Understanding low-cost airline users' expenditure patterns and volume

BERTA FERRER-ROSELL

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

GLÒRIA MATEU-FIGUERAS

Department of Computer Science, Applied Mathematics and Statistics, University of Girona, Campus Montilivi, 17071 Girona, Spain. E-mail: gloria.mateu@udg.edu

VERA PAWLOWSKY-GLAHN

Department of Computer Science, Applied Mathematics and Statistics, University of Girona, Campus Montilivi, 17071 Girona, Spain. E-mail: vera.pawlowsky@udg.edu.

The expansion of low-cost airlines (LCAs) has led to a reduction of flight costs and has enabled tourists to redistribute their trip budget and spend a higher proportion at the destination. Analyses of absolute trip spending by parts (such as transportation) and of trip budget share (for example, the proportion of transportation within the total trip budget) serve different objectives. The first type of analysis refers to how much tourists spend and the second to how they spend. The aims of this article are, first, to provide a new methodological tool to combine the analysis of budget share and absolute expenditure and, second, to analyse the determinants of absolute expenditure on transportation and the relative importance

of transportation expenditure compared to other trip budget compo-
nents. The main findings are that traveller characteristics affecting
high absolute transportation expenditure are different from those
affecting a high share of transportation in the trip budget. This has
implications for LCA pricing strategies.

The reduction in flight costs resulting from the increased market of low-cost
airlines (LCAs) has had a twofold effect. First, it has made air travel available
to all budgets (Mason and Alamdari, 2007; Dobruszkes, 2013; Graham, 2013).
Second, it has affected expenditure patterns by enabling tourists to redistribute
their expenditure and spend a higher proportion of their trip budget at the
destination, for instance, on activities (Martínez-Garcia and Raya, 2008). Rel-
evant research questions thus include the determinants of transportation ex-
penditure both in absolute terms (Wang *et al*, 2006; Kim *et al*, 2010; Marcussen,
2011; Zheng and Zhang, 2013) and relative to the total trip budget volume
or to other trip budget parts (Ferrer-Rosell *et al*, 2015).

Beyond the trivial analysis of one single overall expenditure variable, up to
now available statistical techniques have made it possible to study either:

- Absolute trip expenditure on each trip budget part (expenditure on trans-
  portation, accommodation, activities, shopping, and so on). Since part ex-
  penditure in absolute terms belongs to total expenditure, a repeated finding
  is that some variables affect all budget parts in the same direction: for
  example, Marcussen (2011) found that hotel accommodation, eating in
  restaurants and staying at the capital city led to higher expenditure in all
  trip budget parts. Furthermore, a large absolute expenditure on a given
  budget part may have a dual interpretation: it may correspond either to a
  tourist with a large overall budget or to a tourist who devotes a large share
  to that part.
- Trip budget composition, or in other words the share of the trip budget
  allocated to the budget parts (proportion of the total trip budget spent on
  transportation, accommodation, activities, shopping and so on). Variables
  affecting in the same direction all budget parts are never obtained when
  analysing expenditure allocation in relative terms. Share devoted to one
  budget part can only increase if share devoted to at least one other part
  decreases. The study of trip budget share with compositional data analysis
  (CODA) isolates the effects on how a given trip budget is allocated (Ferrer-
  Rosell *et al*, 2015). On the negative side, information about absolute ex-
  penditure (how much tourists spend) is lost.

The same absolute amount spent on transportation can, thus, hide different
budget distributions, and the same budget share can hide different cardinal
amounts spent on transportation. Both phenomena can only be disentangled
when absolute expenditure on transportation is analysed together with trip
budget composition. The objective of this article is twofold.

- To provide the needed new methodological tools to combine the analysis of

budget share (how tourists distribute expenditure) and absolute expenditure (how much tourists spend) while distinguishing the effects of exogenous variables on budget distribution from their effects on absolute expenditure.

- To analyse the traveller characteristics affecting absolute expenditure on transportation and those affecting the relative importance of transportation expenditure compared to accommodation and activities for tourists arriving in Spain by LCAs. Spain is an ideal case to study this phenomenon, being ranked as the third 2014 destination in the world by the World Tourism Organization (WTO), and given the fact that in 2012, which is when the data in this article were collected, the majority of foreign tourists arrived by air (80% according to ITE (2013a)) and LCAs dominated the air market (58% according to ITE (2013b)).

