spain; which is also presented as the method illustration.

3 DATA AND METHODS USED

This chapter is structured into two distinct sections based on methodology. The first section is related to the data used, while the second is related to the innovative methods introduced and used in the thesis which make, from our point of view, novel and relevant contributions to length of stay and tourist expenditure composition research.

3.1 Data

Data used are 2010 secondary official statistics data, provided by the IET. The IET is an official agency of the Spanish Ministry of Industry, Energy and Tourism and produces the bulk of tourism data in Spain. The survey used for this thesis is known as the *Encuesta de Gasto Turístico* (*EGATUR*), in which tourism expenditure and other tourist information, such as trip information and tourist socio-demographic characteristics are gathered. The EGATUR survey is conducted in 23 major Spanish airports, using CAPI (Computer Assisted Personal Interview) to interview tourists leaving the country. The sample is non-proportionally stratified by country of residence, airport and month (Table 1).

Table 1. EGATUR sample design

Universe	Foreign visitors arriving by air to Spain (island groups included), through 23 major airports
Regularity	Monthly
Sampling method	Random Stratified by airport (Aiports grouped in 9 strata) Quotas by country of residence and month
Sample size	73,000 air travellers per year
Data collection	Computer Assisted Personal Interviews conducted at the airport while tourists waited for the flight home

Source: IET – EGATUR methodology.

For the articles included in this thesis, we have used a subset of EGATUR universe which consists of European leisure visitors arriving by air and spending at least one night in Spain. For the sake of comparability between airline types, we have excluded flights from outside Europe because LCAs mostly operate short-haul flights. We focused the study on only those trips with one single destination, thus excluding multi-stage trips, as the decision process regarding length of stay and expenditure composition

for these trips is expected to fundamentally differ from that of single-stage trips. Stays of over 120 days have also been excluded. Table 2 summarizes which individuals have been used in each article.

Table 2. Individuals taken into account according to the objectives of each article

	First article	Second article	Third article
Package travellers	✓ Included	✗ Excluded	✗ Excluded
LCA travellers	✓ Included	✓ Included	✓ Included
Legacy airline travellers	✓ Included	✓ Included	✗ Excluded
Business travellers	✗ Excluded	✗ Excluded	✗ Excluded
Study travellers	✗ Excluded	✗ Excluded	✗ Excluded
Travellers staying at owned accommodation or with friends or relatives	✓ Included	✗ Excluded	✗ Excluded
Travellers who do not pay for the trip themselves (presents, contests, etc.)	✓ Included	✗ Excluded	✗ Excluded

Although detailed information is given in the respective papers, we provide some information about the sample characteristics. Regarding the composition of the largest sample, which is that selected for the first article, 49.4% were female, 8.7% were aged 65 and above, 29% between 45 and 64, 49.6% between 25 and 44, and 12.8% under 24. 30.3% came from the United Kingdom, 16.3% from Germany, 10.8% from Italy, and the rest were mostly from Benelux and Scandinavian Countries. They mostly had university education (62.8%). 28.6% were package travellers, 49.0% flew with a LCA and 22.4% flew with a legacy airline.

3.2 Statistical Methods

In the research done about the temporal restriction, that is, length of stay at destination (first article), the analysis presented in this thesis has used an ordered logit model as a means to account for the observed multimodality in the length-of-stay variable. Even though there were other authors who had already encountered the multimodality of the length of stay variable and used a binary logit model (Alegre and Pou, 2006 –for only

two modes–), a multinomial logit model (Ferrer-Rosell, Martínez-Garcia and Coenders, 2012 –general case–, see appendix), the literature generally ignores the multimodality issue.

Since tourists' length of stay is integer and positive, the most common methods used up until now seemed, on the surface, to be appropriate. These methods are the survival model (Barros, Butler and Correia, 2010; Barros, Correia and Crouch, 2008; Barros and Machado, 2010; Gokovali, Bahar and Kozak, 2007; Hong and Jang, 2005; Machado, 2010; Martinez-Garcia and Raya, 2008; 2009; Menezes et al., 2008; Peypoch, Randriamboarison, Rasoamananjara and Solonandrasana, 2012; Thrane and Farstad, 2012; Wang, Little and Delhomme-Little, 2012), Poisson regression (Rodríguez et al., 2003) or negative binomial regression (Nicolau and Más, 2006). Thrane (2012) suggested that, since the data generation process assumed by survival models is not tenable for length of stay, there is not much to be gained over simpler and more widely understood methods such as linear OLS regression. All these methods assume unimodality.

In this thesis a further step is taken by using the ordered logit model to account for the multimodality of the duration variable. We have encountered five duration modes, and the duration variable has been grouped into 5 categories around these modes. The advantage of the ordered logit model over the multinomial is basically parsimony, which is essential for heavily parametrized models with moderating effects. Ferrer-Rosell et al. (2012, see appendix) attempt to use the multinomial model on the same data and some key moderating effects fail to be identified.

The research carried out on budget restriction or tourist expenditure (articles 2 and 3), has been analysed globally, in absolute terms per budget parts, as a part in itself in family budgets and, as in this thesis, in relative terms (share) per budget parts.

To treat the expenditure variable in relative terms, this thesis is again innovative in the method used for analysing the tourist expenditure composition, because it uses the Compositional Data Analysis (CODA) methodology, which analyses ratios of components. Statistical analysis of budget compositions is a methodologically challenging task. Share in budgets, as any other composition, are expressed as

proportions or percentages of a total, whose sum can only be 1 (or 100). Compositional data lie in a restricted space and only convey information regarding the relative size of components to one another. The seminal work of Aitchison (1986) started a fruitful tradition in CODA of which the most widely used technique is the transformation of compositions. This is achieved by means of logarithms of ratios. Working with log-ratios not only has methodological implications but also substantive ones. Without log-ratios, components are estimated and interpreted separately from one another as if they could vary independently (*ceteris paribus*), which is impossible: the relative importance of one component (budget share) can only increase if the relative importance of at least one other decreases. Most methods used to model demand and budget share such as the *almost ideal demand system-* AIDS (Deaton and Muellbauer, 1980) ignore, at least partly, the constraints and distributional nature of compositional data. The AIDS is designed to analyse the interdependence of budget allocations, thus, it overcomes the limitation of single-equation modelling. It fits share as dependent variables in a set of simultaneous regressions. The approach has been mainly focussed on estimating price and income elasticities, using macroeconomic data (Divisekera, 2007; 2009; 2010; Fujii, Khaled and Mak, 1985; O'Hagan and Harrison, 1984; Syriopoulos and Sinclair, 1993; Wu, Li and Song, 2011). In microeconomic studies, the AIDS is also used to estimate elasticities, and individual characteristics are used only as controls (Coenen and van Eekeren, 2003 and Fleisher et al., 2011). To ensure compositional coherence of predicted budget share (i.e. unit sum and non-negativity) a set of parameter constraints is imposed to the AIDS. However, the presence of an error term with an unbounded distribution (usually normal), results in a non-zero probability that actual share lies outside the [0,1] interval (McLaren, Fry and Fry, 1995; Fry, Fry and McLaren, 1996; Fry, 2011). In other words, the bounded distribution of budget share results in a misspecification of the almost ideal demand system and of any model fitting percentage share with an unbounded error distribution. Despite the limitations of the AIDS for analysing compositions, compositionally coherent alternatives have very rarely been used in general budget research (McLaren et al., 1995; Fry et al., 1996; Fry, 2011; Houthakker's 1960; Bewley, 1982; Bewley and Fiebig, 1988) and never in travel budget research.

With the CODA methodology, once components have been transformed by means of log-ratios, standard statistical techniques may be used. This makes for a wide

applicability and usefulness for fixed sum data in tourism. This includes, for instance, micro tourist time use research (24 hour sum) or macro analysis of variables, such as outbound market share (country of origin variable, 100 percentage point sum).

Another innovation regarding the CODA methodology is introduced in this thesis, specifically in the third article. It refers to the Isometric Log-Ratio transformation (ILR), which has never been used for budget share research in any field. The ILR transformation of compositions is more flexible (the denominator does not have to be the same component in all ratios) and interpretable than the additive log-ratio transformation (ALR) proposed by Fry et al. (1996) and Fry, Fry and McLaren (2000). The ILR can be tailored to the research questions and it is distance preserving, which is a key point for the classification purpose. The ILR defines weighted log-ratios of geometric means of components called balances (Egozcue and Pawlowsky-Glahn, 2005).

Still on the subject of the third article, there is no scholarly work fulfilling the aim of segmenting tourists (in our case LCA users) based on their travel budget share, and thus no useful previous methodology reference. A wide range of methods is encountered when it comes to tourism segmentation problems using other variables (Dolnicar, 2002), among which cluster analysis is the most common. Then some statistical model (e.g. a multinomial logit model) can be used to explain class-membership from external antecedent variables and/or explain some outcome variables from the clusters. However, this strategy has two main drawbacks: estimates are biased and standard errors are incorrect (Clark and Muthén, 2009; Liu, Kiang and Brusco, 2012; Muthén, 2002). This is so because the analysis does not take into account the uncertainty of the classification but rather treats it as an observed variable. Instead, latent class methods for clustering perform random model-based classifications (e.g. Magidson, 2002; McLachlan and Basford, 1988). They employ finite mixture models to define a set of unobserved subpopulations.

In the last article of the thesis, we have hand-picked the most suitable statistical techniques related to the CODA methodology and latent class models, to propose a new method which is able to analyse the expenditure profiles of tourists by segmenting them according to their expenditure composition. This combination consists of an up-to-date

mix which especially suits the tourist budget segmentation problem. Firstly, we suggest using ILR transformations, which allows researchers to build tailor-made log-ratios which are easy to interpret and suit the research questions. Secondly, regarding the latent class models, we have adapted the flexible specification of the model to the log-ratios by keeping in the model both variances and covariances constrained to be equal across classes. On the other hand, when it comes to selecting the number of classes, it is useful to take into account the standard statistical criteria which can help to decide that number, but, it is also relevant to consider the interpretability of market segments, since for management purposes the latter is much more important, and it is what we have emphasized in our research. Finally, regarding the statistical inference and the relationship between the segments found and the external variables, we also innovate in using the most recent multiple-imputation based methods. By using this new method (pseudo class-draws) predictors are included in the model without modifying the latent classes by means of a multinomial logit and class membership is recognised as a random variable (Clark and Muthén, 2009). What has to be decided carefully when using this method is the treatment of each particular variable. Variables which may be decided at the same time as the trip budget composition (for example type of accommodation, activities undertaken, destination, and duration) should not be included as predictors but as outcomes. The pseudo class-draw method developed by Clark and Muthén (2009) can be adapted to comparing means and proportions of outcomes across classes.

4 TRANSCRIPTION OF THE THESIS' ARTICLES

This thesis is presented as a compendium of the three articles whose references follow below. All other articles on LCA demand published during my PhD studies are included in the appendix to the thesis.

Authors: Berta Ferrer-Rosell; Esther Martinez-Garcia; Germà Coenders

Title: Package and no-frills air carriers as moderators of length of stay

Journal: *Tourism Management*

Status: Published

Doi: 10.1016/j.tourman.2013.11.002

Year, volume, pages: 2014, 42, 114-122

Journal quality index: I.F.: 2.571, Q1 (34th out of 174 in management; 3rd out of 35 in hospitality, leisure, sport & tourism)

Authors: Berta Ferrer-Rosell; Germà Coenders; Esther Martínez-Garcia

Title: Determinants in tourist expenditure composition – the role of airline type

Journal: *Tourism Economics*

Status: Accepted

Doi: not available yet

Year, volume, pages: not available yet

Journal quality index: I.F.: 0.800, Q3 (24th out of 35 in hospitality, leisure, sport & tourism)

Authors: Berta Ferrer-Rosell; Germà Coenders; Esther Martínez-Garcia

Title: Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes

Journal: *Current Issues in Tourism*

Status: Under review

Journal quality index: I.F.: 1.307, Q2 (9th out of 35 in hospitality, leisure, sport & tourism)

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

Abstract

This article analyses the determinants of length of stay among inbound tourists arriving by air in one of the world's most popular tourist countries, Spain. Special emphasis is placed on the effects of whether tourists booked the trip themselves or as part of a package and whether they travelled by low cost (LCA) or legacy airline. An ordered logit model is estimated. Relevant explanatory variables are related to tourist preferences and characteristics, trip characteristics, stay characteristics, and activities at destination. One of the main relevant results concerns the moderating effects. For instance, length of stay among package travellers is most affected by type of destination (city vs. coast), length of LCA trip by age, and length of legacy airline trip by accommodation type.

Keywords

Ferrer-Rosell, B., Coenders, G. and Martínez-Garcia, E. (accepted). Determinants in tourist expenditure composition – the role of airline type. *Tourism Economics.*

Determinants in tourist expenditure composition - the role of airline types

Berta Ferrer-Rosell[a*], Germà Coenders[b], Esther Martínez-Garcia[c]
[a,b,c] Department of Economics, University of Girona
[a, c] Faculty building of Tourism.
Edifici Sant Domènec. Plaça Ferrater Mora, 1, 17071 Girona - Spain
[b] Faculty building of Economics and Business.
Campus Montilivi, 17071 Girona – Spain
[a]berta.ferrer@udg.edu, [b]germa.coenders@udg.edu, [c]esther.martinez@udg.edu
*corresponding author: berta.ferrer@udg.edu
Phone: +34972419719 Fax: +34972418032

Abstract

The reduction in transportation costs when travelling with a low cost airline (LCA) seems to have modified the composition of the trip budget. An understanding of expenditure composition when comparing LCA and legacy airline travellers is vital for destination marketers. Using micro official statistics data for air travellers to Spain in 2010 and the compositional data analysis (CODA) methodology, this study analyses the determinants of trip budget composition and its differences between airline types. We consider transportation expenses, as well as basic (accommodation and food) and discretionary (activities, shopping, etc.) at-destination expenses. Log-ratios of budget share are fitted to a MANOVA, with traveller's attributes as explanatory factors along with the moderating effect of the airline type. Findings include that high income LCA travellers spend relatively more at-destination, LCA tourists travelling with friends have a larger share of discretionary expenses, and highly educated travellers have a larger share of discretionary expenses for both airline types.

Key words: compositional data analysis (CODA), tourist expenditure, low cost airline, tourism demand

Introduction

A growing tourist segment is the independent tourist. The independent tourist does not travel on a package deal and can organize the entire trip by him or herself. They also have easier access to information and usually prefer to undertake more activities. While 48% of all tourists to Spain in 2004 travelled without a package deal, this rose to 66% in 2010. The growth of the independent tourist is accompanied by an increasing expansion and consolidation of no-frills airlines, also referred to as low cost airlines (LCA). More LCA users than legacy airline passengers travel independently of the package tour. In 2010, 74% of all tourists who arrived in Spain with an LCA did not book a package deal, IET (2004; 2010).

LCAs' cheaper fares result in a significant reduction in transportation costs, which is expected to modify the composition of the trip's budget (Martínez-Garcia and Raya, 2008). This seems to have been the case in Spain, a major world tourist destination where in 2010 55.7% of all air travellers arrived by LCA and spent 32% of their trip budget on transportation and 54% on accommodation and food. In comparison, legacy passengers spent relatively more on transportation (40%) and somewhat less on accommodation and food (46%), IET (2010). The reduction in transportation costs can also affect the distribution of non-transportation expenditures, in other words, at-destination expenditures on accommodation, activities, and other budget components.

LCAs have received growing attention in the literature which either focuses on the demand differences between LCA and legacy airlines (Chiou and Chen, 2010; Ferrer-Rosell *et al*, 2014; Forgas *et al*, 2010; O'Connel and Williams, 2005), or on LCA demand and its heterogeneity (Castillo-Manzano and Marchena-Gómez, 2010; Martínez-Garcia *et al*, 2012; Kim and Lee, 2011; Martínez-Garcia and Royo-Vela, 2010; Raya-Vilchez and Martínez-Garcia, 2011). However, this growing literature on LCA demand and LCA compared to legacy airline demand has not, to our knowledge, focussed on tourist expenditure and its composition.

 Expenditure analyses have been quite frequent in the tourism literature, since it is of major concern for destination management offices (DMO), marketers, tourist agencies, and in general to all those engaged in tourism. Tourist expenditure, rather than numbers of tourists received, is becoming much more important for destinations and the economic impact of tourism. The analysis of expenditure composition provides valuable information for destination management, (over that given by the analysis of absolute expenditure), in terms of type of tourist according to how they distribute their travel budget. The travel budget can be broken down, as with household budget studies, between non-discretionary and discretionary components. The non-discretionary are those in accommodation, transportation and food, where a minimum amount has to be spent, whereas discretionary expenditure is that which includes extra activities, shopping, etc. Depending on personal, economic, trip and socio-demographic characteristics, tourists may be more or less willing to embark on visits, activities, excursions, shopping, etc. and thus the proportion of discretionary tourist budget will change accordingly. This distinction between the expenditure components is relevant for destinations as they are more interested in local spending, than in expenses paid directly to tour operators; as is the case for package trips and transportation expenses. DMOs are also interested in how the different tourist profiles within LCA and legacy users allocate their budget. For instance, if DMOs seek to promote activities, they

should focus their marketing efforts on those tourist types spending more on that component of their budget, and which may differ between airline types.

 The objective of this article is to study the determinants of the composition of tourist expenditure, in other words, the share of tourism expenditure allocated to the different categories of a trip budget, by taking into account tourist heterogeneity and distinguishing between legacy and LCA market segments. Two main research questions are addressed:

- How does travelling with an LCA vs. a legacy airline affect the distribution of trip expenses between transportation and other costs and the distribution of the non-transportation (or also called at-destination) expenses between discretionary and non-discretionary?

- Are passenger characteristics affecting budget composition? If so, do they affect it in the same way for both types of airline? In other words, does the airline type have a moderating effect?

With this purpose in mind, we build a statistical model explaining budget composition from passenger characteristics along with the moderating effect of airline type. The following budget parts are distinguished:

- Proportion or share of total trip expenditure devoted to transportation, where savings from low cost airlines arise (amount paid for transportation from the airport of origin to the point of accommodation, and the return trip)

- Share of total trip expenditure devoted to accommodation and food (basic, i.e., non-discretionary)

- Share of total trip expenditure devoted to doing activities, moving around at the destination and shopping (discretionary tourist expenditure).

Statistical analysis of budget compositions is a methodologically challenging task. Share in budgets, as any other composition, is expressed as proportions or percentages of a total, whose sum can only be 1 or 100. Compositional data lie in a restricted space and only convey information regarding the relative size of components to one another. The seminal work of Aitchison (1986) started a fruitful tradition in *compositional data analysis* (CODA) and of which the most widely used technique is the transformation of compositions. This is achieved by means of logarithms of ratios. Working with log-ratios not only has methodological implications but also substantive ones. Without log-ratios, components are estimated and interpreted separately from one another as if they could vary independently (*ceteris paribus*), which is impossible: the relative importance of one component (budget share) can only increase if the relative importance of at least one other decreases. Most methods used to model budget share such as, the *almost ideal demand system* (Deaton and Muellbauer, 1980) ignore, at least partly, the constraints and distributional nature of compositional data. To the best of our knowledge there is no scholarly study of the composition of tourist budget using an appropriate methodology for compositional data analysis.

