

# Determinants in tourist expenditure composition – the role of airline types

BERTA FERRER-ROSELL

*Department of Economics, Faculty of Tourism, University of Girona, Edifici Sant Domènec, Plaça Ferrater Mora 1, 17071 Girona, Spain. E-mail: berta.ferrer@udg.edu.  
(Corresponding author.)*

GERMÀ COENDERS

*Department of Economics, Faculty of Economics and Business, University of Girona, Campus Montilivi, 17071 Girona, Spain. E-mail: germa.coenders@udg.edu.*

ESTHER MARTÍNEZ-GARCIA

*Department of Economics, Faculty of Tourism, University of Girona, Edifici Sant Domènec, Plaça Ferrater Mora 1, 17071 Girona, Spain. E-mail: esther.martinez@udg.edu.*

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. The authors 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 travellers' attributes as explanatory factors along with the moderating effect of the airline type. Among the findings are that high-income LCA travellers spend relatively more at the destination, LCA tourists

The authors are pleased to acknowledge the support of the Spanish Institute of Tourism Studies (*Instituto de Estudios Turísticos*) in providing them with the raw data. Any lack of accuracy or reliability in the data analysis is the authors' responsibility. The first author carried out this research with the support of a grant from the University of Girona (BR 18/2011). The authors also wish to thank the members of the Research Group 'Statistics and Data Analysis' and the 'Research Group on Compositional Data Analysis and Other Restricted Sample Spaces, CODA-RSS' at the University of Girona, Juan Luis Nicolau at the University of Alicante, and Jorge González Chapela at the University of Girona for their helpful comments on previous versions of this article.

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.

*Keywords:* tourism demand; tourist expenditure; low-cost airline; compositional data analysis (CODA)

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

The cheaper fares offered by LCAs result in a significant reduction in transportation costs, which is expected to modify the composition of the trip's budget (Martínez-García and Raya, 2008). This seems to have been the case in Spain, a major world tourist destination: 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 (O'Connell and Williams, 2005; Chiou and Chen, 2010; Forgas *et al.*, 2010; Ferrer-Rosell *et al.*, 2014), or on LCA demand and its heterogeneity (Castillo-Manzano and Marchena-Gómez, 2010; Martínez-García and Royo-Vela, 2010; Kim and Lee, 2011; Raya-Vilchez and Martínez-García, 2011; Martínez-García *et al.*, 2012). However, this growing literature on LCA demand and comparisons with legacy airline demand has not, to our knowledge, focused on tourist expenditure and its composition.

Expenditure analyses have been quite frequent in the tourism literature, since they are of major concern for destination management offices (DMO), marketers, tourist agencies, and in general to all those engaged in tourism. Tourist expenditure, rather than the number 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 the type of tourist classified by how they distribute their travel budget. The travel budget can be broken down, as with household budget studies, between non-discretionary and discretionary components. Non-discretionary components are accommodation, transportation and food, where a minimum amount has to be spent, whereas discretionary expenditure includes

extra activities, shopping and so on. Depending on personal, economic, trip and socio-demographic characteristics, tourists may be more or less willing to embark on visits, activities, excursions, shopping and so on and thus the proportion of discretionary tourist budget will change accordingly. This distinction between expenditure components is relevant for destinations as they are naturally 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 of their budget on that component, 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 versus 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 LCAs 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, that is, 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. Shares 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. Aitchison's (1986) seminal work 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 has not only 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 increase only 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 are 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 (Downward and Lumsdon, 2000, 2003; Cannon and Ford, 2002) to advanced econometric techniques. For instance, Hung *et al* (2012) use quantile regression 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. With respect to 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 (Wang *et al*, 2006; Sainaghi, 2012). Nicolau (2009) includes individual price sensitivity estimated in a previous model.

Another stream of research is that which analyses tourist expenditure per tourism product (for example, lodging, food, transportation and sightseeing/entertainment). A common argument for studying tourist expenditure patterns per tourism product 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 (Pyo *et al*, 1991; Cai *et al*, 1995; Oppermann, 1996; Cai, 1998, 1999; Lee, 2001; Lehto *et al*, 2001; Jang *et al*, 2004; Wang *et al*, 2006). Socio-demographic variables are the most common

significant variables. Age affects expenditure on 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 roughly 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 analysing budget share.