The article is structured as follows. It first presents a literature review including both the relevant methodology for analysing multiple expenditure variables, and applied research on transportation expenditure. It next describes the CODA approach and the existent possibilities to combine budget composition and absolute expenditure. It subsequently provides the new proposed extensions to the method needed for the second objective of the article, whose results in the Spanish LCA user case are then presented. The last section discusses the results and concludes.

Literature review

Microeconometric analysis of individual tourist spending as a function of traveller characteristics has focused on a single expenditure variable, on absolute expenditure by trip budget parts and on expenditure in relative terms (share) by trip budget parts. The vast majority of microeconomic tourist expenditure studies (24 out of 27 in the review of Wang and Davidson (2010); and 64 out of 77 in the review of Brida and Scuderi (2013)) takes into account one single expenditure variable.

Another stream of research is that which analyses absolute tourist expenditure by trip budget parts (for example, the amount spent on lodging, food, transportation and sightseeing/entertainment). A common argument for studying tourist expenditure by budget parts is that it provides vital information to travel organizers and destination marketers. Not all budget parts have the same economic impact on each destination or on each tourism industry (Hadjikakou *et al,* 2014). Besides, knowing expenditure by tourist product is useful when designing appropriate marketing strategies. For this purpose, researchers have used several methods. Within the least squares family (Lehto *et al*, 2001; Wang *et al*, 2006; Marcussen, 2011) the most common is multivariate analysis of variance (MANOVA) when predictors are qualitative. When predictors are numeric or include a combination of both types, researchers may use multivariate analysis of covariance (MANCOVA) or seemingly unrelated linear regressions (SUR), although running separate ordinary least squares regressions for each budget part often leads to the same results (Thrane, in press). Beyond least squares methods, Tobit is the dominant approach (Hong *et al*, 1996; Barquet *et al,* 2011; Kim *et al,* 2011; Brida *et al,* 2012; Zheng and Zhang, 2013). Some authors compared Tobit and least squares (Kim *et al*, 2010).

A not uncommon finding in the studies of absolute expenditure by budget parts is that some of the explanatory variables affect all budget components in the same direction. For example, Hong *et al* (1996) found that higher levels of education led to higher expenditure in all budget parts, Marcussen (2011) concluded similarly about hotel accommodation, eating in restaurants and staying at the capital city, Lehto *et al* (2001) found the same for certain countries of origin and Zheng and Zhang (2013) for tourists who were white and married, had higher education and high employment status, and travelled in summer.

The empirical analysis of budget share commonly implies estimating an almost ideal demand system of equations (Deaton and Muellbauer, 1980). With this system, effects of one variable in the same direction on all budget components are impossible to obtain; more precisely the effects of a variable on all components add up to zero. It thus properly reflects the trade-offs between parts that are always encountered when analysing share. This system directly fits trip budget share as dependent variables in a set of simultaneous regressions and is gaining popularity in macroeconomic tourism demand analysis (Song *et al*, 2012). A few notable microeconomic applications are also to be found in the literature (Coenen and van Eekeren, 2003; Fleischer *et al*, 2011; Chang *et al*, 2013; Lee *et al*, 2015).

The almost ideal demand system has a serious methodological drawback. Let x_{i1}, $x_{i2}, \ldots, x_{iD}$, be the absolute expenditure of individual i in each of the D budget parts. Compared to absolute data, budget share lies in a constrained D–1 dimensional space. A D-term budget share measured on individual i z_{i1}, $z_{i2}, \ldots, z_{iD}$ is constrained to be non-negative and add up to one. This conforms to what the literature refers to as compositional data. Aitchison (1986), Pawlowsky-Glahn and Buccianti (2011) and Pawlowsky-Glahn *et al* (2015b) warn against the serious problems that arise when using standard statistical analysis tools on compositional data. Compositional data are bounded and heteroscedastic (close to the unit boundary the variance can only be lower). The presence of an error term with an unbounded distribution (for example, normal) results in a non-zero probability that actual share lies outside the [0,1] interval (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. Ferrer-Rosell *et al* (2015) report as many as 84% of cases with invalid prediction intervals when fitting share directly to statistical models as the almost ideal demand system does.