This article is structured as follows. First, we present a review of the main approaches in modelling tourist expenditure, which is then followed by a description of the CODA

methodology. Then, we introduce the method and data, which is followed by the results, and finally by the overall conclusions and discussion.

Major approaches in modelling tourism expenditure

The study of expenditure in terms of composition (relative share of each part of the budget) is not the same as the study of expenditure in absolute terms and the variables affecting absolute expenditure may differ from those affecting relative expenditure. Tourism budgets have been approached from both perspectives in the literature. More precisely, tourism expenditure has been analysed globally, in absolute terms per budget parts, in relative terms (share) per budget parts and as a part in itself in family budgets.

The vast majority of microeconomic tourism demand studies (24 out of 27 in the review of Wang and Davidson, 2010) concern the prediction of one single aggregated expenditure variable. Methods range from mean comparison tests (Craggs and Schofield, 2009), OLS and WLS regression (Cannon and Ford, 2002; Downward and Lumsdon, 2000, 2003), to advanced econometric techniques. For instance, Hung *et al* (2012) use quantile regression in order to build equations predicting not only typical expenditures but also the highest and lowest. Alegre *et al* (2009), Eugenio-Martin (2003), Hong *et al* (1999), and Nicolau and Más (2005) propose double-hurdle, Heckit and related models to separate the decision whether to spend on tourism from the decision of how much to spend. Concerning the explanatory variables used, the most common are income, age, gender, marital status, education, place of residence, length of stay, travel group size and composition, accommodation, main trip purpose and activities (Marcussen, 2011). In general, explanatory variables can be grouped into economic variables (prices and income), socio-demographic variables, and trip or travel-related variables (Sainaghi, 2012; Wang *et al*, 2006). Nicolau (2009) includes individual price sensitivity estimated in a previous model.

Another stream of research is that which analyses tourist expenditure per tourism product (e.g. lodging, food, transportation and sightseeing/entertainment). A common argument for studying tourist expenditure patterns per tourism products is that it provides vital information to travel organizers and destination marketers when designing the appropriate marketing strategies. For this purpose, researchers have used several methods, such as Tobit, MANOVA, or seemingly unrelated linear regressions (Cai, 1998, 1999; Cai *et al*, 1995; Jang *et al*, 2004; Lee, 2001; Lehto *et al*, 2001; Oppermann, 1996; Pyo *et al*, 1991; Wang *et al*, 2006). Socio-demographic variables are the most common significant variables. Age affects expenditure related to meals, as does travel group, which also affects transportation expenditure. Marital status usually affects food and accommodation expenditure.

Since part expenditure in absolute terms is related to total expenditure, a common finding in these studies is that some of the explanatory variables affect all budget elements about equally. For example, Cai *et al* (1995) found that the higher the level of education, the higher the expenditure in all budget parts, and Wang *et al* (2006) conclude similarly about household income. Such results will never be obtained when analyzing budget share.

The empirical analysis of budget share, both for tourism expenditure and general family budgets, commonly implies estimating an almost ideal demand system of equations (Deaton and Muellbauer, 1980). The almost ideal demand system is designed to analyse the interdependences of budget allocations, thus overcoming the limitations of single-equation modelling, and has received much attention in the last decade (Song *et al*, 2012). It directly fits share (i.e. compositions) as dependent variables in a set of simultaneous regressions. The approach has been mainly focussed on estimating price and income elasticities. Studies which use macroeconomic data include amongst others, the ones by Divisekera (2007, 2009, 2010), Fujii *et al* (1985), O'Hagan and Harrison (1984), Syriopoulos and Sinclair (1993), and Wu *et al* (2011).

The almost ideal demand system had been applied both to data from a given origin to multiple destinations (i.e., ex ante, see Divisekera, 2009; Li *et al*, 2004) or from multiple origins to a given destination, which is the case in our article (i.e. ex post, see Divisekera, 2009). In ex post studies it is commonly assumed that various commodities can be aggregated to broad bundles of products, provided that prices in a bundle move in parallel, and that the utility function with respect to tourism and other goods is weekly separable. This makes it possible to conceptualize tourism consumption as a multi-stage process. In the first stage, tourists allocate a household budget part to tourism consumption, in the second a tourism budget part to each trip/destination, in the third a destination budget part to each good and service, including transportation, accommodation, etc. Ex post studies, like ours, only model the last stage. There are, however, some studies of how tourism competes against other categories of discretionary expenditure using individual micro data. For example, Melenberg and Van Soest (1996) use different parametric and semi-parametric Tobit models to explain the vacation budget share from household characteristics and Dolnicar *et al* (2008) analyse how households allocate discretionary income between tourism and competing uses.

The review of Wang and Davidson (2010) concludes that since the vast majority of tourism demand studies are conducted at macro level (this holds even more for budget share analysis), there is room for more micro-econometric studies in this area as the only manner of accounting for demand heterogeneity. The almost ideal demand system approach, when applied to micro data, can include individual characteristics. For example, Coenen and van Eekeren (2003) and Fleisher *et al* (2011) make ex-ante studies of individual budget share in Sweden and Israel respectively and use previous Heckit- type selection equations to model the decision whether to travel or not. Coenen and van Eekeren (2003) include household size and income as individual characteristics. Fleisher *et al* (2011) add age, education, real state ownership, place of birth and internet use. The basic aim of those articles is to estimate elasticities, and they use the individual characteristics only as controls. Fleisher *et al* (2011) are the only to provide the estimates of equations predicting budget share from traveller characteristics as a by-product of their elasticity estimates. Being born in the country of origin is reported to increase the transportation share and reduce the share of on-site expenditures. Education and household real estate assets reduce the share of on-site expenditures.

The aim and approach of our article are similar to those of Coenen and van Eekeren (2003) and of Fleisher *et al* (2011) regarding the use of individual characteristics to predict budget share, but differ in three important respects:

- Our study is ex-post, so that Heckit modelling is unfeasible.

- The effect of individual characteristics on budget composition (share) is the core of the analysis. Explanatory variables are individual characteristics rather than prices. The results will include the effects of traveller heterogeneity rather than demand equations.

- Our analysis takes into account the compositional restrictions of the data by using the CODA methodology.

The CODA methodology

Compared to absolute data, compositional data such as budget share lie in a constrained space. A *D*-term composition measured on individual *i* $x_{i1}, x_{i2},...,x_{iD}$ has the following constraints:

$$0 \leq x_{id} \leq 1 \text{ and } \sum_{d=1}^{D} x_{id} = 1 \tag{1}$$

Aitchison (1986) and Pawlowsky-Glahn and Buccianti (2011) warn against the serious problems that arise when using standard statistical analysis tools on compositional data. Compositional data are non-normal and heteroskedastic. One component can only increase if some other(s) decreases. This results in negative spurious correlations among the components and prevents interpreting effects of linear models in the usual way "keeping everything else constant".

Even if specialized CODA techniques are starting to appear (e.g. Ronning, 1992; Thió-Henestrosa and Martín-Fernández, 2005), the easy way (Aitchison, 1986; McLaren *et al*, 1995; Fry *et al*, 1996) involves transforming compositional data so that they can be subject to standard and well-understood statistical techniques. This is the approach we take in this article. In short, this implies using the transformed share by means of logarithms of ratios, instead of the raw share.

To ensure compositional coherence of predicted budget share (i.e. unit sum and non-negativity), a set of parameter constraints is imposed to the almost ideal demand system. However, the presence of an error term with an unbounded distribution (usually normal), results in a non-zero probability that actual share lies outside the [0,1] interval (McLaren *et al*, 1995; Fry, 2011; Fry *et al*, 1996). In other words, the in fact bounded distribution of budget share results in a misspecification of the almost ideal demand system and of any model fitting percentage share with an unbounded error distribution.

The fact that the error term in any proper system of demand equations applied to budget share should take into account the data compositional nature has been widely acknowledged (Aitchison, 1986; Fry *et al*, 1996; Ronning, 1992). However, the review in Fry (2011) reports few studies which have used the CODA methodology to demand equations and to the study of household budgets (Fry *et al*, 1996, 2000; McLaren *et al*, 1995).

When fitting demand equations to log-ratios of expenditure components Engel's aggregation condition, that for each household total expenditure should equal the sum of components, is automatically satisfied. The model can easily be extended to a full

consumer demand analysis by the introduction of prices and other covariates such as consumer characteristics (Aitchison, 1986). McLaren *et al* (1995) relate the CODA analysis with log-ratio transformation to the almost ideal demand system and conclude that CODA makes it possible to reach the same objectives with normal and homoskedastic error terms.

Related developments are the indirect addilog system (Houthakker, 1960) and the generalised addilog system (Bewley, 1982). When applied to compositions (e.g. Bewley and Fiebig, 1988) the latter is equivalent to the CODA methodology in which the log-ratios of each component over the geometric means of all components are the dependent variables in the set of simultaneous regressions.

To the best of our knowledge, there is no study of tourism budget share using the CODA methodology or any other methodology accounting for the compositional constraints in compositional data.

Method and data

Statistical approach

The simplest CODA approach involves applying standard statistical techniques on logarithms of ratios of components. Several log-ratio transformations have been suggested in the early CODA literature (Egozcue *et al*, 2003). The additive log-ratio transformation (alr) used by Fry *et al* (1996, 2000) is the most popular and the easiest to compute given that it is simply the log-ratio of each component to the last:

$$y_{id} = \ln(x_{id}/x_{iD}) = \ln(x_{id}) - \ln(x_{iD}) \text{ with } d=1,2,3,...,D\text{-}1 \qquad (2)$$

The centred log-ratio transformation (clr) used by the generalised addilog demand system (Bewley, 1982), computes the log-ratios of each component over the geometric mean of all the components, including itself.

$$y_{id} = \ln\left(\frac{x_{id}}{\sqrt[D]{x_{i1} x_{i2} x_{i3} \cdots x_{iD}}}\right) \text{ with } d=1,2,3,...,D \qquad (3)$$

All log-ratio transformed y_{id} variables recover the full unconstrained $-\infty$ to ∞ range. It must be noted that one dimension is lost in the alr while in the clr, one dimension is a linear combination of the remaining.

The alr is commonly used for statistical modelling and prediction of compositions (e.g. Fry *et al*, 1996). Conversely, the clr transformation is commonly used for statistical techniques which are based on a metric, such a cluster analysis, because of its preservation of distances, even though it leads to a singular covariance matrix. Thus, while, the alr would be appropriate for the purpose of this article, the fact that one component must be used as reference for all others reduces its flexibility and interpretability. Alternatives are presented at the end of this subsection.

While having zero expenditures in absolute data indeed has significant methodological consequences (e.g. Lee, 2001), these consequences are arguably more serious in the

CODA methodology. If the x_{id} variables contain zeros, then log-ratios cannot be computed. An obvious initial procedure to reduce zeros is to amalgamate small and conceptually similar components with many zeros into larger ones. In tourism budget research it can be useful to group together all expenditure on activities or all expenditure on food, for instance.

In certain instances, some zero components result from individual characteristics, which are called essential zeros in the CODA literature (Aitchison, 1986). Another typology of zeros encountered in the CODA literature is the rounding zero, that is, a component which is present but is too small to be detected by the measurement instrument. This is typical in chemical, biological and geological compositions, and the CODA literature offers ample instruments to deal with rounding zeros.

A classic essential zero example in economics is in household budget research when measuring expenditure on tobacco, and will essentially be zero if all members are non-smokers. If tobacco expenditure is really in the researchers' interest, the target population should be redefined to include only smokers. In many instances budget research is often not clear whether zero expenditures come closer to being essential or rounding zeros. In some cases, they can be understood as corner solutions in a utility maximization problem. In others they can be understood as the inherent randomness of human behaviour or as the limitations of the data. Tourists may spend a certain amount on activities on certain trips, but not on others and so surveys of only one trip will unavoidably contain some zeros of this type. Tourists may also forget or fail to report trivial expenses, like post-card shopping, local bus tickets, going to a museum, and the like (see Legohérel, 1998 for a discussion on the instability of tourism expenditure). Fry *et al* (2000) claim that in both situations zeros can be proxied by a very small value, and thus be treated as rounding zeros. In tourist budget research, to treat zero expenses in activities, as rounding zeros implies assuming firstly, that there is basically no tourist type who will never spend anything on activities, and secondly, that tourists who generally spend little on activities are basically similar to those who spend nothing or fail to report small expenses. We find both assumptions to be reasonable.

Fry *et al* (2000) essentially use the same zero replacement strategy that will later be suggested by Martín-Fernández *et al* (2003), namely replacing $x_{id}=0$ with:

$$x'_{id}=k\delta_{id} \text{ with } 0<k<1 \tag{4}$$

Where δ_{id} is the smallest detectable proportion for individual *i* and component *d*. Martín-Fernández *et al* (2003) suggest using $k=0.65$, although a sensitivity analysis of the results on the choice of *k* is always advisable (we used $k=0.30$, and $k=0.99$ with no sizeable change in the estimates). Next, non-zero x_{id} values have to be reduced in order to preserve the unit sum and the ratios among non-zero components. As suggested by Martín-Fernández *et al* (2003) with:

$$x'_{id} = x_{id}\left(1 - \sum_{x_{idm}=0} x'_{id}\right) \tag{5}$$

Simulations show this method performs particularly well if the proportion of zeros is below 10% (Martín-Fernández *et al*, 2011).

In our data set, zeros were present only in one budget category (discretionary expenditure). The minimum amount spent by the non-zero group was one euro, which roughly corresponds to the price of a city bus ticket, the entrance to a subsidized local museum, or a cheap souvenir. Since the total expenditure is known for each individual, we compute δ_{id} by dividing one euro with the total expenditure of individual I (see the appendix for the SPSS command syntax).

In this article we consider alternative log-ratio transformations which are more flexible than the alr and clr in that the denominator does not have to be the same in all ratios. This increased flexibility makes it easier to compute log-ratios which are more interpretable with respect to the researchers' questions or hypotheses.

In general, an interpretable log-ratio transformation is easy to compute whenever there is an interpretable sequential binary partition of components into pairs of groups of components, according to the researchers' objectives or to the conceptual similarity of the components. These partitions start by dividing components into two clusters and then continue by subdividing one of the clusters into two until each component constitutes its own cluster. D components always involve D-1 partitions. These partitions are best understood as a partition tree or dendrogram (Pawlowsky-Glahn and Egozcue, 2011).

A meaningful log-ratio transformation takes ratios of the geometric means of the two component clusters at each partition. Numerators and denominators are interchangeable.

In our article we consider x_{i1}=transportation expenditure, x_{i2}=accommodation and food (basic expenditure), x_{i3}=activities and shopping (discretionary expenditure). An interpretable sequential partition is shown in Figure 1.

Insert Figure 1.

The sequential partition in Figure 1 can have a dual interpretation. It may imply that researchers consider both types of at-destination expenses to be mutually similar and less similar to transportation. Or it may imply that the research questions involve the distribution of total expenditure between transportation and at-destination expenditure and the distribution of at-destination expenditure into basic and discretionary components, as is the case in our article.

The first log-ratio compares transportation expenditure with the geometric mean of accommodation and food (basic expenditure) and activities and shopping (discretionary expenditure). With this ratio, we want to observe how travelling with an LCA or a legacy company affects the share of transportation compared to non-transportation (at-destination) expenses.

$$y_{i1} = \ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right) = \ln(x_{i1}) - \frac{1}{2}\ln(x_{i2}) - \frac{1}{2}\ln(x_{i3})$$

(6)

Positive values show a transport share greater than the geometric mean of the remaining two components. Negative values show the opposite.

The second log-ratio is a ratio of accommodation and food (basic expenditure) over activities and shopping (discretionary expenditure). With this ratio we want to find out what tourists allocate the rest of their trip budget to and whether it is more on basic or discretionary expenditure, once they have paid for (often in advance) their transportation.

$$y_{i2} = \ln\left(\frac{x_{i2}}{x_{i3}}\right) = \ln(x_{i2}) - \ln(x_{i3})$$

(7)

Positive values show a basic (accommodation and food) share which is larger than the discretionary (activities and shopping) share. Negative values illustrate the opposite.

Log-ratio transformations based on sequential binary partitions are proportional to the isometric log-ratio transformation (ilr; see Egozcue *et al* 2003; Egozcue and Pawlowsky-Glahn, 2005).

As compositions are vector variables, they cannot be analysed component-wise, by means of univariate regression, ANOVA models and the like. Seemingly unrelated regression models for continuous explanatory variables or Multivariate ANOVA (i.e. MANOVA) models for categorical explanatory variables are appropriate. MANOVA's multivariate tests and statistics (e.g., Pillai's trace, Hotelling's trace or Wilk's Lambda) are invariant to how components are arranged in Figure 1 to compute log-ratios. Of course, univariate tests referring to each particular log-ratio are not invariant, hence the importance of the interpretability of each log-ratio. The software SPSS 19 is used to estimate the MANOVA model (GLM procedure).

We have to distinguish between three types of variables, moderating, exogenous and endogenous, to be included in the analysis.

Moderating variables are those which modify the effect of exogenous variables. In our article we treat type of airline as a moderator, in other words, we include its interaction terms with all other variables in the MANOVA model.

Exogenous variables are assumed to affect expenditure but not the other way around. Socio-demographic characteristics (e.g. traveller's age, gender, education, income, etc.) are obviously determined prior to any travel decision and hence exogenous with respect to expenditure.

On the other hand, many of the choices tourists make are interdependent (see Dellaert *et al*, 1997) or at least planned simultaneously (see Fesenmaier and Jeng, 2000). So, many trip attributes are arguably decided on at the same time as the expenditure, or at least it is not clear whether they are decided on consistently either before or after expenditure, by all travellers. We consider them to be endogenous variables and therefore do not include them in the MANOVA model as explanatory, in order to prevent endogeneity problems. Instead, we display them graphically in the log-ratio space. The means of both log-ratios within each endogenous category are the category coordinates.

The large sample size (see next subsection) makes it possible to use low p-values. Moderating effects with p-values higher than 0.01 according to either Pillai's trace, Hotelling's trace or Wilks' lambda, were removed from the model. All the variables and moderating effects included in the final model are significant at 0.01.

Sample and variables

In this article we use secondary official statistics data. The data were provided by the *Instituto de Estudios Turísticos* (IET) an official agency of the Ministry of Industry, Energy and Tourism and which produces the majority of tourism data in Spain. The survey is known as the *Encuesta de Gasto Turístico (EGATUR),* in which tourism expenditure and other tourist information such as trip information and tourist sociodemographic characteristics, is studied. The EGATUR survey, conducted in 27 major Spanish airports in 2010, used CAPI (Computer Assisted Personal Interview) to interview tourists leaving the country. The sample is non-proportionally stratified by country of residence, airport and month. See IET (2012) for further details on the EGATUR methodology.

Our universe is a subset of the EGATUR universe which consists of European leisure visitors arriving by air and spending at least one night in Spain. We excluded flights from outside Europe because LCAs mostly operate short-haul flights. We also centred our study on only those trips with one single destination, thus excluding multi-stage trips, as the decision process regarding expenditure composition for these trips is expected to fundamentally differ from that of single-stage trips. Stays of over 120 days were also excluded.