The empirical analysis of budget share, both for tourism expenditure and general family budgets, commonly involves 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 shares (composition) as dependent variables in a set of simultaneous regressions. The approach has focused mainly on estimating price and income elasticities. Studies that use macroeconomic data include, among others, those 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 has been applied both to data from a given origin to multiple destinations (that is, *ex ante* – see Li *et al*, 2004; Divisekera, 2009) or from multiple origins to a given destination, which is the case in our article (*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 weakly 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, and so on. *Ex post* studies, such as ours, model only the last stage. There are, however, 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 Fleischer *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. Fleischer *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. Fleischer *et al* (2011) are the only researchers 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 this article are similar to those of Coenen and van Eekeren (2003) and of Fleischer *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  $i$   $x_{i1}, x_{i2}, \dots, x_{iD}$  has the following constraints:

$$0 \leq x_{id} \leq 1 \text{ and } \sum_{d=1}^D x_{id} = 1. \quad (1)$$

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 heteroscedastic. One component can increase only 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 (for example, 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 involves using the transformed share by means of logarithms of ratios, instead of the raw share.

To ensure compositional coherence of predicted budget share (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 *et al*, 1996; Fry, 2011). 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; Ronning, 1992; Fry *et al*, 1996). However, the review in Fry (2011) reports few studies that have applied the CODA methodology to demand equations and to the study of household budgets (McLaren *et al*, 1995; Fry *et al*, 1996, 2000).

When fitting demand equations to log-ratios of expenditure components, Engel's aggregation condition, which requires that 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 homoscedastic error terms.

Related developments are the indirect addilog system (Houthakker, 1960) and the generalized addilog system (Bewley, 1982). When applied to compositions (for example, 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:

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

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

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

All log-ratio transformed  $y_{id}$  variables recover the full unconstrained  $-\infty$  to  $\infty$  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 (Fry *et al*, 1996). Conversely, the clr transformation is commonly used for statistical techniques which are based on a metric, such as 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 (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 decidedly 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, such as postcard shopping, local bus tickets and going to a museum (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, first, that there is basically no tourist type who will never spend anything on activities and, second, 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 used the same zero replacement strategy that was later suggested by Martín-Fernández *et al* (2003), namely replacing  $x_{id} = 0$  with:

$$x'_{id} = k\delta_{id} \text{ with } 0 < k < 1, \quad (4)$$

where  $\delta_{id}$  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:

$$x'_{id} = x_{id} (1 - \sum_{x_{id}=0} x'_{id}). \quad (5)$$

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 €1, 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  $\delta_{id}$  by dividing €1 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) and  $x_{i3}$  = activities and shopping (discretionary expenditure). An interpretable sequential partition is shown in 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.

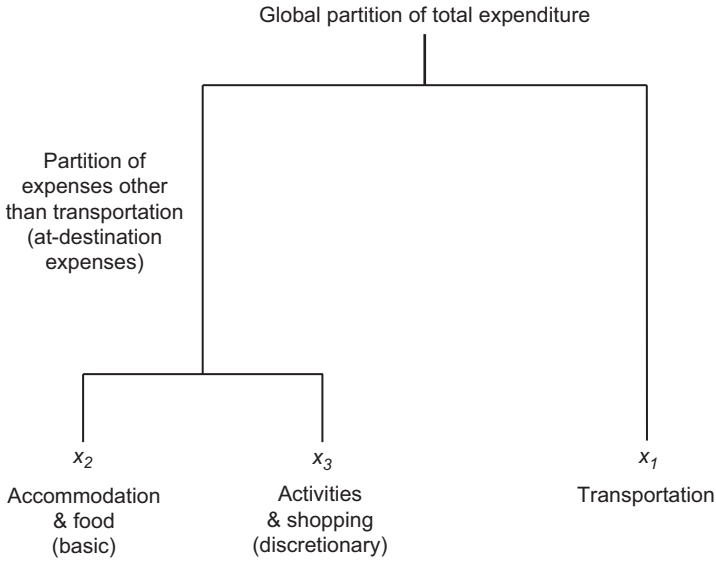


Figure 1. Sequential partition of total expenditure.