The most common approach in compositional data analysis (CODA) is the log-ratio approach, which involves transforming compositional data so that they can be subject to standard and well-understood statistical techniques (Aitchison, 1986; McLaren *et al*, 1995; Fry *et al*, 1996; Ferrer-Rosell *et al*, 2015, 2016). In short, this involves using the transformed share by means of logarithms of ratios, instead of the raw share. Log-ratios recover the full unconstrained $-\infty$ to ∞ range. The CODA methodology has started to be applied to trip budgets (Ferrer-Rosell *et al*, 2015, 2016). Related developments are the indirect addilog system (Houthakker, 1960) and the generalized addilog system (Bewley, 1982).

Regarding applied research on traveller characteristics and absolute transportation expenditure, Kim *et al* (2010) focused on attendants at a local festival who mostly travelled by car from short distances. Among studies including

long-distance travel, Marcussen (2011) analysed transportation expenses on a per-night basis and found that small travel groups, older tourists, repeat visitors, males and high income earners were associated with higher absolute expenditure on transportation, Wang *et al* (2006) also on a per-night basis reached the same conclusions about travel group size and Zheng and Zhang (2013) on a per-trip basis found that high level employment, high education and high income led to a higher absolute transportation expenditure. Whereas, Fleischer *et al* (2011) focused on the relative importance of transportation expenditure and found that being born in the country of origin increased the transportation expenditure share. Ferrer-Rosell *et al* (2015) also focused on transportation share and did so within LCA users. They found that tourists with low education, low income, from the UK and Ireland, travelling in family and homemakers allocated a higher portion of their trip budget to transportation. Lee *et al* (2015), in a study on event tourism, found that repeat attendance to the event and geographical distance from the place of origin increased the transportation expenditure share.

Methodology

This article contributes with a method to predict trip budget share and trip budget volume within the same MANOVA or MANCOVA model. The suggested approach is an extension of that of Pawlowsky-Glahn *et al* (2015a) combining a total with the log-ratios. The article presents new possibilities of such combination, and derives the properties of a MANOVA or MANCOVA model. With these combinations, meaningful hypotheses about trip budget volume and allocation can be tailored to the researchers' own questions and statistically tested. Once share and total volume have been transformed, the analysis is no more complicated than standard MANOVA or MANCOVA, and any standard software can be used.

Usual approaches to the analysis of absolute values by parts

Economic analysis of trip expenditure by budget parts within the least squares framework usually considers logarithms of absolute expenditure values of the *i*th individual (x_i) by each of the D budget parts as dependent variables:

$$\ln(x_{i1}), \ \ln(x_{i2}) \ ,\ldots, \ \ln(x_{iD}). \tag{1}$$

The use of logarithms is attractive for economic analysis of expenditure for a number of reasons (Thrane, 2014) and becomes a must when it is to be combined with allocation and log-ratios, as shown below.

As an alternative to modelling absolute trip expenditure by budget parts, trip budget composition can be analysed together with an absolute expenditure total. As far as the composition is concerned, several log-ratio transformations have been suggested in the CODA literature (Aitchison, 1982; Egozcue *et al*, 2003). Any log-ratio may be computed either from absolute expenditure (x_i) or from share (z_i). The centred log-ratio transformation (clr) computes the log-ratios y_{id} of each *d*th component over the geometric mean of all D components, including itself:

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

The clr transformation makes Euclidean distances in the transformed space equivalent to the standard distance in compositional data (Aitchison *et al*, 2000). Thus, the clr transformation is commonly used for statistical techniques that are based on a metric, such as cluster analysis (Van Der Boogaart and Tolosana-Delgado, 2006; Martín-Fernández *et al*, 2015a).