For this study, we did not consider:

- tourists who have essential zeros in accommodation (tourists who own a house at the destination or tourists who stay with friends or relatives)

- tourists who do not decide how much they spend on certain components (business and study trips)

- tourists who do not pay for the trip themselves (trips paid for by the company, family/friends, contests, offers, etc.)

- tourists for whom composition is wholly or partly unobserved (package tourists).

The final sample size was *n*=19,359.

From the expenditure variables included in the EGATUR survey data base and used in the model as budget components we firstly put the amount paid for transportation (x_1). This component has no zeros.

Secondly, the amount of money paid for accommodation and food is undistinguishable for full-board, half-board accommodation or bed & breakfast. Therefore we define a joint accommodation and food component (basic expenditure). In this component we include not only the amount paid for consumption in bars and restaurants, but also for

buying groceries and everyday products in supermarkets (x_2). This component has no zeros.

Finally, EGATUR provides an aggregated expenditure on activities and shopping (except groceries and everyday products). To this we added the conceptually similar amount paid for moving around the destination (public transportation and/or car rented at the destination) in order to build an activities and shopping component (x_3). This component had 9.8% zeros, which stresses its discretionary character. Zeros were replaced, as explained in the statistical approach subsection.

We consider as at-destination (or also called, non-transportation) expenses the basic component (food and accommodation) and the discretionary component (activities and shopping).

Share and log-ratios are described in Table 1.

Insert Table 1.

Explanatory variables are the level of education, income, country of residence, gender, travel group, professional status (for pensioners we consider their last professional position) and age. The age variable is built as a combination of the original age variable and the variable referring to the individual economic situation. This is done in order to have a specific pensioner category, taking into account the varying retirement ages across countries and professions, and thus capturing those tourists who have more free time to undertake a trip, regardless of physical age. Table 2 shows their categories and frequencies.

As endogenous variables, which are not included in the MANOVA model but are included in the log-ratio space, we include activities undertaken at the destination, accommodation, length of stay, total expenditure quartiles, quartiles of expenditure made at destination per day (basic plus discretionary expenses) and a combination of motivation and destination. This last variable has been built as a combination of the motivation and the destination variable. We distinguish between, firstly, those tourists travelling for cultural tourism to singular cities, secondly, those who come just for leisure, either in the countryside, or more commonly at the seaside, and, thirdly, other typologies.

Insert Table 2 and Table 3.

Results

Results for the exogenous variables

Insert Table 4.

Table 4 presents the MANOVA results. The univariate R-squared, corrected for the degrees of freedom, is 0.045 (first log-ratio) and 0.105 (second log-ratio). The multivariate uncorrected R-square is $1-\Lambda=0.178$ and corrected for the degrees of freedom is 0.176.

We checked that the results did not change when removing the 4 outliers with the highest Cook's distance.

Table 4 presents the estimates for the two log-ratios, the log-ratio of transportation over the other two categories (y_1) , and the log-ratio of basic expenditure (accommodation and food) over discretionary expenditure (activities and shopping) (y_2), respectively. Because of the moderating effects, within each log-ratio there are two columns, representing the effects when flying with an LCA or with a legacy airline (main and moderating effects have already added together for easier reading). If the LCA and Legacy columns are different there is a significant moderating effect (p-value < 0.01).

As we are interpreting log-ratios, a positive estimate of a given predictor category means that tourists in that category spend more on the numerator compared to the denominator, than tourists in the reference predictor category. A negative estimate shows the opposite.

The intercept term shows the main effect of company type, i.e. the predicted log-ratio for each type of airline within the reference category of all variables (university education, medium income, resident in UK or Ireland, 25-44 years old, male, travelling with partner and mid-level employee). Within the reference categories, legacy users spend a higher proportion on transportation compared to other expenses, and LCA users spend more on basic expenses compared to discretionary.

The level of education seems to have more to do with activities undertaken, than with transportation. Results show that a lower level of education results in higher expenses in basic expenses compared to discretionary, and is almost equal for users of both types of airline. For mainly or only LCA users, a lower level of education results in a higher share in transport compared to the other two categories. If we put it in another way, there seems to be a distinct highly educated LCA user segment, which utilizes the savings in transport to increase expenditure in non-transportation expenses, and even more so in discretionary.

Level of income, in contrast, affects both log-ratios. The higher the income, the lower the transport share. The higher the income, the lower the basic/discretionary expense ratio, that is, a higher portion of the non-transport expense is devoted to the discretionary budget.

As far as the country of residence is concerned, results show that it affects mainly the basic/discretionary log-ratio, and exhibits strong airline type moderating effects. LCA users tend to differ, depending on their country of residence, when distributing non-transport expenses into basic versus discretionary. In general, when compared to the UK and Ireland resident category (reference category) all other European residents have a larger discretionary share. More precisely, French, Italians and Portuguese have the highest discretionary share within LCA users, and other European countries (mainly Central-Eastern European) have the highest discretionary share within legacy airline users.

Gender has a small effect which is constant for both types of airline. Female travellers tend to slightly increase the transport share, and the share of basic expenses within the non-transport expenses.

The travel group seems to affect both log-ratios, and has a moderating effect with airline type. Those tourists who travel in a family group have a higher share in transport expenses when flying with legacy airlines. In the second log-ratio, travelling as a family with legacy airlines increases the basic expenses share compared to the share in discretionary. Travelling with friends increases the share in non-transport expenses for both airline types. Besides, those who travel with friends spend more in the discretionary share compared to basic expenses, and those who fly with LCA even more so. Finally, travelling alone increases the basic expenses share compared to the share in discretionary expenses for both airline types.

When professional status is considered, results show that it mostly affects the basic/discretionary log-ratio, and there are some relevant airline type moderating effects. Legacy users with a low professional status (homemaker, unemployed or low-level employee) spend more on transport compared to the non-transport expenses. Conversely, students using legacy airlines spend more on non-transport expenses, and these non-transport expenses are, to a greater extent, devoted to discretionary expenses compared to basic expenses. Homemakers increase basic expenditure compared to discretionary, with respect to the reference category (mid-level employee) for both airline types. LCA users who are self-employed also increase basic expenditure compared to discretionary. Low–level employees flying with an LCA, increase the discretionary share compared to the basic, and high-level employees flying with legacy airlines also spend relatively more on the discretionary part.

Results show that age affects the second log-ratio more than the first, and there is no moderating effect with the airline type. Being a pensioner results in a higher share in basic expenses compared to discretionary expenses for both airline types. The remaining categories have hardly any differences from the reference 25-44 years old.

Insert Figure 2.

As introduced in the methodology section, Figure 2 shows how the endogenous variables behave in the log-ratio space. The origin of the graph is represented by the mean log-ratios in Table 1, 0.1506 for the transport/other ratio and 1.7657 for the basic/discretionary ratio, and according to this the graph shows categories with higher or lower ratios than the average.

Regarding accommodation, it seems that rented apartments and other types of accommodation behave similarly, and have the lowest budget share in transport. In the case of hotels, as star category increases, expenditure on basic compared to discretionary increases, and expenditure on transport compared to at-destination decreases.

As far as the length of stay is concerned, longer stays increase the share of non-transport expenditure in basic expenses and reduce the share in transport.

As regards to destination and motivation, tourists going to a seaside or countryside destination for leisure purposes, behave quite differently from tourists going to an urban destination for urban or cultural tourism. The former group spends relatively more on transport and less on the discretionary portion. However, those coming for urban tourism show an increase in the share of non-transport expenses as well as in the discretionary share, compared to basic expenses.

The total expenditure variable gives relevant information. It is ordered along the first log-ratio axis, from top to bottom, meaning that those tourists having a lower total expenditure spend relatively more on the transport share, and those who have a higher total expenditure, spend relatively more on the other two parts of the budget. Regarding the expenditure made at destination per day, it is ordered diagonally, from the upper-right corner to the bottom-left corner, meaning that tourists having a higher at-destination expenditure per day, spend relatively less on transportation and at the same time, more on the discretionary part within at-destination expenses.

Finally, activities undertaken are located in the lower left quadrant of the graph, meaning that undertaking activities (discretionary expenses) decreases the share in transport expenses and, although they continue to spend more on basic expenses than on discretionary, the discretionary share increases. We have highlighted some of them. Those tourists attending sports and cultural events are those spending the most on discretionary compared to basic expenses. Hiking or some nautical sports could be free of charge, thus tourists doing these activities have a basic to discretionary ratio close to the mean. As for the transportation share, all activities tend to reduce this share about equally, with the exception of golf, which leads to a much more substantial reduction.

Conclusions and discussion

The main purpose of this article was to study the composition of tourist expenditure and its determinants, that is, the drivers of the share of tourist expenditure allocated to the different categories of a travel budget, and with special emphasis on the distinction between legacy versus LCA travellers.

The main differences between legacy and LCA users are that income, some countries of residence, the travel group and some of the occupation situations are differently affected depending on the type of airline. Tourists with a medium or low income who travel with legacy airlines tend to spend a greater share of their travel budget on transportation and basic expenses but those travelling on LCAs tend to spend relatively more at the destination, and more specifically on discretionary expenditure. LCA tourists residing in the "other European countries" spend relatively more on transportation. Families flying on legacy airlines spend more on the transportation share compared to the at-destination share, and also spend relatively more on basic expenses compared to discretionary ones. Finally, in terms of occupational status, legacy users with low-level employment spend more on transportation, whereas the opposite occurs with LCA travellers, who spend more on the discretionary component. Students and high-level employees flying with legacy airlines spend more on the discretionary part of the trip budget.

As far as implications for management are concerned, assuming that destination management offices (DMO's) objectives are to increase the at-destination component of the trip budget (i.e. lowering the first log-ratio), some of the recommended actions would be to increase marketing efforts directed at medium and high income earners, especially at those travelling on LCAs; to increase marketing strategies in some of the European outbound markets such as, Germany (especially those using LCA) and Italy and the Netherlands (especially those flying with legacy airlines). DMOs should also focus on those tourists travelling with friends (airline type is inconsequential) and students using legacy companies. As for the endogenous variables, DMOs should direct their efforts towards those tourists undertaking activities, towards those staying a little bit longer than a week and towards those with high total expenditure. Furthermore, if DMOs are also interested in capturing the groups who spend relatively more on discretionary expenses (i.e. lowering the second log-ratio), they should focus their marketing efforts on those tourists who are highly educated as well as those who have medium to high incomes and who use LCAs. Additionally, they should focus on other markets than UK and Ireland. Young tourists and tourists who travel with friends (where airline type used does not matter), low-level employees and students using LCAs as well as students and high-level employees flying on legacy airlines, should also be targeted. In relation to the endogenous variables, DMOs must take into account those tourists undertaking payable activities, those on shorter stays or coming for urban and cultural tourism or other purposes.

The appeal of the CODA methodology for studying tourism budgets lies in the fact that, once the variables have been transformed, the researcher can use standard and well understood statistical models with unbounded error term distributions, while ensuring that the results will be compositionally coherent and that the standard statistical assumptions will hold. Conversely, results are not compositionally coherent when we use the MANOVA model on raw share without transforming it into log-ratios. We have seen with our data that 84% of individuals included in the sample have at least one limit of the 90% share prediction intervals outside the [0,1] range.

Contrary to the classic and highly restrictive alr and clr transformations, the CODA methodology offers the potential to construct tailor-made log-ratios which are intuitive to interpret and suit the research questions at hand. A classification tree of components is a clear and useful tool in this respect. We encourage researchers to use this approach in further research on tourism expenditure when budget share division is fundamental to the researchers' questions.

In this study, we have encountered mainly two limitations. Firstly, contrary to what the case is in studies using macro data, in micro data studies the coexistence of full-board, half board, bed & breakfast and accommodation-only tourists makes it impossible to meaningfully separate accommodation and food expenditure. Other subdivisions (e.g. discretionary expenditure on activities, non-grocery shopping and moving around the destination) are technically feasible but can only increase the percentages of zeros. The CODA methodology involves some degree of amalgamation of components. Secondly, even though there are some advantages to using a database from an official statistics institution, as in this case (large sample size and scope), the main disadvantage is that the set of available variables cannot be controlled by the researcher. A further issue to be addressed is the time dimension. As the EGATUR survey is conducted annually, further research can be done to include a repeat cross-section analysis in order to capture trends in the effects of predictors.

Appendix. SPSS syntax code used to replace zeros

```
***** x1 x2 x3 contain raw share
***** x1_prime x2_prime x3_prime contain zero-replaced share
***** total_expenditure contains absolute expenditure
***** delta is 1/total_expenditure
***** the zero replacement constant is k=0.65

compute x1_prime = x1.
compute x2_prime = x2.
compute x3_prime = x3.
if (x3 = 0) x3_replacement = 0.65*1/total_expenditure .
if (x3 = 0) x3_prime = sum(x3, x3_replacement).
if (X3 = 0) x2_prime = x2*(1-x3_replacement).
if (X3 = 0) x1_prime = x1*(1-x3_replacement).
execute.
```

References

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability,* Chapman and Hall, London. (Reprinted in 2003 with additional material by The Blackburn Press).

Alegre, J., Mateo, S., and Pou, LL. (2009), 'Participation in tourism consumption and the intensity of participation: an analysis of their socio-demographic and economic determinants', *Tourism Economics*, Vol 15, No 3, pp 531–546.

Bewley, R. (1982), 'The generalised addilog demand system applied to Australian time series and cross section data', *Australian Economic Papers*, Vol 21, No 38, pp 177–192.

Bewley, R., and Fiebig, D.G. (1988), 'Estimation of price elasticities for an international telephone demand model', *The Journal of Industrial Economics*, Vol 36, No 4, pp 393–409.

Cai, L.A. (1998), 'Analyzing household food expenditure patterns on trips and vacations: a tobit mode', *Journal of Hospitality and Tourism Research*, Vol 22, No 4, pp 338–358.

Cai, L.A. (1999), 'Relationship of household characteristics and lodging expenditure on leisure trips', *Journal of Hospitality and Leisure Marketing*, Vol 6, No 2, pp 5–18.

Cai, L.A., Hong, G-S., and Morrison, A.M. (1995), 'Household expenditure patterns for tourism products and services', *Journal of Travel and Tourism Marketing*, Vol 4, No 4, pp 15–39.

Cannon, T.F., and Ford, J. (2002), 'Relationship of demographic and trip characteristics to visitor spending: an analysis of sports travel visitors across time', *Tourism Economics*, Vol 8, No3, pp 263–271.

Castillo-Manzano, J.L., and Marchena-Gómez, M. (2010), 'Analysis of determinants of airline choice: profiling the LCC passenger', *Applied Economics Letters*, Vol 18, No 1, pp 49–53.

Chiou, Y-C., and Chen Y-H. (2010), 'Factors influencing the intentions of passengers regarding full service and low cost carriers: a note', *Journal of Air Transport Management*, Vol 16, No 4, pp 226–228.

Coenen, M., and van Eekeren, L. (2003), 'A study of the demand for domestic tourism by Swedish households using a two-staged budgeting model', *Scandinavian Journal of Hospitality and Tourism*, Vol 3, No 2, pp 114–133.

Craggs, R., and Schofield, P. (2009), 'Expenditure-based segmentation and visitor profiling at the Quays in Salford, UK', *Tourism Economics*, Vol 15, No 1, pp 243–260.

Deaton, A., and Muellbauer, J. (1980), 'An almost ideal demand system', *American Economic Review*, Vol 70, No 3, pp 312–326.

Dellaert, B., Borgers, A., and Timmermans, H. (1996), 'Consumer activity pattern choice: Development and test of stage-dependent conjoint choice experiments. *Journal of Retailing and Consumer Services*, Vol 4, No 1, pp 25–37.

Divisekera, S. (2007), *Modelling and Estimation of Tourism Demand Elasticities: A Study of Tourism Expenditure Allocation in Australia*, Cooperative Research Centre for Sustainable Tourism (CRCST), Gold Coast.

Divisekera, S. (2009), 'Ex Post demand for Australian tourism goods and services', *Tourism Economics*, Vol 15, No 1, pp 153–180.

Divisekera, S. (2010), 'Economics of leisure and non-leisure tourist demand: a study of domestic demand for Australian tourism', *Tourism Economics*, Vol 16, No 1, pp 117–136.

Dolnicar, S., Crouch, G.I., Devinney, T., Huybers, T., Louviere, J.J., and Oppewal, H. (2008), 'Tourism and discretionary income allocation. Heterogeneity among households', *Tourism Management*, Vol 29, No 1, pp 44–52.

Downwoard, P., and Lumsdon, L. (2000), 'The demand for day-visits: an analysis of visitor spending', *Tourism Economics*, Vol 6, No 3, pp 251–261.

Downward, P., and Lumsdon, L. (2003), 'Beyond the demand for day-visits: an analysis of visitor spending', *Tourism Economics*, Vol 9, No 1, pp 67–76.

Egozcue, J.J., and Pawlowsky-Glahn, V. (2005), 'Groups of parts and their balances in compositional data analysis', *Mathematical Geology*, Vol 37, No 7, pp 795–828.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003), 'Isometric logratio transformations for compositional data analysis', *Mathematical Geology*, Vol 35, No 3, pp 279–300.

Eugenio-Martin, J.L. (2003), 'Modelling determinants of tourism demand as a five-stage process: A discrete choice methodological approach', *Tourism and Hospitality Research*, Vol 3, pp 341–354.

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

Fesenmaier, D.R., and Jeng, J. (2000), 'Assessing structure in the pleasure trip planning process', *Tourism Analysis*, Vol 5, No 1, pp 13–27.

Fleischer, A., Peleg, G., and Rivlin, J. (2011), 'The impact of changes in household vacation expenditures on the travel and hospitality industries', *Tourism Management*, Vol 32, No 4, pp 815–821.

Forgas, S., Moliner, M.A., Sánchez, J., and Palau, R. (2010), 'Antecedents of airline passenger loyalty: low-cost versus traditional airlines', *Journal of Air Transport Management*, Vol 16, No 4, pp 229–233.

Fry, T. (2011), 'Applications in economics', in Pawlowsky-Glahn, V., and Buccianti, A., eds, *Compositional Data Analysis. Theory and Applications*, Wiley, New York, pp 318–326.

Fry, J.M., Fry, T.R.L., and McLaren, K.R. (1996), 'The stochastic specification of demand share equations: restricting budget share to the unit simplex', *Journal of Econometrics*, Vol 73, No 2, pp 377–385.

Fry, J.M., Fry, T.R.L., and McLaren, K.R. (2000), 'Compositional data analysis and zeros in micro-data', *Applied Economics*, Vol 32, No 8, pp 953–959.

Fujii, E.T., Khaled, M., and Mak, J. (1985), 'An almost ideal demand system for visitor expenditures', *Journal of Transport Economics and Policy*, Vol 19, pp 161–171.

Hong, G-S., Abdel-Ghany, M., and Kim, S.Y. (1999), 'Modeling tourism spending decisions as a two-step process', *European Advances in Consumer Research*, Vol 4, pp 216–223.

Houthakker, H.S. (1960), 'Additive preferences', *Econometrica*, Vol 28, No 2, pp 244–256.