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

Positive values show a transportation 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 to what tourists allocate the rest of their trip budget, and whether more is allocated to basic or discretionary expenditure, once they have paid for (often in advance) their transportation.

$$y_{i2} = \ln \left( \frac{x_{i2}}{x_{i3}} \right) = \ln(x_{i2}) - \ln(x_{i3}). \quad (7)$$

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 (MANOVA) models for categorical explanatory variables are appropriate. MANOVA's multivariate tests and statistics (for example,

Pillai's trace, Hotelling's trace or Wilk's lambda) are invariant to how components are arranged in Figure 1 to compute log-ratios (Mateu-Figueras *et al.*, 2013). 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 converse. Socio-demographic characteristics (traveller's age, gender, education, income, etc) are obviously determined prior to any travel decision and hence exogenous with respect to expenditure.

Many of the choices tourists make are interdependent (Dellaert *et al.*, 1997) or at least planned simultaneously (Fesenmaier and Jeng, 2000). So, expenditure is arguably decided on at the same time as many trip attributes. At least, it is unclear whether expenditure is consistently decided on after trip attributes 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 one 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, are 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);
- who do not decide how much they will spend on certain components (business and study trips);
- who do not pay for the trip themselves (trips paid for by employers, family/friends, prizes, offers, etc);
- 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 database and used in the model as budget components we, first, put the amount paid for transportation ( $x_1$ ). This component has no zeros.

Second, the amount of money paid for accommodation and food is undistinguishable for full-board, half-board or bed and breakfast accommodation. 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. 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 constructed as a combination of the motivation and the destination variable. We distinguish between, first, those tourists travelling for cultural tourism to singular cities, second, those who come just for leisure, either in the countryside, or more commonly at the seaside and, third, other typologies (Table 3).

Table 1. Percentage share and log-ratio descriptive statistics.

	Min	Max	Mean	SD	Skewness	Kurtosis
Transportation component ( $x_1$ )	0.03	96.11	28.46	13.87	0.837	0.789
Basic component ( $x_2$ )	0.21	99.63	54.72	16.13	-0.114	-0.450
Discretionary component ( $x_3$ )	0.01	86.68	16.81	12.92	0.988	1.174
Transportation/at-destination log-ratio ( $y_1$ )	-7.57	5.48	0.1506	1.0702	0.714	1.404
Basic/discretionary log-ratio ( $y_2$ )	-5.30	8.72	1.7657	1.8173	1.573	2.151

## Results

### *Results for the exogenous variables*

Table 4 presents the MANOVA results. The univariate  $R^2$ , corrected for the degrees of freedom, is 0.045 (first log-ratio) and 0.105 (second log-ratio). The multivariate uncorrected  $R^2$  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 four 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; that is, the predicted log-ratio for each type of airline within the reference category for all variables (university education, medium income, resident in the 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 expenditure on 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.

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

Variables	Count	Percentage
Level of education		
Up to high school	6,019	31.1
University <sup>a</sup>	13,340	68.9
Income category		
Medium/low	797	4.1
Medium <sup>a</sup>	13,186	68.1
Medium/high	4,376	22.6
High	1,000	5.2
Country of residence		
Portugal	624	3.2
Other European countries	850	4.4
Belgium	1,006	5.2
Netherlands	1,069	5.5
Austria, Switzerland and Liechtenstein	1,122	5.8
France	1,326	6.8
Scandinavian countries	1,789	9.2
Germany	2,363	12.2
Italy	3,113	18.1
United Kingdom and Ireland <sup>a</sup>	6,097	31.5
Gender		
Female	8,530	44.1
Male <sup>a</sup>	10,829	55.9
Travel group		
Alone	2,484	12.8
In family	2,980	15.4
With friends	4,440	22.9
With partner <sup>a</sup>	9,446	48.8
Professional status		
Homemaker	517	2.7
Unemployed	602	3.1
Student	1,432	7.4
Low-level employee	915	4.7
Mid-level employee <sup>a</sup>	10,964	58.6
High-level employee	2,138	11.0
Self-employed	2,791	14.4
Age		
Over 45 years' old and pensioner	1,484	7.7
Over 45 years' old and not pensioner	5,009	25.9
25–44 years' old <sup>a</sup>	10,569	54.6
15–24 years' old	2,297	11.9
Type of airline (moderating variable)		
Legacy	5,884	30.4
Low cost	13,475	69.6