As far as including the total is concerned, Pawlowsky-Glahn *et al* (2015a) study the properties of the space defined by a composition and a total and conclude that the centred log-ratio transformation (clr) together with a total computed as $\sqrt{D}$ times the logarithm of the geometric mean of all absolute values leads to the same Euclidean distances among individuals as the variables in Equation (1):

$$\ln\left(\frac{x_{i1}}{\sqrt[D]{x_{i1}x_{i2}\cdots x_{iD}}}\right),\ln\left(\frac{x_{i2}}{\sqrt[D]{x_{i1}x_{i2}\cdots x_{iD}}}\right),\cdots,\ln\left(\frac{x_{iD}}{\sqrt[D]{x_{i1}x_{i2}\cdots x_{iD}}}\right),\sqrt{D}\ln\left(\sqrt[D]{x_{i1}x_{i2}\cdots x_{iD}}\right). \qquad (3)$$

The last term in Equation (3) includes information on how much tourists spend, while the first D terms with the clr log ratios include information on how they distribute their trip budget. Distances between any two individuals can be large if either their log-ratios are different or if their total is different. As opposed to Equation (1), only one of the terms in Equation (3) – the last – contains absolute information.

On the negative side, it must be noted that the clr transformation computes D log-ratios for a composition lying in a D–1 dimensional space. Consequently one log-ratio is a linear combination of the remaining. Thus, the clr is not appropriate for modelling because it leads to a singular covariance matrix.

Egozcue *et al* (2003) and Egozcue and Pawlowsky-Glahn (2005) introduced the isometric log-ratio transformation (ilr) which leads to the same distances as the clr transformation but uses only D-1 log-ratios, so that covariance matrices can be inverted. It can thus be used in all standard statistical analyses (Mateu-Figueras *et al*, 2011). Contrary to the clr, this alternative transformation is more flexible in that the denominator is not the same in all log-ratios. For this reason, ilr log-ratios can easily be tailored to economically meaningful research questions and hypotheses and suited to the needs of each particular tourism expenditure research. Thus, unlike the clr, the ilr does not have a unique algebraic expression but its computation will depend on each case according to the research questions. The next subsection gives the requirements an ilr transformation has to fulfil; while the 'chosen log-ratios and absolute expenditure' section provides an example in Equations (10) and (11), and simple guidelines to tailor the transformation to one's own research interests with a simple graphical tool.

Pawlowsky-Glahn *et al* (2015a) show that an ilr transformation can be substituted into Equation (3) with identical results. This follows their approach. The additional contribution of this article is twofold: to extend it to MANOVA and MANCOVA modelling and to consider alternatives to the total with the

purpose of including absolute information, which can also be tailored to meaningful research questions.

MANOVA and MANCOVA of log-ratios with alternatives to the total expenditure

Pawlowsky-Glahn *et al* (2015a) have already pointed to the fact that alternative totals may make sense for particular purposes, once their statistical properties have been derived. This article shows that for MANOVA and MANCOVA, modelling ilr log-ratios and the logarithm of the raw geometric mean of the absolute expenditure in any number of budget parts is equivalent to modelling the logarithms in Equation (1). Besides, the rather hard-to-interpret constant $\sqrt{D}$ in Equation (3) is not needed. Equivalence stems from the fact that MANOVA and MANCOVA results are preserved under change of basis (Martín-Fernández *et al*, 2015b).

This makes it possible to build tailor-made log-ratios chosen by the researcher to be interpretable with respect to his or her particular research objectives or questions (in this article, a log-ratio of transportation expenditure over the geometric mean of at-destination expenditure to find the drivers of the allocation of a part of the budget on transportation) and to focus on tailor-made absolute expenditure, which is more relevant to the problem at hand (in this article, the log absolute expenditure on transportation).

If log-ratios and absolute expenditure are chosen following the guidelines above, the transformed data (y_{i1} to y_{iD-1} log-ratios and t_i absolute expenditure) and the logarithms of all absolute expenditure in Equation (1) are related through the Ψ transformation matrix:

$$
\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iD-1} \\ t_i \end{pmatrix} = \Psi \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \vdots \\ \ln(x_{iD-1}) \\ \ln(x_{iD}) \end{pmatrix}. \tag{4}
$$