Hung, W-T., Shang, J-K., and Wang, F-C. (2012), 'Another look at the determinants of tourism expenditure', *Annals of Tourism Research*, Vol 39, No 1, pp 495–498.

IET (2004), *Compañías Aéreas de Bajo Coste. Informe Anual 2004*, Instituto de Estudios Turísticos, Ministerio de Industria, Turismo y Comercio, Madrid. http://www.iet.tourspain.es/es—ES/estadisticas/otrasestadisticas/companiabajocoste/anuales/Informe%20Compa%C3%B1%C3%ADas%20Bajo%20Coste%202004.pdf [Downloaded 14.1.2013].

IET (2010), *Compañías Aéreas de Bajo Coste. Informe Anual 2010*, Instituto de Estudios Turísticos, Ministerio de Industria, Turismo y Comercio, Madrid. http://www.iet.tourspain.es/es—ES/estadisticas/otrasestadisticas/companiabajocoste/anuales/Informe%20Compa%C3%B1%C3%ADas%20Bajo%20Coste%202010.pdf [Downloaded 14.1.2013].

IET (2012), *Egatur. Encuesta de gasto turístico. Metodologia*, Instituto de Estudios Turísticos, Ministerio de Industria, Energía y Turismo, Madrid. www.iet.tourspain.es/es−ES/estadisticas/egatur/metodologia/Referencia%20Metodolgica/Nota%20Metodol%C3%B3gica%20Encuesta%20de%20Gasto%20Tur%C3%ADstico.pdf [downloaded 22.11.2013].

Jang, S.C.S., Bai, B., Hong, G.S., and O'Leary, J.T. (2004), 'Understanding travel expenditure patterns: A study of Japanese pleasure travellers to the United States by income level', *Tourism Management*, Vol 25, No 3, pp 331–341.

Kim, Y.K., and Lee, H.R. (2011), 'Customer satisfaction using low cost carriers', *Tourism Management*, Vol 32, No 2, pp 235–243.

Lee, C. (2001), 'Determinants of recreational boater expenditures on trips', *Tourism Management*, Vol 22, No 6, pp 659–667.

Legohérel, P. (1998), 'Toward a market segmentation of the tourism trade: expenditure levels and consumer behaviour instability', *Journal of Travel and Tourism Marketing*, Vol 7, No 3, pp 19–39.

Lehto, X.Y., Morrison, A.M., and O'Leary, J.T. (2001), 'Does the visiting friends and relatives' typology make a difference? A study of the international VFR market to the United States', *Journal of Travel Research*, Vol 40, No 2, pp 201–212.

Li, G., Song, H., and Witt, S.F. (2004), 'Modelling tourism demand: a dynamic linear AIDS approach', *Journal of Travel Research*, Vol 43, No 2, pp 141–150.

Marcussen, C.H. (2011), 'Determinants of tourist spending in cross-sectional studies and at Danish destinations', *Tourism Economics*, vol 17, No 4, pp 833–855.

Martín-Fernández, J.A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003), 'Dealing with zeros and missing values in compositional data sets using non-parametric imputation', *Mathematical Geology*, Vol 35, No 3, pp 253–278.

Martín-Fernández, J.A., Palarea-Albaladejo, J., and Olea, R.A. (2011), 'Dealing with zeros', in Pawlowsky-Glahn, V., and Buccianti, A., eds, *Compositional Data Analysis. Theory and Applications*, Wiley, New York, pp 47–62.

Martínez-Garcia, E., Ferrer-Rosell, B., and Coenders, G. (2012), 'Profile of business and leisure travelers on low cost carriers in Europe', *Journal of Air Transport Management*, Vol 20, pp 12–14.

Martínez-Garcia, E., and Raya, J.M. (2008), 'Length of stay for low-cost tourism', *Tourism Management*, Vol 29, No 6, pp 1064–1075.

Martínez-Garcia, E., and Royo-Vela, M. (2010), 'Segmentation of low-cost flights users at secondary airports', *Journal of Air Transport Management*, Vol 16, pp 234–237.

McLaren K., Fry J., and Fry, T. (1995), 'A simple nested test of the almost ideal demand system', *Empirical economics*, Vol 20, No 1, pp 149–161.

Melenberg, B., and Van Soest, A. (1996), 'Parametric and semi-parametric modelling of Vacation expenditures', *Journal of Applied Econometrics*, Vol 11, No 1, pp 59–76.

Nicolau, J.L. (2009), 'The smile of the tourist: the relationship between price sensitivity and expenses', *The Service Industries Journal*, Vol 29, No 8, pp 1125–1134.

Nicolau, J.L., and Más, F.J. (2005), 'Heckit modelling of tourist expenditure: evidence from Spain', *International Journal of Service Industry Management*, Vol 16, No 3, pp 271–293.

O'Connell, J.F., and Williams, G. (2005), 'Passengers' perception of low cost airlines and full service carriers: a case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines', *Journal of Air Transport Management*, Vol 11, pp 259–272.

O'Hagan, J.W., and Harrison, M.J. (1984), 'Market share of US tourist expenditure in Europe: an econometric analysis', *Applied Economics*, Vol 16, pp 919–931.

Oppermann, M. (1996), 'Visitation of tourism attractions and tourist expenditure patterns - repeat versus first-', *Asia Pacific Journal of Tourism Research*, Vol 1, No 1, pp 61–68.

Pawlowsky-Glahn, V., and Buccianti, A. (2011), *Compositional Data Analysis. Theory and Applications*, Wiley, New York.

Pawlowsky-Glahn, V., and Egozcue, J.J. (2011), 'Exploring compositional data with the CoDa-dendrogram', *Austrian Journal of Statistics*, Vol 40, No 1, pp 103–113.

Pyo, S.S., Uysal, M., and McLellan, R.W. (1991), 'A linear expenditure model for tourism demand', *Annals of Tourism Research*, Vol 18, No 3, pp 443–454.

Raya-Vilchez, J.M., and Martínez-Garcia, E. (2011), 'Nationality and low-cost trip duration. A microeconometric analysis', *Journal of Air Transport Management*, Vol 17, pp 168–174.

Ronning, G. (1992), 'Share equations in econometrics: a story of repression, frustration and dead ends', *Statistical Papers*, Vol 33, No 1, pp 307–334.

Sainaghi, R. (2012), 'Tourist expenditures: the state of the art', *Anatolia: An International Journal of Tourism and Hospitality Research*, Vol 23, No 2, pp 217–233.

Song H., Dwyer, L., and ZhengCao, G.L. (2012), 'Tourism economics research: a review and assessment', *Annals of Tourism Research*, Vol 39, No 3, pp 1653–1682.

Syriopoulos, T.C., and Sinclair, T. (1993), 'An econometric study of tourism demand: the AIDS model of US and European tourism in Mediterranean countries', *Applied Economics*, Vol 25, No 12, pp 1541–1552.

Thió-Henestrosa, S., and Martín-Fernández, J.A. (2005), 'Dealing with Compositional Data: The Freeware CoDaPack', *Mathematical Geology*, Vol 37, No 7, pp 773–793.

Wang, Y., and Davidson, M.C.G. (2010), 'A review of micro-analyses of tourist expenditure', *Current Issues in Tourism*, Vol 13, No 6, pp 507–524.

Wang, Y., Rompf, P., Severt, D., and Peerapatdit, N. (2006), 'Examining and identifying the determinants of travel expenditure patterns', *International Journal of Tourism Research*, Vol 8, No 5, pp 333–346.

Wu, D.C., Li, G., and Song, H. (2011), 'Analyzing tourism consumption: a dynamic system-of-equations approach', *Journal of Travel Research*, Vol 50, No 1, pp 46–56.

Tables and Figures

Table 1. Percent share and log-ratio descriptive statistics

	Min.	Max.	Mean	St. Dev.	Skewness	Kurtosis
Transportation component (x_1)	.03	96.11	28.46	13.87	.837	.789
Basic component (x_2)	.21	99.63	54.72	16.13	-.114	-.450
Discretionary component (x_3)	.01	86.68	16.81	12.92	.988	1.174
Transportation/at-destination log-ratio (y_1)	-7.57	5.48	.1506	1.0702	.714	1.404
Basic/discretionary log-ratio (y_2)	-5.30	8.72	1.7657	1.8173	1.573	2.151

Table 2. Frequency distributions of exogenous and moderating variables in the MANOVA model

Variables	Count	Percent
Level of education		
Up to high school	6019	31.1
University[a]	13340	68.9
Income category		
Medium/Low	797	4.1
Medium[a]	13186	68.1
Medium/high	4376	22.6
High	1000	5.2
Country of residence		
Portugal	624	3.2
Other European Countries	850	4.4
Belgium	1006	5.2
Netherlands	1069	5.5
Austria, Switz. and Liech.	1122	5.8
France	1326	6.8
Scandinavian Countries	1789	9.2
Germany	2363	12.2
Italy	3113	18.1
United Kingdom and Ireland[a]	6097	31.5
Gender		
Female	8530	44.1
Male[a]	10829	55.9
Travel group		
Alone	2484	12.8
In family	2980	15.4
With friends	4440	22.9
With partner[a]	9446	48.8
Professional status		
Homemaker	517	2.7
Unemployed	602	3.1
Student	1432	7.4
Low-level employee	915	4.7
Mid-level employee[a]	10964	58.6
High-level employee	2138	11.0
Self-employed	2791	14.4
Age		
Over 45 years old and pensioner	1484	7.7
Over 45 years old and not pensioner	5009	25.9
25-44[a]	10569	54.6
15-24	2297	11.9
Type of airline (moderating variable)		
Legacy	5884	30.4
Low cost	13475	69.6

[a] Reference categories in the MANOVA model: chosen either because they are the largest categories, or because they are the most standard, conceptually considered.

Table 3. Frequency distributions of endogenous variables (not included in the MANOVA model)

Variables	Count	Percent
Activities		
Golf		
Yes	227	1.2
No	19132	98.8
Hiking		
Yes	364	1.9
No	18995	98.1
Sporting events		
Yes	603	3.1
No	18756	96.9
Nautical sports		
Yes	1645	8.5
No	17714	91.5
Other sports		
Yes	904	4.7
No	18455	95.3
Cultural visits		
Yes	11180	57.8
No	8179	42.2
Cultural events		
Yes	2452	12.7
No	16907	87.3
Other cultural activities		
Yes	4175	21.6
No	15184	78.4
Spa		
Yes	1199	6.2
No	18160	93.8
Theme Parks		
Yes	1745	9.0
No	17614	91.0
Gastronomy		
Yes	1565	8.1
No	17794	91.9
Type of accommodation		
Hotel 4-5*	5056	26.1
Hotel 3*	4922	25.4
Hotel <3*	6673	34.5
Rented appartment	2116	10.9
Other accommodation	592	3.1
Length of stay		
5 or less nights	9911	51.2
6-8 nights	5952	30.7
9-12 nights	1592	8.2
13-15 nights	1261	6.5

16 and more nights	643	3.3
Total expenditure		
Quartile 1	5074	26.2
Quartile 2	5104	26.4
Quartile 3	4986	25.8
Quartile 4	4195	21.7
Expenditure at destination per day		
Quartile 1	4855	25.1
Quartile 2	4832	25.0
Quartile 3	4868	25.1
Quartile 4	4804	24.8
Motivation-Destination		
Leisure on coast /in country side	8083	41.8
Urban tourism	5324	27.5
Others	5952	30.7

Table 4. MANOVA results.

	y_1: log-ratio transportation/ at-destination		y_2: log-ratio basic/ discretionary	
	LCA	Legacy	LCA	Legacy
Intercept	0.225	0.274	2.195	1.881
Up to high school	0.192	0.078	0.406	0.437
University	0	0	0	0
Medium/low income	0.073	0.289	0.037	0.518
Medium income	0	0	0	0
Medium/high income	-0.227	-0.058	-.0269	-0.006
High income	-0.495	-0.190	-0.463	-0.141
Residence in Portugal	0.041	0.069	-1.136	-0.580
Residence in other European Countr.	0.326	-0.012	-0.772	-0.820
Residence in Belgium	-0.127	-0.159	-0.667	-0.461
Residence in Netherlands	-0.120	-0.197	-0.678	-0.530
Residence in Austria+Swiz+Liench.	-0.017	-0.087	-0.687	-0.436
Residence in France	-0.112	-0.111	-0.870	-0.553
Residence in Scandinavian countries	-0.045	-0.038	-0.660	-0.477
Residence in Germany	-0.283	-0.176	-0.452	-0.392
Residence in Italy	-0.209	-0.266	-0.855	-0.503
Residence in UK or Ireland	0	0	0	0
Female	0.094	0.094	0.093	0.093
Male	0	0	0	0
Travelling alone	-0.053	-0.002	0.081	0.126
Travelling in family	0.085	0.193	0.014	0.213
Travelling with friends	-0.237	-0.191	-0.504	-0.282
Travelling with partner	0	0	0	0
Homemaker	0.080	0.160	0.253	0.235
Unemployed	0.030	0.159	-0.036	-0.039
Student	-0.005	-0.104	-0.331	-0.624
Low-level employee	-0.034	0.194	-0.194	0.063
Mid-level employee	0	0	0	0
High-level employee	-0.008	-0.064	-0.001	-0.139
Self-employed	0.029	-0.086	0.193	-0.072
Over 45 years old and pensioner	-0.060	-0.060	0.633	0.633
Over 45 years old and not pensioner	-0.010	-0.010	0.142	0.142
25-44 years old	0	0	0	0
15-24 years old	0.002	0.002	-0.019	-0.019
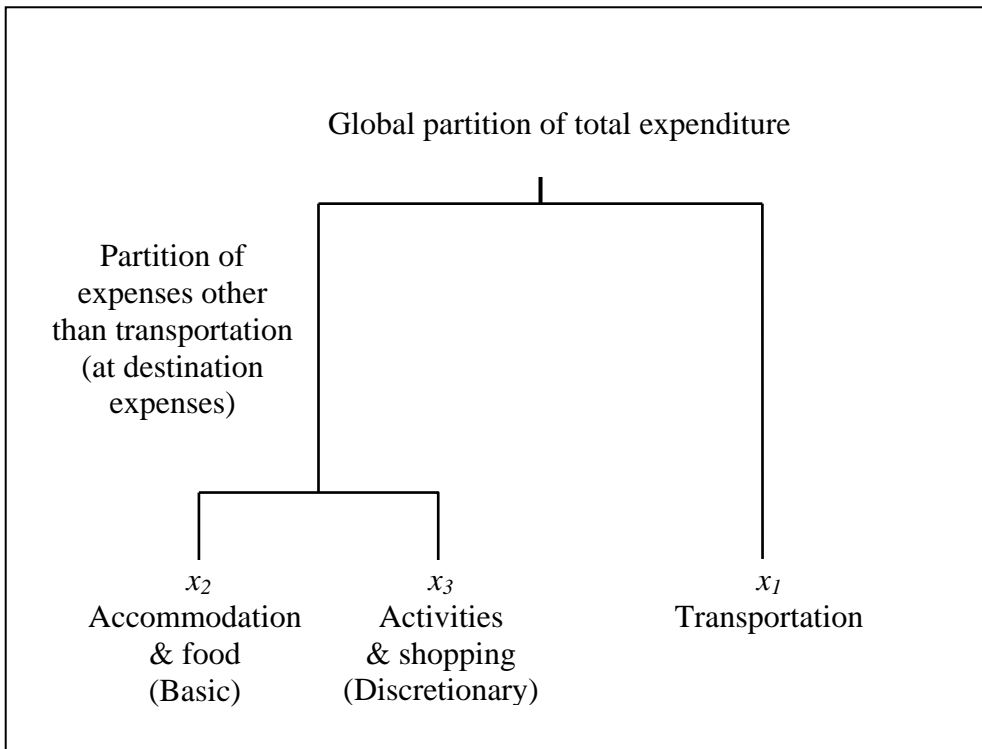
Figure 1. Sequential partition of total expenditure
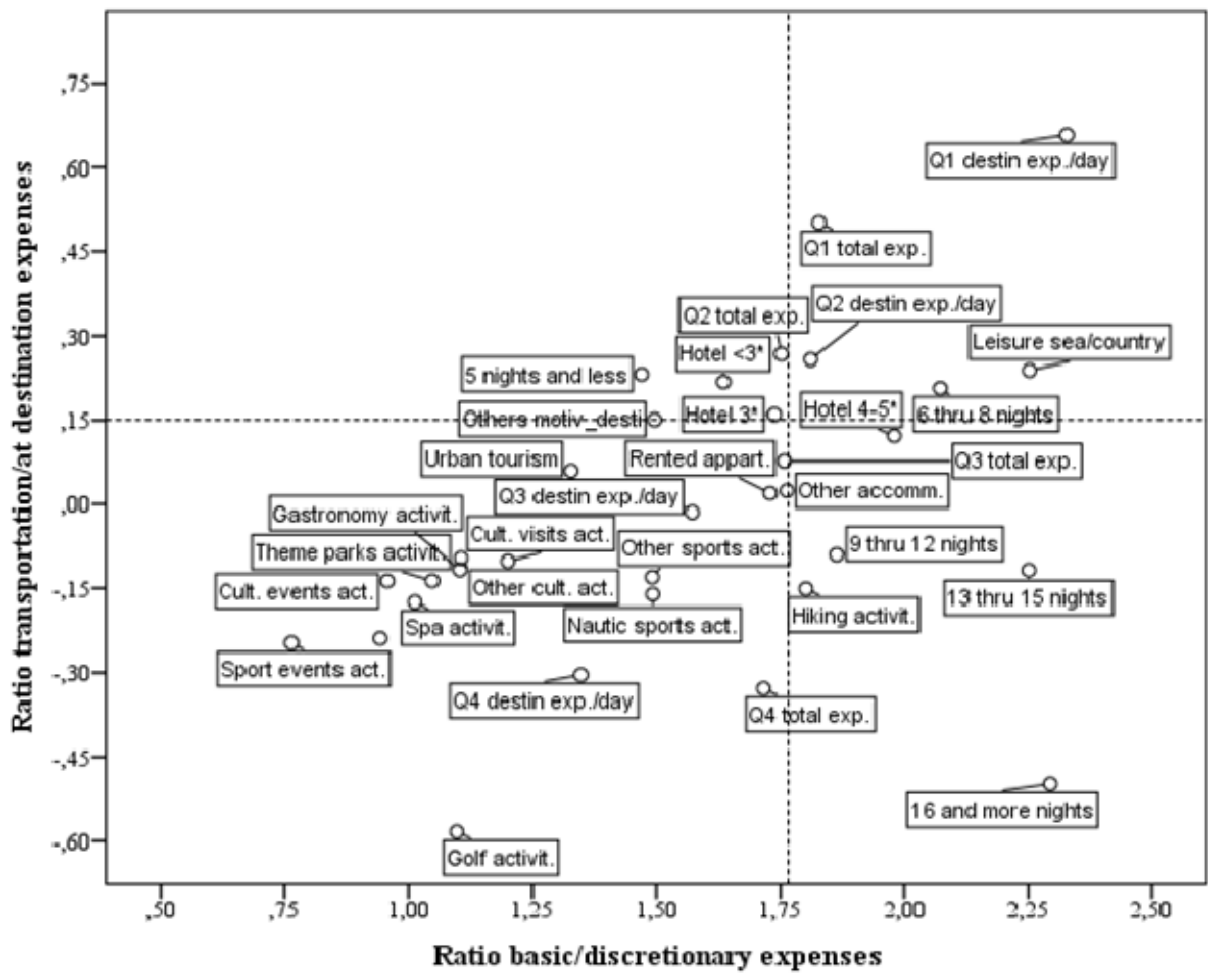
Figure 2. Endogenous variables on the log-ratio space. Mean log-ratio values within each category of the endogenous variables

Ferrer-Rosell, B., Coenders, G. and Martínez-Garcia, .E. (under review). Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes. *Current Issues in Tourism*.