*Note:* <sup>a</sup>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	Percentage
<i>Activities</i>		
<i>Golf</i>		
Yes	227	1.2
No	19,132	98.8
<i>Hiking</i>		
Yes	364	1.9
No	18,995	98.1
<i>Sporting events</i>		
Yes	603	3.1
No	18,756	96.9
<i>Nautical sports</i>		
Yes	1,645	8.5
No	17,714	91.5
<i>Other sports</i>		
Yes	904	4.7
No	18,455	95.3
<i>Cultural visits</i>		
Yes	11,180	57.8
No	8,179	42.2
<i>Cultural events</i>		
Yes	2,452	12.7
No	16,907	87.3
<i>Other cultural activities</i>		
Yes	4,175	21.6
No	15,184	78.4
<i>Spa</i>		
Yes	1,199	6.2
No	18,160	93.8
<i>Theme parks</i>		
Yes	1,745	9.0
No	17,614	91.0
<i>Gastronomy</i>		
Yes	1,565	8.1
No	17,794	91.9
<i>Type of accommodation</i>		
Hotel 4–5*	5,056	26.1
Hotel 3*	4,922	25.4
Hotel < 3*	6,673	34.5
Rented apartment	2,116	10.9
Other accommodation	592	3.1
<i>Length of stay</i>		
5 or less nights	9,911	51.2
6–8 nights	5,952	30.7
9–12 nights	1,592	8.2
13–15 nights	1,261	6.5
16 and more nights	643	3.3

Table 3 continued.

Variables	Count	Percentage
<i>Total expenditure</i>		
Quartile 1	5,074	26.2
Quartile 2	5,104	26.4
Quartile 3	4,986	25.8
Quartile 4	4,195	21.7
<i>Expenditure at destination per day</i>		
Quartile 1	4,855	25.1
Quartile 2	4,832	25.0
Quartile 3	4,868	25.1
Quartile 4	4,804	24.8
<i>Motivation-destination</i>		
Leisure on coast/in countryside	8,083	41.8
Urban tourism	5,324	27.5
Others	5,952	30.7

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, the 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, the French, Italians and Portuguese have the highest discretionary share within LCA users, and other European countries (mainly central and 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.

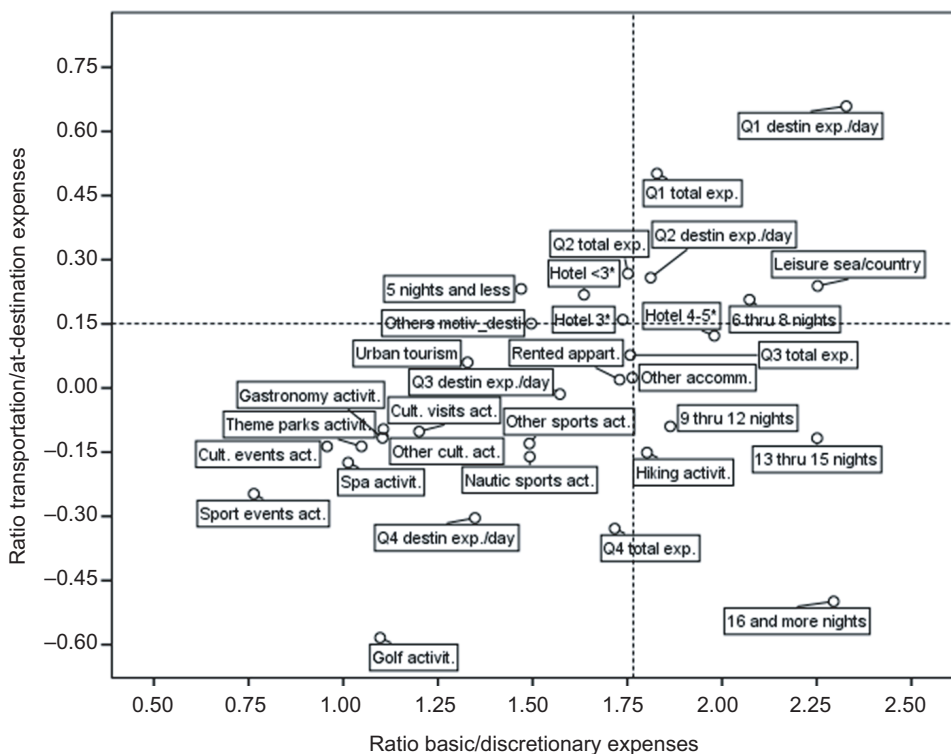
The travel group seems to affect both log-ratios, and has a moderating effect with airline type. Those tourists who travel in a family group have a higher share in transport expenses when flying with legacy airlines. In the second log-ratio, travelling as a family with legacy airlines increases the basic 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, the results show that it mostly affects the basic/discretionary log-ratio, and there are some relevant airline type