If, for instance, there are three expenditure components and t_i is computed as the logarithm of absolute expenditure on the first component x_{i1}, then the transformed data can be the following function of the logarithms of all absolute expenditure in Equation (1):

$$
\begin{pmatrix} y_{i1} \\ y_{i2} \\ t_i \end{pmatrix} = \Psi \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix} = \begin{pmatrix} \sqrt{\dfrac{2}{3}} & -\sqrt{\dfrac{1}{6}} & -\sqrt{\dfrac{1}{6}} \\ 0 & \sqrt{\dfrac{1}{2}} & -\sqrt{\dfrac{1}{2}} \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ln(x_{i1}) \\ \ln(x_{i2}) \\ \ln(x_{i3}) \end{pmatrix}. \tag{5}
$$

In Equation (5), Ψ contains the same ilr transformation and absolute expendi-

ture used in this article (see 'chosen log-ratios and absolute expenditure' section).

The first D–1 rows of $\boldsymbol{\Psi}$ contain information on how tourists distribute expenditure and must fulfil the requirements of any ilr transformation (Egozcue *et al,* 2003):

- each row has zero sum;
- each row has unit sum of squares;
- each pair of rows have zero scalar products.

The last row of $\boldsymbol{\Psi}$ refers to the chosen absolute expenditure (t_i) understood as the geometric mean of one or more absolute expenditure by parts. It can only contain zero values and one or more equal positive values adding up to 1.

These are sufficient conditions for the whole $\boldsymbol{\Psi}$ matrix being a transformation matrix between the space in Equation (1) and the space generated by Equation (3). From this and from Martín-Fernández *et al* (2015b), it follows that MANOVA and MANCOVA multivariate tests and statistics (for example, Pillai's trace, Hotelling's trace or Wilk's Lambda) are invariant to how ilr log-ratios are computed and to the choice of which components are in the geometric mean that is used as absolute expenditure. These tests answer the question whether an explanatory variable is relevant to expenditure, all things considered (share and/or volume).

Univariate tests referring to the absolute expenditure and to each particular log-ratio are not invariant to the choice. If the log-ratios and absolute expenditure have been selected with respect to particular and meaningful research questions, these tests can be directly related to them. In this article, such tests can show a significant relationship between a predictor and the relative or absolute importance of transportation expenditure. This is why the interpretability of each log-ratio and of the selected absolute expenditure is important.

Regarding the univariate tests, separate ANOVA and separate univariate ordinary least squares regression models for each log-ratio and for the absolute expenditure will yield the same results as MANOVA and MANCOVA (Thrane, in press). The main strength of the multivariate approach is precisely the multivariate tests. When the number of variables is large, which is often the case in tourist expenditure research (Thrane, 2014), the number of univariate tests grows large and leads to type-I risk inflation. Performing many tests leads to a higher probability of at least one true null hypothesis being rejected. Multivariate tests reduce error inflation because there is a reduced number of them, and because univariate tests are interpreted only for variables that are statistically significant in the multivariate tests.

Chosen log-ratios and absolute expenditure

This article uses the same ilr transformation implied in Equation (5), with an emphasis on interpretability with respect to the tourism researchers' own questions or hypotheses. The same transformation has already been used by Ferrer-Rosell *et al* (2016).

In general, an interpretable log-ratio transformation is easy to compute whenever there is an interpretable sequential binary partition of budget components into pairs of groups of components, according to the researchers' own
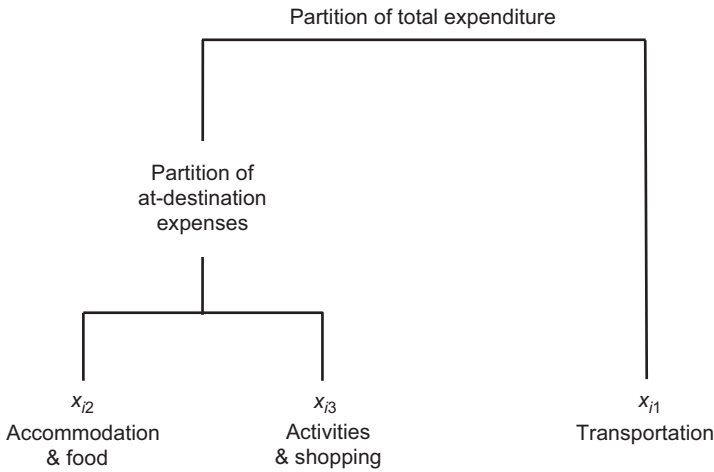
Partition of total expenditure

Partition of
at-destination
expenses

x_{i2}
Accommodation
& food

x_{i3}
Activities
& shopping

x_{i1}
Transportation

Figure 1. Sequential partition of budget parts used in this article.

objectives or hypotheses (Pawlowsky-Glahn and Egozcue, 2011). 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 with a tree diagram.