Ferrer-Rosell, B., Coenders, G. and Martínez-Garcia, .E. "Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes". *Current Issues in Tourism* [under review]

Abstract

In this article, we propose a new method for segmenting tourists according to budget share (in other words, the proportions of total expenditure devoted to different items, such as transportation, accommodation, activities undertaken, and so on). Share allocated to budget parts lies in a restricted sample space (is positive and has constant sum), thus requiring compositional data analysis (CODA) tools. Our proposal bridges the traditions of CODA -new to tourism research- and mixture models (latent class models) which, by contrast with traditional methods (e.g. cluster analysis) make it possible to carry out statistical inference with regard to segment characteristics and their relationships with external variables, such as tourist background. The way we use the CODA and latent class methodologies is intuitively appealing and boils down to building transformed variables which can be tailored to the specific research questions. Among the available method variants and statistical tools we handpick the most appropriate for tourism budget segmentation purposes. We apply the suggested method combination to analyze the spending profiles of tourists arriving in Spain on low-cost airlines, with relevant and interpretable results.

Keywords

segmentation; compositional data analysis (CODA); latent class models; tourist expenditure; budget share; low cost tourism

5 SUMMARY AND DISCUSSION OF THE MAIN RESULTS

In this thesis, two of the key aspects of tourism demand have been analysed; length of stay and travel budget. As seen in the literature review, there were some significant gaps regarding these two main variables, especially when relating them to previous research on airline and Spanish tourism demand. So, the objectives of the present dissertation, which in our opinion have been accomplished, were firstly to compare legacy airline and LCA users in terms of the length of stay by using the variable how the trip was booked (by package, or by the tourists themselves deciding on the type of airline) as a moderating effect; secondly, to study how airline tourists allocate their trip budget to the different trip budget categories (transportation, and non-discretionary and discretionary expenses at destination); and finally, to develop a sound and easy-to-use method to segment tourists based on travel budget share. In this case we also use the intended LCA users' segmentation in Spain as an illustration of the new method.

Significant contributions to the length of stay and tourist expenditure research have been made in terms of substantive results, management implications and methodology.

We would like to start with the most readily applicable contributions accomplished by the present thesis; namely the implications for tourism management. It is vital for destination marketers, Destinations Management Offices (DMO) and the destination's tourism industry to know the profiles of their visitors in order to generate products that meet their most valuable tourists' preferences, make the highest possible profit from them in terms of the economic impact of their consumption at destination, and generate actions to attract them. As has already been pointed out in the articles, DMOs have the objective of lengthening the stay of their tourists in order to also increase their expenditure, and at the same time, destinations and local tourism firms are more interested in how tourists distribute at-destination spending than in spending on transportation. Besides, segmenting the majority of tourists who arrive at a destination (56% of airline tourists used LCA in 2010) makes it possible to carefully depict the characteristics of the heterogeneous tourists groups, so that tourism promoters, DMOs and the industry may contemplate specific effective actions for the differing segments.

Thus, assuming that DMOs are interested in lengthening the stay of tourists and in increasing the expenses at-destination, the results of the three articles show that they should focus their marketing efforts on those tourists travelling with LCAs who undertake activities; those from European countries other than the UK and Ireland, i.e. Scandinavians, Germans (especially using LCA), Italians and visitors from the Benelux countries (especially those flying with legacy airlines); and they should also focus on tourists coming for urban-cultural tourism purposes and mostly travelling with LCAs. Overall, DMOs should not be any less interested in LCA travellers than in legacy airline travellers.

Amongst LCA users, the third article has identified relevant distinct consumption profiles of tourists according to their expenditure composition, which can constitute valuable marketing targets, and which are significantly related to trip and traveller characteristics.

We found six segments; the largest segment includes 76.8% of the sample and is thus close to the average profile.

Two segments together account for 11% of the sample and include mostly tourists with zero discretionary components (activities, shopping and moving around at the destination). Most trips involve no activities and are taken for leisure purposes. Typical predictor profiles include residents of the UK and Ireland, self-employed, female, with a lower level of studies, travelling with their partner, alone or with the family, and low income earners.

Another segment (7.6%) includes those tourists who spend the least in the discretionary component after the former two. This is related to residents of the UK, Ireland, Germany and Scandinavia, as well as to pensioners.

The smallest segment (less than 1% of the sample) is a valuable target despite its mere size as it includes the relatively heavier spenders in discretionary components such as activities done at destination. Being a pensioner was a significant predictor. Associated trip characteristics include the longest stays, the highest overall expenditures and the highest propensity to undertake activities.

The final segment (3.4%) includes tourists who spend a relatively high amount on transportation, and relatively more on discretionary than on basic expenses (accommodation and food). It is related to tourists from countries other than the UK, Ireland, Benelux and Germany, and to tourists who undertake urban trips, who embark on cultural activities and who have the shortest stays and lowest total expenditure. It seems to us that they respond to the so-called "city-break" travellers.

Globally, we can say that although most LCA tourists are grouped into one market segment, and hence in one type of consumption pattern, we can distinguish other consumption patterns (the other five segments), which differ in their total budget distribution between transportation and at-destination expenses, and in the distribution of their at-destination expenditures into basic and discretionary.

To conclude the section regarding practical implications, the results encountered in this dissertation could be transferred to the tourism industry to help it develop better marketing strategies and focus on those segments of tourists who have a major impact on the tourism destination. Knowing how tourists behave at destination is essential for destination marketers. DMOs can hereby dispose actual information on how long different segments of tourists stay, understand tourist expenditure composition, which is yet an under-used variable in tourism research, and adapt tourism products to tourist needs.

As regards the methodological contribution, this dissertation not only uses up-to-date methods, some of which have never been used in tourism research before, but also adapts them to specific tourism research problems. The articles included in this thesis, thus, contain an in-depth literature review on methodology, on length of stay, on survival and logit models, on tourist expenditure and tourist expenditure composition, as well as on CODA and latent class models. We also challenge some methods which have been extensively used previously in the areas we have studied and provide strong statistical arguments that they are not actually appropriate for research objectives similar to ours.

As already mentioned in the third chapter of this thesis, what can be considered as a methodological contribution in the first article, is the use of an ordered logit model to analyse the length of stay variable. The most common method is the survival model, but based on the research by Thrane (2012) providing a critical look at this method, and also based on the fact that the variable is multimodal, we found it more appropriate to use a method which accounts for this multimodality while being more parsimonious than the alternative multinomial logit model.

As far as tourist expenditure is concerned, (i.e. what was outlined in the second and third articles), we used a methodology which has been widely used in geology, biology and chemistry, but never before in tourism. This methodology is called Compositional Data Analysis (CODA) and as it is specifically designed to analyze share, it suits the objective of decomposing the total trip budget into the various expenditure categories (accommodation, shopping, transportation, and activities; among others) and finding differences in how airline tourists allocate their trip budget. Unlike methods which directly fit share to models assuming normality, CODA allows researchers to make correct statistical inferences and, once data are transformed by means of log-ratios, to use any standard statistical method.

The main objective of the second article was to study the determinants of the composition of tourist expenditure, that is, the share of tourism expenditure allocated to the different categories of a trip budget, by taking into account tourist heterogeneity and distinguishing between legacy airline and LCA users. Accordingly, once we transformed the three components of the budget we had (transportation, basic expenses at destination and discretionary expenses at destination) into two log-ratios; we used a standard MANOVA model with the moderating effect of airline type. In summary, what is relevant as a methodological contribution is the first use of CODA in the field of tourism, and more specifically to study a variable which has been extensively analysed as one of the main tourism demand variables, only with methods that did not take into account the compositional restrictions in the data.

In the third article, we take a step further in terms of the methodological contribution. Again we used the variable tourist expenditure composition and CODA, but in this case, the aim of the article was to put forward a method to segment tourists according to their

budget share. Thus, we combined CODA with latent class models by hand-picking the most suitable techniques of these two methods to fulfil the aim of segmenting tourists by their travel budget allocation. With regards to the CODA methodology, we used the Isometric Log-Ratio (ILR) transformation, which defines weighted log-ratios of geometric means of components and preserves distances, which is crucial when it has to do with segmentation purposes. On the other hand, regarding latent class models, we used the most up-to-date techniques to estimate the classes, and to relate them to external variables. Great care has been taken in specifying the variances and covariances within the latent classes in a suitable manner to CODA. Then, to decide the number of classes we estimated, apart from several statistical tools, the most important consideration we took into account was their interpretability. The other technique we hand-picked from latent-class-model variants is the use of the pseudo-class-draw method to introduce predictor and outcome variables without modifying the latent classes with the aim of doing statistical inference with regard to segment characteristics. This method recognizes class membership as random, and it works like multiple imputation. The article is presented as a step-by-step guide for applied researchers interested in segmenting tourists by travel budget share, that is, with similar research questions to ours. In terms of methodology and based on the results found (discussed among the contributions regarding management implications) we conclude that the method we propose is sound, easy to use and can be tailored to the research questions, because of the coherence, interpretability and usability for management of the results.

As regards the substantive results contributing to tourism knowledge, the first article uses moderators to distinguish between (1) booking a package, where the airline is already included and the tourist cannot choose airline type; (2) self-booking and flying by LCA; and (3) self-booking and flying by legacy airline. To our knowledge, the comparison between these types of air travellers and package tourists has never been made, and results showed that legacy and LCA users are quite similar in their behaviour, and the differences in length of stay determinants mainly appeared between the package and non-package traveller. Regarding the variables used, apart from the most common explanatory variables (tourist and trip characteristics), we included the activities undertaken at destination, which had not been included or treated in almost any previous study. We conclude that those who undertake activities at destination stay longer.

In the second article, we show the determinants of tourist expenditure composition, and, to compare both types of air travellers, the airline type was introduced again as a moderating effect. As far as the main results are concerned, we found that the main differences between legacy and LCA users are that income, some countries of residence, the travel group and some of the occupation situations differently affect budget composition depending on the type of airline. There are no relevant differences among air travellers to Spain regarding other explanatory variables. Regarding the manner in which the variables are treated, we would like to highlight the fact there are some variables which may be decided at the same time as the trip budget. In other words, trip budget composition may vary depending on decisions tourists make regarding the type of accommodation, the destination, or the trip duration, among others. Or, depending on the amount of money devoted to the trip, tourists may decide whether to stay in high category or low category hotels, whether to undertake activities, or not, whether to stay for a shorter or longer period, and so on. Since the decision sequence is not clear and is likely to differ from trip to trip, for the sake of preventing endogeneity problems, these variables were not included in the model and we analysed them separately.

The very few differences between LCA and legacy users encountered in the first two articles could be due to the increasing expansion of LCAs in Europe, which are progressively capturing different tourist segments. Moreover, the services offered by the companies themselves (legacy and LCA) are in some cases becoming increasingly similar, which could also contribute to the fewer significant differences observed in our studies. It seems to us that research aiming at discovering differences between LCA and legacy airlines is close to reaching knowledge saturation, at least for the two tourism demand variables analysed in this thesis. The fact that different segments of tourists use both types of airlines leads to heterogeneity within LCA users, and at the same time to homogeneity between both types of airline users. This makes research of heterogeneity within LCA users a more promising field of research in the near future.

Our third article thus, analyses heterogeneity of LCA users. The article shows how different segments of LCA tourists differ in terms of tourist expenditure composition, in other words, segments LCA users based on travel budget share. To our knowledge LCA users have never been segmented before with regards to any expenditure variable, and

expenditure share has never been used for segmenting purposes in any tourism field. Even though the main contribution of the third article is methodological, we have highlighted the results found in the application presented as the example, which appear to be both relevant and useful for DMO's and tourism marketing purposes. An extended presentation of the segments was already made among the implications for management.

Having discussed the practical, methodological and tourism knowledge contributions, we would like to highlight that this thesis has a balance between these three types of contributions, by using common and traditional variables in theoretical tourism demand analysis, together with more rarely used variables which are important for management (budget share and activities, among others), by formally comparing airline types, and by using new methods never before used in the field of tourism. Consequently, we have both contributed to the tourism and airline demand research in Spain, and suggested appropriate methods, which can be used to fulfil similar objectives to ours in other countries, and which we would like to encourage other researchers to use.

Based on the gaps encountered in the most relevant literature about length of stay, tourist expenditure and Spanish airline tourism demand, and using a large country wide data base, we outlined our objectives, which have been accomplished, and we provide useful results both for the academia and tourism destination management, using a combination of methods never before used in the tourism field.

However, there have been some limitations and further research is needed. This thesis opens new research opportunities regarding the variables and methods used as well as the results obtained.

As limitations, we would like to highlight that even though we had an official statistics database with a large sample size and thus ensuring statistically significant results, it represents a limitation when it comes to the range of available variables because it cannot be controlled by researchers. The data availability also has profound consequences on budget share research and in this respect there is a difference between the many studies using macro data and the yet few using micro data. In micro data studies the coexistence of full-board, half board, bed & breakfast and accommodation-

only, not to speak of package tourists, makes it impossible to meaningfully separate accommodation and food expenditure. Other subdivisions (e.g. discretionary expenditure on activities, non-grocery shopping and moving around the destination) are technically feasible but can only increase the percentages of zero responses in some components, which represents a problem when using the CODA methodology. So, to avoid this, CODA on micro data involves some degree of amalgamation of components.

The time dimension represents another important issue to take into account when outlining further research. The EGATUR survey is conducted annually, so further research might be done to include a repeat cross-section analysis in order to capture trends in the effects of the explanatory variables used, as well as the evolution of tourist expenditure composition or its market segments. Focusing on other further research possibilities, it would be interesting to extend the research on Spanish air tourism demand to variables other than expenditure and length of stay. In other words, to check if the conclusion outlined in this thesis regarding the ever closing differences between LCA and legacy airlines applies to other tourism demand variables too, such as destination loyalty, use of the Internet, travelling season or length of time travellers book in advance. Furthermore, even though comparing LCA and legacy users would be interesting regarding the variables just mentioned, it would be more relevant to go further into the heterogeneity amongst LCA users, also using these tourism demand variables. Finally, in order to extend the findings to a wider context it would also be interesting to do similar research for other destinations especially those where LCA have a relevant market share, or are increasing their share. In other words, to test if the conclusions outlined in this thesis regarding the ever closing differences between LCA and legacy airlines applies to other tourism destinations.

Another possible direction for future research refers to the segmentation of airline-company users. We think that segmenting legacy airline users by expenditure composition would also result in relevant varying patterns amongst legacy users. Once we had the segments of LCA and legacy users, we could compare them, just in order to confront the idea that LCA and legacy users are becoming more and more similar. This can also be done by including the variable "type of company used" in a similar analysis to that carried out in the third article.

6 GENERAL CONCLUSIONS

Last, but not least, we would like to conclude this dissertation by highlighting specific findings regarding the variables included and the methods considered. In this chapter we bring to the fore the main conclusions which are closely related to the inclusion of specific and new variables (such as activities undertaken at destination in the first article, the use of moderators in both the first and second articles, and budget share in the second and third articles), the use of new methods, such as the ordered logit models (first article), CODA (second and third articles) or latent class models (third article) which constitute an innovation when analyzing variables which have been widely studied (length of stay, expenditure, tourist background and trip characteristics).

In the findings of the first article, we focus on the main contributions of introducing the activities undertaken and the moderating effect of how the trip is booked:

- ➢ Duration is longer for almost all activities undertaken when compared with no activities undertaken.
- ➢ In general terms, tourists doing nautical sports, hiking, visiting to a spa, going to a theme park, cultural visits, nightlife or visiting friends and relatives spend more days at the destination.
- ➢ Those tourists who do other sports generally stay longer at the destination, but those who book the trip as part of a package stay less time than those who book it themselves, regardless of the type of airline. For those who attended sporting and cultural events, only LCA users have substantially extended trip durations.
- ➢ Visitors using legacy airlines and LCAs show the main increases in duration for home ownership, family/friends' housing and other types of accommodation. Notwithstanding, package travellers have longer stays in other kinds of hotel.
- ➢ Booking the trip as part of a package makes stays in main cities even shorter.
- ➢ How the trip is booked moderates length of stay for all socio-demographic variables. When older travellers use LCAs or legacy airlines, there is a greater increase in their length of stay than when they book a package deal. As for level of income, high-income tourists who use a legacy airline have the longest increase in stay.

> ➢ Tourists from Germany travelling as part of a package or using an LCA stay longest, followed by tourists from Scandinavia using both LCA and legacy airlines. On the other hand, tourists from the Benelux countries stay longer when they travel on package trips and behave similarly to UK visitors when they book the trip themselves.

Since the second article analyzes a new variable altogether, and does so with a new method, below we highlight all the main relevant conclusions, not only those regarding moderating effects and activities:

> ➢ Legacy users spend a higher proportion on transportation compared to other expenses, and LCA users spend more on basic expenses compared to discretionary.
>
> ➢ Educated LCA users utilize the savings in transport to increase expenditure on non-transportation expenses, and even more so on discretionary expenses.
>
> ➢ The higher the income, the lower the transport share, and within the at destination expense a higher portion is devoted to the discretionary component.
>
> ➢ French, Italians and Portuguese have the highest discretionary share as compared to the basic share, amongst LCA users. Other European countries have the highest discretionary share amongst legacy airline users.
>
> ➢ Travelling as a family with legacy airlines increases the basic expenses share within at-destination expenses. Travelling with friends increases the overall at-destination share for both airline types. Besides, those who travel with friends and LCA spend more in the discretionary share compared to the basic share.
>
> ➢ Being a pensioner results in a higher part of at-destination expenses being devoted to basic expenses for both airline types.

Finally, with reference to the third article, the market segments have already been extensively presented in the previous chapter. In addition to that, we can conclude:

> ➢ We have found heterogeneity in how LCA users distribute their trip budget, that is, there are significantly different segments.
>
> ➢ Segments present differences in both how the trip budget is distributed between transportation and at-destination expenses and, in the case of at-destination expenses, in how they are distributed between basic and discretionary components.

➢ Segments are meaningfully related to both tourist background and trip characteristics.

7   REFERENCES

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

Aguiló, E., Riera, A. and Rosselló, J. (2005). The short-term price effect of a tourist tax through a dynamic demand model: the case of the Balearic Islands. *Tourism Management, 26*, 359-365.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability.* London: Chapman and Hall. (Reprinted in 2003 with additional material by The Blackburn Press).

Alegre, J. and Cladera, M. (2010). Tourist expenditure and quality: why repeat tourists can spend less than first-timers. *Tourism Economics, 16*, 517-533.

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

Alegre, J. and Garau, J. (2010). Tourist Satisfaction and dissatisfaction. *Annals of Tourism Research, 37*, 52-73.