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	-0.269	-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 Countries	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 + Switzerland + Liechtenstein	-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

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



**Figure 2.** Endogenous variables on the log-ratio space. Mean log-ratio values within each category of the endogenous variables.

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.

The 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.

#### *Results for the endogenous variables*

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 the star rating 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 destination and motivation, tourists going to a seaside or countryside destination for leisure purposes behave quite differently from those going to an urban destination for urban or cultural tourism. The former group spends 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 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 with a higher at-destination expenditure per day, spend 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.

## Conclusion 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 occupations have different effects 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 with LCAs tend to spend more at the destination, and more specifically on discretionary expenditure. LCA tourists residing in the 'other European countries' spend more on transportation. Families flying with legacy airlines spend more on the transportation share compared to the at-destination share,

and also spend 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 the objectives of DMOs are to increase the allocation to the at-destination component of the trip budget (that is, 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 (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 markets other than the UK and Ireland. Young tourists and tourists who travel with friends (where airline type is inconsequential), 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  $\ln$  and  $\ln$  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 two limitations. First, contrary to studies that use macro data, in micro data studies the coexistence of full-board, half board, bed and breakfast and accommodation-only tourists makes it impossible to separate accommodation and food expenditure meaningfully. Other subdivisions (for example, 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. Second, 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.

## 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, L.L. (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 generalized 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), 'Analysing 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, No 3, 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. (1997), '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.
- Downard, 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-García, 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/anales/Informe%20Compa%C3%B1a%20ADas%20Bajo%20Coste%202004.pdf>, accessed 14 January 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/anales/Informe%20Compa%C3%B1a%20ADas%20Bajo%20Coste%202010.pdf>, accessed 14 January 2013).
- IET (2012), *Egatur. Encuesta de Gasto Turístico. Metodología*, Instituto de Estudios Turísticos, Ministerio de Industria, Energía y Turismo, Madrid (<http://www.iet.tourspain.es/es-ES/estadisticas/egatur/metodologia/Referencia%20Metodologica/Nota%20Metodol%C3%B3gica%20Encuesta%20de%20Gasto%20Tur%C3%ADstico.pdf>, accessed 22 November 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.
- Martínez-García, E., and Raya, J.M. (2008), 'Length of stay for low-cost tourism', *Tourism Management*, Vol 29, No 6, pp 1064–1075.
- Martínez-García, 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.
- Martínez-García, 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.
- 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, NY, pp 47–62.
- Mateu-Figueras, G., Daunis-i-Estadella, J., and Martín-Fernández, J.A. (2013), 'Analysing grouped CoDa: Manova and post-hoc contrasts', paper presented at the workshop on Compositional Data Analysis (CoDaWork), Vorau, 3–7 June.
- McLaren K.R., Fry J.M., and Fry, T.R.L. (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, NY.
- 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-García, E. (2011), 'Nationality and low-cost trip duration. A microeconomic 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), 'Analysing tourism consumption: a dynamic system-of-equations approach', *Journal of Travel Research*, Vol 50, No 1, pp 46–56.

# 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.
```