A meaningful log-ratio transformation starts by taking log-ratios of the geometric means of the two component clusters at each partition. Numerators and denominators are interchangeable. This article considers three budget parts:

x_{i1} = transportation expenditure from the place of origin to the accommodation at destination and back;
x_{i2} = accommodation and food, including both expenditure at bars and restaurants and grocery shopping;
x_{i3} = activities, moving around at destination and non-grocery shopping.

A sequential partition, which is interpretable with respect to the purpose of this article, is shown in Figure 1. The sequential partition in Figure 1 implies that the research questions involve the allocation of total expenditure between transportation and at-destination expenditure and the allocation of at-destination expenditure between accommodation and food on the one hand and activities and shopping on the other. At-destination expenditure is here understood as expenditure on services provided at the destination, regardless of whether they were paid for prior to departure or on the spot.

The first log-ratio compares transportation expenditure with the geometric mean of accommodation and food, on the one hand, and activities and shopping, on the other. This ratio is used to observe the share of transportation compared to at-destination expenses. Larger values show a higher relative importance of transportation expenses:

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

The second log-ratio is a ratio of accommodation and food over activities and shopping. This ratio indicates how tourists allocate at-destination expenses:

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

As regards the absolute expenditure, the focus is placed on studying the economic impact from the point of view of the transportation industry. t_i is computed as the logarithm of absolute expenditure on transportation, as in Equation (5):

$$t_i = \ln(x_{i1}). \tag{8}$$

The log-ratios and absolute expenditure in Equations (6), (7) and (8) would imply the first, second and third rows of the Ψ matrix in Equation (9), which fulfils all conditions for an ilr but the unit sum of squares in the first $D-1$ rows:

$$\Psi = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{9}$$

Hence, each of the $D-1$ first rows is multiplied by the appropriate constants ($\sqrt{2/3}$, and $\sqrt{1/2}$) from which the matrix expression in Equation (5) is obtained. The final log-ratios are:

$$y_{i1} = \sqrt{\frac{2}{3}}\left(\ln\left(\frac{x_{i1}}{\sqrt{x_{i2}x_{i3}}}\right)\right) = \sqrt{\frac{2}{3}}\ln(x_{i1}) - \sqrt{\frac{1}{6}}\ln(x_{i2}) - \sqrt{\frac{1}{6}}\ln(x_{i3}), \tag{10}$$

$$y_{i2} = \sqrt{\frac{1}{2}}\left(\ln\left(\frac{x_{i2}}{x_{i3}}\right)\right) = \sqrt{\frac{1}{2}}\ln(x_{i2}) - \sqrt{\frac{1}{2}}\ln(x_{i3}). \tag{11}$$

The choice of the relevant absolute expenditure is as dependent on the objectives and research questions as the choice of log-ratios. For instance, if the focus would have been on the economic impact on the destinations then $t_i = \ln\left(\sqrt{x_{i2}x_{i3}}\right) = (1/2)\ln(x_{i2}) + (1/2)\ln(x_{i3})$ would have been chosen.

Zero expenditure

If the x_{id} variables contain zero expenditure, neither logarithms nor log-ratios can 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 aggregate expenditure on particular activities or expenditure on particular food items, for instance.

Nevertheless, one must be aware that amalgamation is a decision that prevents the separate analysis of the amalgamated components.