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

Alegre, J., Mateo, S. and Pou, LL. (2009). Participation in tourism consumption and the intensity of participation: an analysis of their socio-demographic and economic determinants. *Tourism Economics, 15*, 531-546.

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

Alegre, J. and Pou, L. (2004). Micro-economic determinants of the probability of tourism consumption. *Tourism economics, 10*, 125-144.

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

Alegre, J. and Pou, L. (2007). Microeconomic determinants of the duration of stay of tourists. In Á. Matias, P. Nijkamp and P. Neto (Eds.), *Advances in Modern Tourism Research*, (pp. 181-206). Heidelberg: PysicaVerlag.

Archer, B.H. and Fletcher, J. (1996). The economic impact of tourism in the Seychelles. *Annals of Tourism Research, 23*, 32-47.

Archer, B.H. and Shea, S. (1975). Length of stay problems in tourism research. *Journal of Travel Research, 13*, 8-10.

Artal-Tur, A., García-Sánchez, A. and Sánchez-Garcia, J.F. (2008). The length of stay determinants for sun-and-sand tourism: an application for the region of Murcia. *Rect@, Acta,* 16, 801-813.

Barros, C.P., Butler, R. and Correia, A. (2010). The length of stay of golf tourism: a survival analysis. *Tourism Management, 13*, 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*, 692-706.

Bewley, R. (1982). The generalised addilog demand system applied to Australian time series and cross section data. *Australian Economic Papers*, *21*, 177-192.

Bewley, R. and Fiebig, D.G. (1988). Estimation of price elasticities for an international telephone demand model. *The Journal of Industrial Economics*, *36*, 393-409.

Brida, J.G. and Scuderi, R. (2013). Determinants of tourist expenditure: a review of microeconometric models. *Tourism Management Perspectives, 6*, 28-40.

Bull, A. (1995). *The Economics of Travel and Tourism*. Melbourne: Longman.

Calderón, M.E. and Ruiz, A.V. (2002). La demanda del destino turístico Burgos: situación actual y retos futuros. *Cuadernos de Turismo, 9*, 5-18.

Castillo-Manzano, J.I. and Marchena-Gómez, M. (2011). Analysis of determinants of airline choice: profiling the LCC passenger. *Applied Economic Letters, 18*, 49-53.

Chiou, Y-C. and Chen Y-H. (2010). Factors influencing the intentions of passengers regarding full service and low cost carriers: a note. *Journal of Air Transport Management*, *16*, 226-228.

Clark, S. and Muthén, B.O. (2009). Relating latent class analysis results to variables not included in the analysis. Retrieved from www.statmodel.com/download/relatinglca.pdf 03/10/2012.

Coenen, M. and van Eekeren, L. (2003). A study of the demand for domestic tourism by Swedish households using a two-staged budgeting model. *Scandinavian Journal of Hospitality and Tourism*, *3*, 114-133.

Crouch, G.I. (1994). The study of international tourism demand: a review of findings. *Journal of Travel Research, 33*, 12-23.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review, 70*, 312–326.

Díaz-Pérez, F.M., Bethencourt-Cejas, M. and Álvarez-González, J.A. (2005). The segmentation of Canary Island tourism markets by expenditure: implications for tourism policy. *Tourism Management*, *26*, 961-964.

Divisekera, S. (2007). *Modelling and Estimation of Tourism Demand Elasticities: A Study of Tourism Expenditure Allocation in Australia*. Gold Coast: Cooperative Research Centre for Sustainable Tourism (CRCST).

Divisekera, S. (2009). Ex post demand for Australian tourism goods and services. *Tourism Economics*, *15*, 153-180.

Divisekera, S. (2010). Economics of leisure and non-leisure tourist demand: a study of domestic demand for Australian tourism. *Tourism Economics*, *16*, 117-136.

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing, 12,* 1-22.

Dwyer, L., Forsyth, P. and Dwyer, W. (2010). *Tourism Economics and Policy*. Cheltenham: Channel View Publications.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, *37*, 795-828.

Espino, R., Martín, J.C. and Román, C. (2008). Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context. *Transportation Research, Part E, 44*, 593-606.

Ferrer-Rosell, B., Martínez-Garcia, E. and Coenders, G. (2012). Does travelling with low cost affect trip duration? In *Papers of the First Congress of Leisure and Tourism*, (pp. 129-142). Mataró: Escola Universitària del Maresme.

Fleischer, A., Peleg, G. and Rivlin, J. (2011). The impact of changes in household vacation expenditures on the travel and hospitality industries. *Tourism Management*, *32*, 815-821.

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

Forgas, S., Moliner, M.A., Sánchez, J. and Palau, R. (2010). Antecedents of airline passenger loyalty: low-cost versus traditional airlines. *Journal of Air Transport Management, 16*, 229-233.

Fry, T. (2011). Applications in economics. In V. Pawlowsky-Glahn, and A. Buccianti (Eds.), *Compositional Data Analysis. Theory and Applications*, (pp. 318-326). New York: Wiley.

Fry, J.M., Fry, T.R.L. and McLaren, K.R. (1996). The stochastic specification of demand share equations: restricting budget share to the unit simplex. *Journal of Econometrics*, *73*, 377-385.

Fry, J.M., Fry, T.R.L and McLaren, K.R. (2000).Compositional data analysis and zeros in micro-data. *Applied Economics*, *32*, 953-959.

Fujii, E.T., Khaled, M. and Mak, J. (1985). An almost ideal demand system for visitor expenditures. *Journal of Transport Economics and Policy*, *19*, 61-171.

Garcia-Ferrer, A. and Queralt, R.A. (1997). A note on forecasting international tourism demand in Spain. *International Journal of Forecasting, 13*, 539-549.

Garín-Muñoz, T. (2006). Inbound international tourism to Canary Islands: a dynamic panel data model. *Tourism Management, 27*, 281-291.

Garín-Muñoz, T. (2007). German demand for tourism in Spain. *Tourism Management, 28*, 12-22.

Garín-Muñoz, T. (2011). La demanda de turismo británico en España. *Boletín Económico de ICE, 3010*, 49-62.

Garín-Muñoz, T. and Perez-Amaral, T. (2000). An econometric model for international flows to Spain. *Applied Economics Letters, 7*, 525-529.

Gokovali, U., Bahar, O. and Kozak, M. (2007). Determinants of length of stay: a practical use of survival analysis. *Tourism Management, 28*, 736-746.

González, P. and Moral, P. (1995). An analysis of international tourism demand. *International Journal of Forecasting, 11*, 233-251.

González, P. and Moral, P. (1996). Analysis of tourism trends in Spain. *Annals of Tourism Research, 32*, 739-754.

Hong, S.K. and Jang, H. (2005). Factors influencing purchasing time of a new casino product and its managerial implications: an exploratory study. *Journal of Travel Research, 43*, 395-403.

Houthakker, H.S. (1960). Additive preferences. *Econometrica*, *28*, 244-256.

Kim, Y.K. and Lee, H.R. (2011). Customer satisfaction using low cost carriers. *Tourism Management*, *32*, 235-243.

Legohérel, P. (1998). Toward a market segmentation of the tourism trade: expenditure levels and consumer behaviour instability. *Journal of Travel and Tourism Marketing*, *7*, 19-39.

Legohérel, P. and Wong, K.K.F. (2006). Market segmentation in the tourism industry and consumer's spending: what about direct expenditures? *Journal of Travel and Tourism Marketing, 20*, 15-30.

Li, G., Song, H. and Witt, S.F. (2004). Modelling tourism demand: a dynamic linear AIDS approach. *Journal of Travel Research*, *43*, 141-150.

Li, G., Song, H. and Witt, S.F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research, 44*, 82-99.

Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research*, 24, 835-849.

Lim, C. (2006). Tourism demand modelling: issues and implications. In L. Dwyer and P. Forsyth (Eds.), *International Handbook of Tourism Economics*, (pp. 45–72). London: Edward Elgar.

Liu, Y., Kiang, M. and Brusco, M. (2012). A unified framework for market segmentation and its applications. *Expert Systems with Applications, 39*, 10292-10302.

Lopez-Delgado, P., Alarcón-Urbistondo, P. and Martín-Armario, J. (2000). El posicionamiento de las comunidades autónomas en la demanda turística extranjera. *Estudios de Economía Aplicada, 16*, 111-131.

Machado, L.P. (2010). Does destination image influence the length of stay in a tourism destination? *Tourism Economics, 16*, 443-456.

Magidson, J. (2002). Latent class models: a comparison with k-means. *Canadian Journal of Marketing Research, 20*, 37-44.

Mak, J. and Moncur, J. (1979). The choice of journey destinations and length of stay: a micro analysis. *The Review of Regional Studies, 10*, 38-48.

Mak, J., Moncur, J. and Yonamine, D. (1977). Determinants of visitor expenditure and visitor length of stay: a cross-section analysis of U.S. visitors to Hawaii. *Journal of Travel Research, 15*, 5-8.

Martínez-Garcia, E., Ferrer-Rosell, B. and Coenders, G. (2012). Profile of business and leisure travelers on low cost carriers in Europe. *Journal of Air Transport Management, 20*, 12-14.

Martínez-Garcia, E. and Raya, J.M. (2008). Length of stay for low-cost tourism. *Tourism Management, 29*, 1064-1075.

Martínez-Garcia, E. and Raya, J.M. (2009). Determinantes de la demanda temporal de turismo: una aproximación macroeconómica con un modelo de duración. *Investigaciones Económicas, 33*, 271-302.

Martínez-Garcia, E. and Royo-Vela, M. (2010). Segmentation of low-cost flights users at secondary airports. *Journal of Air Transport Management*, *16*, 234-237.

McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models. Inference and Applications to Clustering*. New York: Marcel Dekker.

McLaren, K., Fry, J. and Fry, T. (1995). A simple nested test of the almost ideal demand system. *Empirical economics*, *20*, 149-161.

Medeiros, M.C., McAleer, M., Slottje, D., Ramos, V. and Rey-Maquieira, J. (2008). An alternative approach to estimating demand: neural network regression with conditional volatility for high frequency air passenger arrivals. *Journal of Econometrics, 147*, 372-383.

Medina-Muñoz, D.R. and Medina-Muñoz, R.D. (2012). Determinants of expenditures on wellness services: the case of Gran Canaria. *Regional Studies, 46*, 309-319.

Menezes, A.G., Moniz, A. and Vieira, J.C. (2008). The determinants of length of stay of tourists in the Azores. *Tourism Economics, 14*, 205-222.

Mihalič, T. (2002). Tourism and economic development issues. In R. Sharpley and D.J. Telfer (Eds.), *Tourism and Development: Concepts and Issues,* (pp. 81–111). Clevedon, UK: Channel View Publications.

Muthén, B.O. (2002). Beyond SEM: general latent variable modelling. *Behaviormetrika*, *29*, 81–117.

Nicolau, J.L. (2011). Testing prospect theory in airline demand. *Journal of Air Transport Management, 17*, 241-243.

Nicolau, J.L. and Más, F.J. (2005). Heckit modelling of tourist expenditure: evidence from Spain. *International Journal of Service Industry Management*, *16,* 271-293.

Nicolau, J.L. and Más, F.J. (2006). Elección de la duración de las estancias turísticas: una aproximación con modelos de recuento. *Revista Europea de Dirección y Economía de la Empresa, 5*, 99-115.

Nicolau, J.L. and Más, F.J. (2009). Simultaneous analysis of whether and how long to go on holidays. *The Service Industries Journal, 29*, 1011-1092.

O'Connell, J.F. and Williams, G. (2005). Passengers' perception of low cost airlines and full service carriers: a case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines. *Journal of Air Transport Management*, *11*, 259-272.

O'Hagan, J.W. and Harrison, M.J. (1984). Market share of US tourist expenditure in Europe: an econometric analysis. *Applied Economics*, *16*, 919-931.

Pardellas de Blas, X.X., Padín-Fabiero, C. and Bouzada-Fernández, X. (2003). Comportamiento de la demanda turística en la ría de Pontevedra. *Cuadernos de Tursimo, 11*, 155-167.

Peypoch, N., Randriamboarison, R., Rasoamananjara, F. and Solonandrasana, B. (2012). The length of stay of tourists in Madagascar. *Tourism Management, 33,* 1230-1235.

Pulido-Fernández, J.I. and Sánchez-Rivero, M. (2010). Attitudes of the cultural tourist: a latent segmentation approach. *Journal of Cultural Economics, 34*, 111-129.

Raya-Vilchez, J.M. and Martínez-Garcia, E. (2011). Nationality and low-cost trip duration. A microeconometric analysis. *Journal of Air Transport Management*, *17*, 168-174.

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

Rodriguez, S., Dávila, D. and Rodríguez, A. (2003). Determinantes de la demanda de turismo vacacional en las Islas Canarias. *Investigaciones Regionales, 3*, 81-98.

Rosselló, J., Riera, A. and Sansó, A. (2004). The economic determinants of season patterns. *Annals of Tourism Research, 31*, 697-711.

Sainaghi, R. (2012). Tourist expenditures: the state of the art. *Anatolia: An International Journal of Tourism and Hospitality Research*, *23*, 217-233.

Silberman, J. (1985). A demand function for length of stay: the evidence form Virginia Beach. *Journal of Travel Research, 4*, 16-23.

Sinclair, T.M. (1998). Tourism and economic development: a survey. *Journal of Development Studies, 34*, 1-15.

Song, H., Dwyer, L., Li, G. and Cao, Z. (2012). Tourism economics research: a review and assessment. *Annals of Tourism Research, 39*, 1653-1682.

Song, H. and Li, G. (2008). Tourism demand modelling and forecasting – a review of recent research. *Tourism Management, 29*, 203-220.

Syriopoulos, T.C. and Sinclair, T. (1993). An econometric study of tourism demand: the AIDS model of US and European tourism in Mediterranean countries. *Applied Economics*, *25*, 1541-1552.

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. and Farstad, E. (2012). Tourists' length of stay: The case of international summer visitors to Norway. *Tourism Economics, 18,* 1069-1082.

Thumberg, E.M. and Crotts, J.C. (1994). Factors affecting traveler's overnight stay behavior. *Journal of Travel & Tourism Marketing, 3*, 1-17.

Wang, Y. and Davidson, M.C.G. (2010). A review of micro-analyses of tourist expenditure. *Current Issues in Tourism*, *13*, 507-524.

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.

WTO (2013) United Nations World Tourism Organisation Tourism Highlights, 2013 Edition. Retrieved from http://dtxtq4w60xqpw.cloudfront.net/sites/all/files/pdf/unwto_highlights13_en_hr.pdf 10/06/2013.

Wu, D.C., Li, G. and Song, H. (2011). Analyzing tourism consumption: a dynamic system-of-equations approach. *Journal of Travel Research*, *50*, 46-56.

Yang, Y., Wong, K.K.F. and Zhang, J. (2011). Determinants of length of stay for domestic tourists: case study of Yixing. *Asia Pacific Journal of Tourism Management Research, 16*, 619-633.

8 APPENDIX. TRANSCRIPTION OF OTHER PUBLICATIONS ON LCA DEMAND WRITTEN DURING THE PHD

Martínez-Garcia, E., Ferrer-Rosell, B., Coenders, G. (2012). Profile of business and leisure travelers on low cost carriers in Europe. *Journal of Air Transport Management*, 20, 12-14.

Ferrer-Rosell, B., Martínez-Garcia, E., Coenders, G. (2012). Does travelling with low cost affect trip duration? In: *Papers of the First Congress of Leisure and Tourism*. Escola Universitària del Maresme, Mataró: 129-152. ISBN 978-84-695-3343-7.

Martínez-Garcia, E., Ferrer-Rosell, B. and Coenders, G. (2012). Profile of business and leisure travelers on low cost carriers in Europe. *Journal of Air Transport Management*, 20, 12-14.

Abstract

Low cost carriers are increasingly attracting a market segment traditionally associated with legacy airlines, the business traveler. Little, however, is known regarding the characteristics of this new demand. This study analyzes the differences between business and leisure travelers flying by low cost airlines. To do this, we surveyed a sample of tourists in Catalonia, one of Europe's most popular tourist destinations. While some differential characteristics emerge between the traveler types, still more similarities are found.

Keywords

Ferrer-Rosell, B., Martínez-Garcia, E. and Coenders, G. (2012). Does travelling with low cost affect trip duration? In: Papers of the First Congress of Leisure and Tourism. Escola Universitària del Maresme, Mataró: 129-152. ISBN 978-84-695-3343-7.

ACTES PRIMER CONGRÉS D'OCI I TURISME

ACTAS PRIMER CONGRESO DE OCIO Y TURISMO

PAPERS OF FIRST CONGRESS OF LEISURE AND TOURISM

OCITUR 2012

Escola Universitària del Maresme

Centre adscrit a:

UNIVERSITAT POMPEU FABRA

TecnoCampus Mataró-Maresme

RESUM

El treball analitza els determinants de la durada de l'estada del turisme receptor en un dels països més turístics del món, Espanya, amb especial èmfasi en els efectes que pot tenir el fet d'organitzar el viatge amb paquet turístic o sense i viatjar amb una aerolínia de baix cost (LCA) o una de tradicional. S'estima un model logit on les variables explicatives rellevants estan relacionades amb les preferències i caracterítiques dels turistes, les característiques del viatge, les característiques de l'estada i les activitats en destinació. Un dels principals resultats fa referència als efectes moderadors que té viatjar amb un paquet turístic o amb una companyia de baix cost, en particular.

Paraules clau: Logit ordinal, turismo de baix cost, efecte moderador, durada de l'estada

ABSTRACT

The paper analyses the determinants of the length of stay of inbound tourism in one of the world most touristic countries, Spain, with special emphasis on the effects of organizing the flight with package or by oneself and travelling with a low cost airline (LCA) or a legacy one. An ordered logit model is estimated. Relevant explanatory variables are related to tourists' preferences and characteristics, trip characteristics, stay characteristics, and activities at destination. One of the main relevant results concerns the moderating effects of travelling with an organised package and with a LCA, specifically.

Keywords: Ordered Logit, Low Cost tourism, moderating effect, length of stay

1. INTRODUCTION

The increasing expansion and consolidation of no-frills airlines, also referred to as low cost airlines (LCA), results in a significant reduction in travel costs, thereby bringing destinations closer to a greater number of tourists while allowing an increasing number of trips throughout the year, especially abroad, as opposed to what might be a single annual holiday typically made in summer. All of the above would appear to contribute to a trend in reducing the average length of stay at destination.

The aforementioned expansion of LCA flights is particularly evident in Spain, one of the world's most important tourist destinations. According to IET (*Instituto de Estudios Turísticos* – the Spanish Institute for Tourism Studies), Spain received 52.7 million tourists in 2010, 77% of whom travelled by air. Of the latter, 56% flew with an LCA and 84.3% lived in a European country.