In certain instances, some zero components result from individual characteristics, which are called essential zeros, structural zeros, or absolute 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. A classic essential zero example in household budget research is expenditure on tobacco: it will essentially be zero if all members are non-smokers. In many instances the analyst does not have enough information to know whether zero expenditure comes closer to being essential or rounding zeros. Some tourists may indeed systematically choose to spend nothing on activities and shopping. Other tourists may spend a certain amount on activities and shopping on certain trips, but not on others and so surveys of only one trip will unavoidably contain some zeros. Tourists may also forget or fail to report trivial expenses, like post-card shopping, local bus tickets, going to a museum, and the like (Legohérel, 1998). Fry *et al* (2000) claim that, even if uncertain about the nature of zeros, they can be proxied by a value below the smallest non-zero expenditure, which is tantamount to treating them as rounding zeros. Simulations show this method performs well if the proportion of zeros is below 10% (Martín-Fernández *et al*, 2011). When the number of zeros is large, this approach indeed falls short, and analysing zero spenders as a separate group is preferable.

Sample and variables

This article uses secondary official statistics data from the Encuesta de Gasto Turístico (EGATUR) conducted by the Instituto de Turismo de España (ITE), an official agency of the Spanish Ministry of Industry, Energy and Tourism. The EGATUR survey asks about tourist expenditure, and trip and traveller characteristics. All expenditure variables are provided per person (not per travel group). The sample is non-proportionally stratified by country of residence, airport and month. In 2012 it was conducted in 19 major Spanish airports, by computer-assisted personal interview when tourists were about to leave the country.

The universe in this article is defined as a subset of the EGATUR universe, which consists of European leisure visitors arriving by LCA and spending between one and 120 nights in Spain. Flights from outside Europe are excluded because these airlines mostly operate short-haul flights. Multi-stage trips are excluded because the decision on expenditure for these trips is expected to fundamentally differ from that of single-stage trips. Tourists who have essential zeros in accommodation (tourists who own a house at the destination or tourists who stay with friends or relatives) are not considered. Finally tourists who do not decide how much they spend on certain components or do not pay by themselves (package tourists, and trips paid for by gifts or contests) are not considered either because of sheer inability to observe all expenditure components. The final sample size is $N = 14,446$.

The following traveller characteristics are used as predictors: travel group, country of residence, education, self-reported income category, being repeat or

first-time visitor, gender, age and professional status. Since all variables are qualitative, MANOVA is applied.

Regarding treatment of zeros, in this article only one budget category (x_{i3}, activities and shopping) contains zeros (4.7%). The minimum amount spent by the non-zero group is €2.00. Zeros are replaced with €1.30 which roughly corresponds to the price of a city bus ticket, the entrance to a subsidized local museum, or a cheap souvenir. Given the small proportion of zeros in this category, no distortion is expected due to the replacement.

Results

Regarding the question whether explanatory variables are relevant for tourist expenditure all things considered, gender is the only variable which is not globally significant (p-value > 0.01 in the first column of Table 1). For this variable no further tests are interpreted.

Univariate tests referring to the chosen absolute expenditure (how much) and to each particular log-ratio (how) show that all variables influence absolute expenditure on transportation (p-value < 0.01 in the fourth column in Table 1); that income category is not significantly related to y_2 (that is, to the allocation of at-destination budget between accommodation-food and activities-shopping); and that previous visit to Spain does not significantly predict either y_1 or y_2 (p-value > 0.01 in the second and third columns in Table 1).

Table 2 shows that, compared to tourists travelling with a partner, tourists travelling with family spend relatively more on transportation and less at the destination. Within at-destination expenses, tourists travelling with family spend comparatively more on activities and shopping and less on accommodation and food, as tourists travelling with friends do. In terms of absolute expenditure on transportation, those travelling alone and with family spend more than those travelling with a partner.

Regarding how LCA tourists from different countries allocate their trip budget, results show that tourists from the UK and Ireland are the ones who spend relatively the most on transportation, compared to at-destination expenses; and within at-destination expenses, they are the ones who spend relatively the least on activities and shopping. As far as the transportation expenditure volume is concerned, compared to tourists from the UK and Ireland, tourists from Scandinavia, Austria, Switzerland, Liechtenstein and Germany spend more, and tourists from Italy, other European countries and France spend less.

LCA tourists with lower education spend absolutely more on transportation than tourists with university education. Regarding how tourists distribute their trip budget, tourists with lower education spend comparatively more on transportation and less at destination, and within the expenditure at destination, they spend comparatively more on accommodation and food and less on activities and shopping.