The relevance of low-cost tourism today is also reflected in the increasing number of academic studies emerging on the subject, although these are still among a minority. For its part, microeconomic research on the determinants of trip duration has increased in recent years (see Alegre, Mateo & Pou, 2011; Alegre & Pou, 2006; Barros & Machado, 2010; Martínez-Garcia & Raya, 2008; Menezes, Moniz & Vieira, 2008; Yang, Kevin & Zhang, 2011, among others) probably due to destinations' growing interest in obtaining more information on said determinants, given the observed reduction in trip length and in many cases its associated lower expenditure. However, to the best of our knowledge no study has yet analyzed the moderating effect on length of stay of booking the trip as a package or not and of travelling with an LCA or a legacy airline.

The main aim of this paper is to study the length of stay at destination of tourists travelling to Spain by air. It analyzes the impact of them booking the trip themselves and travelling with an LCA, booking it themselves and flying with a legacy company, or travelling as part of an organized package holiday. Besides the usual explanatory variables relating to individual, trip and stay characteristics, the moderating effects between booking the trip and the remaining explanatory variables are also of key interest. This paper has some other new elements with respect to that published to date. Firstly, the inclusion of activities undertaken at the destination. Secondly, the scope of the study, which is for a whole country (Spain) rather than just an airport or airline.

Thirdly, the use of multinomial logit models to account for the multimodality observed in the trip duration variable.

This paper is structured as follows: we first present the literature relevant to the present study and then introduce the methodology used to estimate the model. This is followed by a description of the variables and the results, and then finally the overall conclusions.

2. LITERATURE REVIEW

Research on length of stay has recently attracted the interest of several researchers. There was little written on the subject prior to 2008, but since then the number of published econometric studies has been rising. Despite this increase, few studies have focused explicitly on the air transportation segment, and only those by Martínez-Garcia and Raya (2008, 2009) and Raya-Vilchez and Martínez-Garcia (2011) referred to LCA demand. Most authors have analyzed different specific regions (Barros et al., 2008; Menezes et al., 2008; Peypoch et al., 2011; Yang et al., 2011) and a few of them, specific demand segments, such as golf tourists (Barros et al., 2010).

Although various different theoretical frameworks have been produced to analyze length of stay, most studies agree on including socio-demographic variables in the model, as well as variables like trip and stay characteristics. In this paper we propose a new factor affecting length of stay: activities undertaken at destination. To the best of our knowledge this has not been taken into account in a systematic manner before. Only Barros and Machado (2010) included a limited number of activities at the destination in their model.

Socio-demographic factors and trip and stay characteristics as determinants of length of stay have already been found in studies by Oppermann (1995, 1997), Seaton and Palmer (1997) and Sung, Morrison, Hong and O'Leary (2001), Peypoch et al. (2011) and Gokovali et al. (2007). Alegre and Pou (2006) and Alegre et al. (2011) also included expenditure, price, loyalty and trip motivation, Barros et al. (2008) destination attributes, Menezes et al. (2008) trip attributes and destination image, Martínez-Garcia and Raya (2008) time constraints, budget restrictions and prices. Martinez-Garcia and Raya (2009)

destination, motivation and season, Barros et al. (2010) climate and events, and Machado (2010) image.

The variable we use as moderator in our study – how the trip is booked – has been employed as an explanatory variable with only main effects in some studies. Alegre and Pou (2006), Martínez-Garcia and Raya (2008, 2009) found that booking a package holiday was not significant. On the other hand, Menezes et al. (2008) claimed that taking a charter flight increases expected length of stay. Yang et al. (2011) went one step further by analyzing subsamples in a particular area in China and found that there were differences in the factors affecting length of stay depending on how the trip was booked (package or individual tourists). Individual tourists stayed significantly longer when their reason for travel was visiting friends and relatives rather than sightseeing. However, package tourists on vacations stayed longer than sightseeing package tourists, while those travelling individually did not. These findings encourage the formal inclusion and testing of moderating effects.

3. METHODOLOGY

The length of stay problem can be approached from general population surveys or from surveys of tourists carried out at their destinations. The type of survey has an influence on the econometric approach, the set of available variables, and the type of tourist decisions which can be modelled. The survey we use was carried out at destination and the methodology review below is centred on this framework.

Estimating equations predicting tourists' length of stay poses considerable econometric challenges due to the limited dependent variable, which is integer and positive. Drawing on analogies with survival research, survival models have been used in a major stream of research (Barros et al., 2008; Barros et al., 2010; Barros & Machado, 2010; Gokovali et al., 2007; Hong & Jang, 2005; Machado, 2010; Martinez-Garcia & Raya, 2008; Menezes et al., 2008; Peypoch et al, 2012; Raya-Vilchez & Martínez-Garcia, 2011). Drawing on the sheer distributional characteristics of the dependent variable, other authors use Poisson regression (e.g. Rodríguez, Dávila & Rodríguez, 2003) or negative binomial regression (e.g. Nicolau & Más, 2006).

Each of these approaches assumes specific data generation processes, which may be considered somehow unrealistic for a model predicting length of tourist stays (Thrane, 2012). In addition, all of the aforementioned methods assume unimodal distributions for the dependent variable. Lengths of stay usually feature multiple modes around a few typical durations (e.g. long week-end, week, two weeks, and the like). Alegre and Pou (2006) encounter a bimodal distribution and group it into two categories to use a binary logit model. Alegre et al. (2011) use finite mixture models to accommodate the multiple modes.

Other feasible approaches are ordered logit models and multinomial logit models, for three or more modes around which a categorical length of stay variable with three or more categories is created (for an overview, see Hosmer and Lemeshow, 2000). To our knowledge, only Yang et al. (2011) have thus far used the ordered logit model, and only Nicolau and Más (2009) have used the multinomial logit model, albeit within the context of a general population survey.

The multinomial model is the most general, as it implies estimating a separate equation for each category, making it possible for each specific length of stay to have predictors operating in a specific way. For *K* duration categories we have:

$$\frac{\Pr ob(Y = k)}{\Pr ob(Y = 1)} = \exp(\mathbf{x_i}\boldsymbol{\beta_k}) \quad \text{with } k=2,3,...,K \tag{1}$$

where $\mathbf{x_i}$ is a row vector of characteristics of the trip made by the *i*th individual and $\boldsymbol{\beta_k}$ is the column vector of coefficients for the *k*th category of the dependent variable. The first category is assumed to be the reference category and does not have any associated $\boldsymbol{\beta}$ vector. The exponential function of each β_l coefficient is interpreted as an odds ratio of belonging to the *k*th or 1st category, when increasing the corresponding x_l predictor by one unit or when moving to a given category of a categorical binary-coded predictor (1) from the reference category (0). All variables in our model are, in fact, categorical:

$$\frac{\left.\dfrac{\Pr ob(Y = k \, / \, x_l = 1)}{\Pr ob(Y = 1 \, / \, x_l = 1)}\right.}{\left.\dfrac{\Pr ob(Y = k \, / \, x_l = 0)}{\Pr ob(Y = 1 \, / \, x_l = 0)}\right.} = \exp(\beta_{kl}) \qquad (2)$$

Moderating effects, also called interaction effects, can be accommodated in the usual manner as products of variables. In this paper we consider interactions between how the trip is booked (as a package, without package with legacy airline, and without package with LCA) and all other variables in the model.

The multinomial model with moderating effects requires very large sample sizes: the three-way contingency tables between each predictor, each duration and the moderator variable can contain no zeroes. Despite the large sample size used in our study, the moderating effect of type of accommodation could not be estimated.

We draw on a representative sample of European leisure tourists arriving by air at 27 major Spanish airports during 2010 and staying overnight (EGATUR, which stands for *Encuesta de Gasto Turístico* – Survey of tourism expenditure, conducted by the *Instituto de Estudios Turísticos* – Institute for Tourism Studies). Consistent with this definition, we did not consider business trips, study trips, and trips without an overnight stay. We excluded flights from outside Europe because LCA mostly operate regional flights. We also centred our study only on those trips with one single destination, thus excluding multi-stage trips.

Stays of over 120 days (0.3 % of cases) are excluded. The final sample size was *n*=61,334. The large sample size makes it possible to use low p-values. All the variables and moderating effects included in the final model are significant at 0.01. SPSS 19 software is used to estimate all the models.

4. RESULTS

As observed in Figure 1, the duration of stay is multimodal, with modes of 3-4 nights, 7 nights, 10 and 14 nights. Thus, we have created the categorized length of stay variable around these modes: 5 nights and fewer, from 6 to 8 nights, from 9 to 12 nights, and more than 13 nights.

Figure 1. Dependent variable length of stay
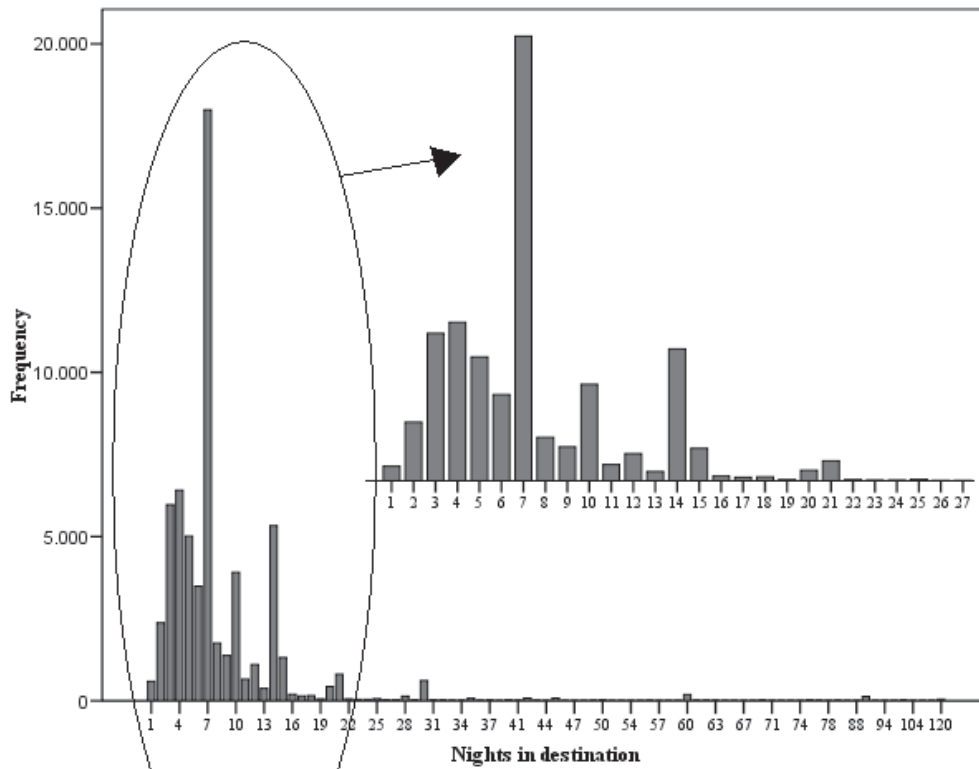
Table 1. Categorized dependent variable by how trip is booked

Length of stay	Legacy	Package	LCA
5 and fewer*	44.7%	11.0%	40.9%
6 to 8	26.8%	57.9%	31.2%
9 to 12	11.1%	11.9%	11.4%
13 and over	17.3%	19.1%	16.5%

* Reference dependent variable category in the multinomial logit model

Table 1 shows the bivariate relationship between length of stay and how the trip is booked. It can be seen that only small differences seem to exist between those tourists arriving with a legacy airline and those arriving with an LCA. The main difference is found between booking the trip oneself (and using either type of airline) and travelling as part of a package.

Table 2 shows frequency distributions of all explanatory variables. As far as activities at destination are concerned, eleven activities are included in the model. Regarding trip characteristics, six variables are included. Previous stays in Spain; the season (coded as binary: summer season vs. otherwise); type of accommodation; travel group; destination (the two Spanish major cities, Madrid and Barcelona, other unique capital cities and other locations); and, finally, time of booking, which has two categories. Socio-demographic variables are age, gender, self-reported income category, country of residence and level of education.

Table 3 presents the multinomial logit results. The large sample size makes it unsurprising that the estimated model was globally significant (overall likelihood ratio test for all variables 33,271 with 243 d.f.). The goodness of fit measures are more informative, Nagelkerke's R-squared being 0.453 and McFadden's 0.211. The large sample size also makes all variables significant individually.

The fact that all variables are treated as categorical and coded as 0-1 dummy variables makes the size of odd ratios comparable across variables. Since all odd ratios are significant, we concentrate our interpretation on those which are furthest from unity. Reference categories are indicated for multinomial predictors in Table 2. Main and moderating effects have already been summed before computing odd-ratios, for an easier reading of Table 3, in which each column represents the odd ratio within a particular way of booking the trip when shifting from the reference category to a given category of the explanatory variable. If the column values differ for a given row, there is a significant moderating effect (p-value < 0.01). Within each variable, rows represent a given trip length, the shortest trips (5 days or shorter) acting as reference category, in other words, as denominators in the odd ratios.

The coefficients of the length categories are nearly always in ascending or descending order within the explanatory variables' categories and booking form. This means that characteristics increasing the probability ratios of greater trip lengths, increase these ratios further for the greatest lengths. Characteristics decreasing the probability ratios of greater trip lengths, decrease these ratios further for the greatest lengths. When the ascending or descending order is broken, certain characteristics are related to specific trip durations, which are neither the longest nor the shortest. This type of

unordered patterns can only be uncovered by the multinomial logit model, and not with any other of the models mentioned in the methodology section.

Table 2. Frequency distributions of explanatory variables

Activities			
Nautical sports		Other cultural activities	
Yes	3.9%	Yes	16.9%
No*	96.1%	No*	83.1%
Hiking		Spa	
Yes	2.2%	Yes	5.6%
No*	97.8%	No*	94.4%
Other sports		Theme Parks	
Yes	6.3%	Yes	9.8%
No*	93.7%	No*	90.2%
Attendance at sport events		Nightlife	
Yes	2.7%	Yes	24.6%
No*	97.3%	No*	75.4%
Cultural visits		Visiting friends	
Yes	45.8%	Yes	15.8%
No*	54.2%	No*	84.2%
Attendance at cultural events			
Yes	10.9%		
No*	89.1%		
Trip Characteristics			
Has been to Spain before		Season	
Yes	20.5%	Not in summer	70.9%
No*	79.5%	Summer*	29.1%
Travel group		Accommodation used	
With friends	14.5%	Other accommodation	6.3%
In family	16.5%	Home ownership	7.3%
Alone	25.6%	Hotel 3-star	16.5%
With partner*	43.4%	Other kind of hotel	17.4%
		Family/friends' housing	27.2%

		Hotel 4-5 star*	25.2%
Destination		**How trip was booked**	
Unique capitals**	8.5%	Package	28.6%
Madrid	9.0%	No package with LCA	49.0%
Barcelona	14.2%	No package with legacy*	22.4%
Other locations (mostly coastal)*	68.3%		
Time of booking			
Weeks before or less	39.8%		
Months before or longer*	60.2%		

Socio-demographic profile			
Age		**Country of residence**	
65 and more	8.7%	Austria, Switz. & Liechtenstein	5.9%
15-24	12.8%	France	6.9%
45-64	29.0%	Other European Countries	8.5%
25-44*	49.6%	Nordic Countries	10.2%
		Italy	10.8%
		Benelux	11.2%
		Germany	16.3%
		United Kingdom*	30.3%
Gender		**Level of education**	
Male	50.6%	Up to high school	37.2%
Female*	49.4%	University*	62.8%
Income category			
Low	5.5%		
High	27.4%		
Medium*	67.1%		

* Reference categories in the logit model: chosen either because they are the largest categories, or because they are the most standard, conceptually considered.

** The capital towns considered unique are: Bilbao, Córdoba, Girona, Granada, Salamanca, Santiago de Compostela, San Sebastián, Sevilla, Tarragona, Toledo and Valencia.

Table 3. Multinomial Logit results. Odd ratios within how trip was booked and airline type

Variable/category	Duration*	Package	Low cost	Legacy
Nautical sports	6 to 8	2,228	2,228	2,228
	9 to 12	2,298	2,298	2,298
	>12	2,323	2,323	2,323
Other sports	6 to 8	1,702	1,702	1,702
	9 to 12	2,032	2,032	2,032
	>12	2,237	2,237	2,237
Spa visited	6 to 8	1,468	1,468	1,468
	9 to 12	2,109	2,109	2,109
	>12	2,314	2,314	2,314
Visited theme parks	6 to 8	1,547	1,547	1,547
	9 to 12	2,079	2,079	2,079
	>12	2,377	2,377	2,377
Hiking	6 to 8	1,402	1,402	1,402
	9 to 12	1,380	1,380	1,380
	>12	1,644	1,644	1,644
Attended sporting events	6 to 8	0,530	0,770	0,929
	9 to 12	0,353	1,274	0,930
	>12	0,368	1,379	0,913
Attended cultural events	6 to 8	1,285	1,285	1,285
	9 to 12	1,366	1,366	1,366
	>12	1,285	1,285	1,285
Nightlife	6 to 8	1,040	1,040	1,040
	9 to 12	1,377	1,377	1,377
	>12	1,207	1,207	1,207
Cultural visits	6 to 8	1,021	1,021	1,021
	9 to 12	1,228	1,228	1,228
	>12	1,172	1,172	1,172

Other cultural activities	6 to 8	0,730	0,928	1,264
	9 to 12	0,691	0,978	1,058
	>12	0,627	0,927	1,254
Visited friends/relatives	6 to 8	1,093	1,093	1,093
	9 to 12	1,219	1,219	1,219
	>12	1,376	1,376	1,376
Destination unique capitals	6 to 8	0,075	0,350	0,285
	9 to 12	0,040	0,259	0,275
	>12	0,024	0,255	0,164
Destination Madrid	6 to 8	0,008	0,174	0,157
	9 to 12	0,011	0,132	0,171
	>12	0,007	0,119	0,098
Destination Barcelona	6 to 8	0,019	0,205	0,168
	9 to 12	0,014	0,152	0,132
	>12	0,004	0,144	0,085
Other accommodation	6 to 8	2,686	2,686	2,686
	9 to 12	3,904	3,904	3,904
	>12	7,576	7,576	7,576
Home ownership	6 to 8	2,612	2,612	2,612
	9 to 12	4,928	4,928	4,928
	>12	13,805	13,805	13,805
Hotel 3 star	6 to 8	0,881	0,881	0,881
	9 to 12	0,745	0,745	0,745
	>12	0,685	0,685	0,685
Other kind of hotel	6 to 8	1,800	1,800	1,800
	9 to 12	2,100	2,100	2,100
	>12	2,524	2,524	2,524
Family/friends housing	6 to 8	1,992	1,992	1,992
	9 to 12	3,529	3,529	3,529
	>12	4,446	4,446	4,446

Table 3. Continued

Variable/category	Duration*	Package	Low cost	Legacy
Has been to Spain before	6 to 8	0,973	0,973	0,973
	9 to 12	0,695	0,695	0,695
	>12	0,576	0,576	0,576
Season: not in summer	6 to 8	0,427	0,605	0,765
	9 to 12	0,215	0,411	0,548
	>12	0,254	0,360	0,454
Booking weeks before or less	6 to 8	0,771	0,771	0,771
	9 to 12	0,606	0,606	0,606
	>12	0,508	0,508	0,508
Travel group with friends	6 to 8	0,373	0,613	0,790
	9 to 12	0,353	0,549	0,728
	>12	0,245	0,346	0,467
Travel group in family	6 to 8	1,649	1,505	1,570
	9 to 12	1,910	1,817	1,439
	>12	1,996	1,690	1,733
Travel group alone	6 to 8	0,532	0,726	0,85
	9 to 12	0,574	0,824	0,847
	>12	0,543	0,930	0,937
Age 65 and more	6 to 8	3,043	1,564	1,545
	9 to 12	8,298	3,130	2,620
	>12	19,434	7,463	5,675
Age 15-24	6 to 8	1,068	1,088	1,126
	9 to 12	0,935	1,126	1,172
	>12	0,720	1,219	1,112
Age 45-64	6 to 8	1,221	1,241	1,081
	9 to 12	1,680	1,451	1,411
	>12	2,368	1,833	1,495
	6 to 8	1,021	0,798	0,610

Country of residence Austria, Switz. & Liechtenstein	9 to 12	1,395	1,090	0,775
	>12	1,732	1,331	0,688
Country of residence France	6 to 8	1,324	0,882	0,629
	9 to 12	1,160	1,064	0,862
	>12	1,362	1,102	0,694
Country of residence other European Countries	6 to 8	1,978	0,955	0,792
	9 to 12	2,702	1,441	1,165
	>12	2,617	1,495	1,200
Country of residence Nordic Countries	6 to 8	2,280	1,171	1,140
	9 to 12	0,747	1,419	1,507
	>12	2,664	1,726	1,554
Country of residence Italy	6 to 8	1,385	0,701	0,799
	9 to 12	1,297	0,728	0,685
	>12	1,296	0,607	0,614
Country of residence Benelux	6 to 8	1,137	0,755	0,656
	9 to 12	2,883	1,005	1,017
	>12	1,758	0,914	0,748
Country of residence Germany	6 to 8	1,027	1,014	1,443
	9 to 12	3,337	1,665	1,142
	>12	2,583	1,782	1,120
education up to high school	6 to 8	0,873	0,724	0,745
	9 to 12	1,079	0,783	0,954
	>12	0,871	0,696	0,534
Low reported income	6 to 8	1,122	1,122	1,122
	9 to 12	1,142	1,142	1,142
	>12	1,241	1,241	1,241
High reported income	6 to 8	1,021	1,021	1,021
	9 to 12	1,133	1,133	1,133
	>12	1,265	1,265	1,265
	6 to 8	0,908	0,908	0,908

Gender male	9 to 12	0,952	0,952	0,952
	>12	0,883	0,883	0,883

* 5 or fewer days is the reference category

The inclusion of activities undertaken by tourists at destinations is one of the new features included in this paper, and overall it has been found to be relevant to trip duration. Specifically, the effect of undertaking activities on the probability ratios of higher trip durations is mostly positive (odd ratios larger than 1). This means that duration is longer for almost all activities undertaken when compared with the reference of doing no activities. The odd ratios of the length categories are in ascending or descending order within the groups of tourists who undertake activities. For instance, for those who visit Theme Parks, odd ratios increase from 1.547 for the group spending 6 to 8 nights compared to the reference category (<6 nights), to 2.377 for the group spending more than 12 night compared to the reference category. No relevant moderating effects emerge in the case of nautical sports, hiking, cultural visits, going to a spa, going to a theme park, nightlife, visiting friends and relatives, other sports or attendance to cultural events, in other words, on the whole tourists doing these activities spend more days at destination and how the trip is booked is of no importance. For the remaining two activities, moderating effects are found. For those who attended sporting events, only LCA users increase the probability ratios of stays above 9 nights over stays below 6 nights. The remaining probability ratios are below one, thus showing lower probabilities for longer trips for those tourists attending sporting events. For those who undertake other cultural activities, only legacy travellers extend trip duration, even if in that case the order of odd ratios from shorter to longer stays is broken. Legacy carrier flyers attending these activities tend to stay either 6 to 8 nights or above 12 nights.

Regarding trip characteristics, length of stay varies substantially depending on the variable categories. Tourists coming to Spain outside the summer season, or with friends, tend to spend considerably fewer days (odd ratios are smaller than 1) than those coming in summer or those travelling with a partner. In both cases, the order within the length categories is descending. Also in both cases, the moderating effect is significant and the odd ratios are similar

regardless of airline type, but different when booking packages. Conversely, tourists coming with their families, have higher probabilities of longer trips.

As far as accommodation is concerned, ratios are mostly larger than one, thus showing the reference 4-5-star hotels to be related to the shortest stays. Tourists who own a second residence at the destination are those who spend the most days (the coefficients are ascending from 2.612 for the 6 to 8 night group compared to the reference category, to 13.805 for those who stay more than 12 nights); in fact, this is the most distinct category for the length of stay variable. The coefficients of the 3-star hotel category are closest to 1, meaning that tourists using them are indistinguishable from the reference category (4-5-star hotels). As said before, the moderating effect of type of accommodation could not be estimated.

Concerning destination, going to any main city considerably reduces length of stay. Airline type seems to be fairly irrelevant but booking the trip as part of a package makes stays in these main cities even shorter. This reduction is largest in Madrid (odd ratios being 0.008, 0.011 and 0.007), closely followed by Barcelona (being 0.019, 0.014 and 0.004).

No significant moderating effects emerge in the case of having previously been to Spain (odd ratios smaller than 1 and thus negative effect of loyalty on length) nor booking shortly before travelling (negative effect of last-minute booking on length); in both cases, odd ratios are descending orderly (from 0.973 for the 6 to 8 nights group compared to the reference category, to 0.576 for the group staying more than 12 nights, in the case of having been to Spain before, and from 0.771 to 0.508 in the case of last-minute booking).

How the trip is booked moderates length of stay for almost all sociodemographic variables included in the model (only income category and gender have no moderating effect). The oldest tourists stay considerably longer than other age groups, and when they travel as part of a package their stay is even longer (the odd ratio is 19.434 for the group spending more than 12 nights compared to the reference category). The 15-24 age group behaves similarly to the reference 25-44 group. The 45-64 age group behaves similarly to the oldest group, but with lower odd ratios. As with other variables, larger differences are found for package travellers than between airline types.

With regard to the country of residence, the ascending and descending order of odd-ratios is often broken. Differences depending on the way the trip is booked and on airline choice emerge in all directions and differ across countries, reflecting country specific habits regarding trip lengths or country specific policies of tour operators. Large odd ratios are found, for instance, among German and Benelux package travellers staying 9 to 12 nights, among Nordic package travellers staying more than 12 nights, among residents in other European countries booking packages of any length except 5 nights and fewer, and among residents in Germany flying with LCA and staying 9 nights or more.

Tourists who have up to secondary school education tend to stay shorter than those with a university education. The main differences are found between booking oneself and package trips, the latter group showing nearly no effect in terms of education (odd ratios around 1). In this case, as it happens with the country of residence variable, the descending order is broken, and those with high school education or less tend to do stays both below 6 days and from 9 to 12.

Finally, self-reported income and gender have very little impact on duration, all odd ratios being close to 1.

5. CONCLUSIONS

The purpose of this study was twofold. Firstly, to analyze the determinant factors of length of stay at destination for all European tourists arriving in Spain by air (most inbound tourism in Spain is European and travels by air). Secondly, to study the impact of tourists booking the trip themselves and flying with an LCA or legacy airline or travelling as part of an organized package, both in terms of main and moderating effects. The large sample and the multinomial model allow us to analyze how tourists behave regarding each category of the explanatory variables in terms of the probabilities of each length of stay at destination and within each way of booking the trip.

Unlike previous studies, we have subdivided this variable into three categories rather than two (package or not package). Package users and non-package users face different decision constraints for obvious reasons, but we wished to

also highlight differences within non-package users who choose between LCA and legacy airlines.

Some results presented in this study are not unexpected, such as tourists over 65 or those who have a second residence and those accommodated in cheaper accommodation types generally staying longer at destination. In fact, as we observed in the literature review, socio-demographic factors play an important role and country of residence and age are the main variables. In addition to this, we have confirmed other effects such as later bookings producing shorter stays, or travelling outside the summer season also producing fewer days spent at destination.

Another important aspect to highlight, since it is one of the new features included in our research, is that tourists who declare to have done some activity at the destination generally stay longer and no moderating effects emerge for mostly of the activities. When these interactions exist, LCA users are not always the segment with the shortest trips, as it has some times been suggested. For instance, it seems that LCA users are the only ones to increase length of stay when attending a sporting event.

In summary, what this research has told us about air travellers is that the moderating effect of how the trip is booked is generally much more important than that of type of airline. If we compare package travel to choice of airline, many more variables substantially change their effect on length of stay with the former. The fact that LCAs are currently so consolidated appears to have resulted in them gradually capturing more different market segments, leading to increasing similarity between users of the two airline types. The stereotype that LCA users spend less time at their destination is becoming just that, a stereotype. Consequently, Destination Management Offices should not be any less interested in LCA travellers than in legacy company travellers.

REFERENCES

ALEGRE, Joaquín; MATEO, Sara & POU, Llorenç (2011) "A latent class approach to tourists' length of stay". Tourism Management 32, pp. 555-563.

ALEGRE, Joaquín & POU, Llorenç (2006) "The length of stay in the demand for tourism". Tourism Management 27, pp. 1343-1355.

BARROS, Carlos Pestana; BUTLER, Richard & CORREIA, Antónia (2010) "The length of stay of golf tourism: A survival analysis". Tourism Management 13, 13-21.

BARROS, Carlos Pestana; CORREIA, Antónia & CROUCH, Geoffrey (2008) "Determinants of the length of stay in Latin American tourism destinations". Tourism Analysis 13, 329-340.

BARROS, Carlos Pestana & MACHADO, Luiz Pinto (2010) "The length of stay in tourism". Annals of Tourism Research 37, pp. 692-706.

GOKOVALI, Uummuhan; BAHAR, Ozan & KOZAK, Metin (2007) "Determinants of length of stay: A practical use of survival analysis". Tourism Management 28, 736-746.

HONG, Sung-Kwon & JANG, Hochan (2005) "Factors influencing purchasing time of a new casino product and its managerial implications: an exploratory study". Journal of Travel Research 43, 395-403.

HOSMER, David W. & LEMESHOW Stanley (2000) "Applied Logistic Regression". 2nd ed. New York: Wiley.

MACHADO, Luiz Pinto (2010) "Does destination image influence the length of stay in a tourism destination?" Tourism Economics 16, 443-456.

MARTINEZ-GARCIA, Esther & RAYA, Josep Mª (2008) "Length of stay for low-cost tourism". Tourism Management 29, pp. 1064-1075.

MARTINEZ-GARCIA, Esther & RAYA, Josep Mª (2009) "Determinantes de la demanda temporal de turismo: una aproximación microeconómica con un modelo de duración". Investigaciones Económicas XXXIII, pp. 271-302.

MENEZES, António Gomez; MONIZ, Ana & VIEIRA, José Cabral (2008) "The determinants of length of stay of tourists in the Azores". Tourism Economics 14, pp. 205-222.

NICOLAU, Juan Luis & MÁS, Francisco Javier (2006) "Elección de la duración de las estancias turísticas: una aproximación con modelos de recuento". Revista Europea de Dirección y Economía de la Empresa 5, pp. 99-115.

NICOLAU, Juan Luis & MÁS, Francisco Javier (2009) "Simultaneous analysis of whether and how long to go on holidays". The Service Industries Journal 29, 1011-1092.

OPPERMAN, Martin (1995) "Travel life cycle". Annals of Tourism Research 22, 5353-552.

OPPERMAN, Martin (1997) "First-time and repeat visitors to New Zealand". Tourism Management 18, 177-181.

PEYPOCH, Nicolas; RANDRIAMBOARISON Rado; RASOAMANANJARA, Fy & SOLONANDRASANA, Bernardin (2012) "The length of stay of tourists in Madagascar". Tourism Management 33, 1230-1235.

RAYA-VILCHEZ, Josep Maria & MARTINEZ-GARCIA, Esther (2011) "Nationality and low-cost trip duration. A microeconomic analysis". Journal of Air Transport Management 17, 168-174

RODRIGUEZ, Santiago; DÁVILA, Delia & RODRIGUEZ, Alejandro (2003) "Determinantes de la demanda de turismo vacacional en las Islas Canarias". Investigaciones Regionales 3, pp. 81-98.

SEATON, Av & PALMER, Christine (1997). "Understanding VFR tourism behaviour: the first five years of the United Kingdom tourism survey". Tourism Management I, 345-355.

SUNG, Heidi H.; MORRISON, Alistair M.; HONG Gong-Soong & O'LEARY, Joseph T. (2001) "The effects of household and trip characteristics on trip types: a consumer behavioural approach for segmenting the US domestic leisure travel market". Journal of Hospitality & Tourism Research 25, 46-68.

THRANE, Christer (2012) "Analyzing tourists' length of stay at destinations with survival models: A constructive critique based on a case study". Tourism management 33, pp. 126-132.

YANG Yang; KEVIN K.F. Wong & JIE Zhang (2011) "Determinants of Length of stay for domestic Tourists: Case study of Yixing". Asia Pacific Journal of Tourism Management Research 16, pp. 619-633.

$$\beta_s = \frac{\beta}{\sqrt{\text{var}(y^*)}} = \frac{\beta}{\sqrt{\text{var}(\mathbf{x}\boldsymbol{\beta}) + \pi^2/3}}$$

$$x'_{id} = x_{id}\left(1 - \sum_{x_{idm}=0} x'_{id}\right)$$

$$f(y) = \sum_{j=1}^{k} \pi_j f(y|class=j)$$

$$\sum_{d=1}^{D} x_{id} = 1$$

$$= \ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right) = \ln(x_{i1}) - \frac{1}{2}\ln(x_{i2}) - \frac{1}{2}\ln(x_{i3})$$

$$N(\mu_j, \Sigma_j) \to N\left(\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,j} & \sigma_{12,j} \\ \sigma_{12,j} & \sigma^2_{2,j} \end{pmatrix}\right)$$

$$\sum_{d=1}^{D} x_{id} = 1$$

$$\text{with } \mathbf{y}|class = j \to N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \to N\left(\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,j} & \sigma_{12,j} \\ \sigma_{12,j} & \sigma^2_{2,j} \end{pmatrix}\right)$$

$$y_{i1} = \sqrt{\frac{2}{3}}\left(\ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right)\right) = \sqrt{\frac{2}{3}}\ln(x_{i1}) - \sqrt{\frac{1}{6}}\ln(x_{i2}) - \sqrt{\frac{1}{6}}\text{ln}$$

$$\mathbf{y} = \boldsymbol{\Psi}\ln(\mathbf{x})$$

$$y_{i1} = \ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right) = \ln(x_{i1}) - \frac{1}{2}\ln(x_{i2}) - \frac{1}{2}\ln(x_{i3})\beta$$

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + u_i$$

$$y_{id} = \ln\left(\frac{x_{id}}{\sqrt[D]{x_{i1}\cdot x_{i2}\cdot x_{i3}\cdots x_{iD}}}\right)$$

$$f(\mathbf{y}) = \sum_{j=1}^{k} \pi_j f(\mathbf{y}|class = j)$$

$$\beta_s = \frac{\beta}{\sqrt{\text{var}(y^*)}} = \frac{\beta}{\sqrt{\text{var}(\mathbf{x}\boldsymbol{\beta}) + \pi^2/3}}$$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \sqrt{2/3} & -\sqrt{1/6} & -\sqrt{1/6} \\ 0 & \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix}$$

$$\mathbf{y} = \boldsymbol{\Psi}\ln(\mathbf{x})$$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \sqrt{2/3} & -\sqrt{1/6} & -\sqrt{1/6} \\ 0 & \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix}$$

$$f(y) = \sum_{j=1}^{k} \pi_j f(y|class = j)$$

$$y_{i2} = \sqrt{\frac{1}{2}}\left(\ln\left(\frac{x_{i2}}{x_{i3}}\right)\right) = \sqrt{\frac{1}{2}}\ln(x_{i2}) - \sqrt{\frac{1}{2}}\ln(x_{i3})$$

$$\beta_s = \frac{\beta}{\sqrt{\text{var}(y^*)}} = \frac{\beta}{\sqrt{\text{var}(\mathbf{x}\boldsymbol{\beta}) + \pi^2/3}}$$

$$N(\mu_j, \Sigma_j) \to N\left(\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,j} & \sigma_{12,j} \\ \sigma_{12,j} & \sigma^2_{2,j} \end{pmatrix}\right)$$

$$\sum_{d=1}^{D} x_{id} = 1 \quad y_{i1} = \ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right) = \ln(x_{i1}) - \frac{1}{2}\ln(x_{i2}) - \frac{1}{2}\ln(x_{i3})$$

$$\text{with } \mathbf{y}|class = j \to N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \to N\left(\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,j} & \sigma_{12,j} \\ \sigma_{12,j} & \sigma^2_{2,j} \end{pmatrix}\right)$$

$$= \sqrt{\frac{2}{3}}\left(\ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right)\right) = \sqrt{\frac{2}{3}}\ln(x_{i1}) - \sqrt{\frac{1}{6}}\ln(x_{i2}) - \sqrt{\frac{1}{6}}\ln(x_{i3})$$

$$y_{i1} = \ln\left(\frac{x_{i1}}{\sqrt{x_{i2}\cdot x_{i3}}}\right) = \ln(x_{i1}) - \frac{1}{2}\ln(x_{i2}) - \frac{1}{2}\ln(x_{i3})$$

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + u_i$$

$$\mathbf{y} = \boldsymbol{\Psi}\ln(\mathbf{x})$$

$$x'_{id} = x_{id}\left(1 - \sum_{x_{idm}=0} x'_{id}\right)$$

$$\beta_s = \frac{\beta}{\sqrt{\text{var}(y^*)}} = \frac{\beta}{\sqrt{\text{var}(\mathbf{x}\boldsymbol{\beta}) + \pi^2/3}}$$

$$f(\mathbf{y}) = \sum_{j=1}^{k} \pi_j f(\mathbf{y}|class = j)$$

$$y_{id} = \ln\left(\frac{x_{id}}{\sqrt[D]{x_{i1}\cdot x_{i2}\cdot x_{i3}\cdots x_{iD}}}\right)$$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \sqrt{2/3} & -\sqrt{1/6} \\ 0 & \sqrt{1/2} & -\sqrt{1/6} \\ & & -\sqrt{1/2} \end{pmatrix}$$

$$f(y) = \sum_{j=1}^{k} \pi_j f(y|class = j)$$

$$= \sqrt{\frac{2}{3}}\ln(x_{i1}) - \sqrt{\frac{1}{6}}\ln(x_{i2}) + \sqrt{\frac{1}{6}}\ln(x_{i3})$$

$$\sum_{d=1}^{D} x_{id} = 1$$

$$\mathbf{y} = \boldsymbol{\Psi}\ln(\mathbf{x})$$

$$\begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix}$$

$$= \sqrt{\frac{1}{2}}\left(\ln\left(\frac{x_{i2}}{x_{i3}}\right)\right) = \sqrt{\frac{1}{2}}\ln(x_{i2}) - \sqrt{\frac{1}{2}}\ln(x_{i3})$$

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \sqrt{2/3} & -\sqrt{1/6} & -\sqrt{1/6} \\ 0 & \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix}$$