Compared to LCA tourists with medium income, those with high income tend to spend more at the destination than on transportation, and to have higher absolute expenditure on transportation. There is no difference between tourists with low and medium income in how they distribute their trip budget, but

Table 1. MANOVA hypothesis test *p*-values.

Variable	Multivariate Pillai's trace test	Univariate *F*-tests		
		y_1: log-ratio transportation/ at-destination	y_2: log-ratio accommodation/ activities	*t*: log transportation absolute expenditure
Travel group	0.000	0.000	0.000	0.000
Country of residence	0.000	0.000	0.000	0.000
Education	0.000	0.000	0.000	0.000
Reported income category	0.000	0.000	0.192	0.000
Repeat visitor	0.000	0.825	0.051	0.000
Gender	0.190	0.030	0.169	0.268
Age	0.000	0.000	0.000	0.000
Professional status	0.000	0.000	0.000	0.000

Table 2. Category frequencies and MANOVA parameter estimates.

Variable	Parameter	Category Frequency	y_1: log-ratio transportation/ at-destination	y_2: log-ratio accommodation/ activities	t: log transportation absolute expenditure
	Intercept		-0.022	1.089**	4.914**
Travel group	Alone	15.7%	-0.077**	0.026	0.120**
	With family	17.0%	0.103**	-0.111**	0.088**
	With friends	19.8%	-0.083**	-0.238**	0.018
	With partner	47.5%	0	0	0
Country of residence	Germany	10.3%	-0.127**	-0.117**	0.079**
	Italy	17.4%	-0.265**	-0.200**	-0.368**
	France	8.7%	-0.157**	-0.229**	-0.166**
	Benelux	11.4%	-0.053*	-0.133**	-0.022
	Scandinavia	11.0%	-0.004	-0.221**	0.358**
	Austria, Switzerland, Liechtenstein	3.3%	-0.094*	-0.287**	0.169**
	Other	5.7%	-0.203**	-0.259**	-0.224**
	UK and Ireland	32.2%	0	0	0
Education	Up to high school	28.2%	0.068**	0.106**	0.118**
	University degree	71.8%	0	0	0
Reported income category	High income	23.7%	-0.098**	0.016	0.139**
	Low income	3.0%	0.001	-0.072	0.120**
	Medium income	73.3%	0	0	0
Repeat visitor	Yes	23.0%	-0.003	-0.036	-0.081**
	No	77.0%	0	0	0

Gender	Female	46.7%	0.025	0.021	0.010
	Male	53.3%	0	0	0
Age	+45 retired	6.1%	-0.070**	0.192**	0.107**
	15–24	11.7%	-0.109**	-0.171**	0.014
	+45 not retired	28.5%	-0.008	0.057*	0.084**
	25–44	53.6%	0	0	0
Professional status	Homemaker	2.1%	0.186**	0.295**	0.127**
	Unemployed	2.8%	0.041	0.165**	-0.167**
	Low-level employee	4.9%	0.219**	0.236**	-0.083**
	Student	6.6%	0.161**	0.139*	-0.157**
	High-level employee	9.7%	0.050	0.091*	0.115**
	Self-employed	14.2%	0.024	0.019	0.026
	Mid-level employee	59.7%	0	0	0

Notes: The last (most frequent) category in each variable acts as reference. *t-test p-value < 0.01; **t-test p-value < 0.001.

there are significant differences in terms of how much they spend on transportation. Low income earners spend more than medium income earners.

Tourists repeating their visit to Spain spend absolutely less on transportation than those who travel there for the first time. No differences are found between first-time and repeat visitors in how they allocate their trip budget.

Regarding trip budget allocation, retired tourists spend comparatively more at destination and less on transportation, and within at-destination expenses, they spend relatively more on accommodation and food. While young (15–24-year-old) travellers also spend comparatively less on transportation, within at-destination expenses they are the age group which spends comparatively the most on activities and shopping. As far as transportation expenditure volume is concerned, tourists above 45 spend more than middle-age travellers.

Compared to tourists with a mid-level employment, unemployed, low-level employees and students spend absolutely less on transportation. Regarding at-destination expenditure allocation, mid-level employees spend comparatively the most on activities and shopping compared to accommodation and food. Conversely, homemakers, low-level employees and students spend relatively more on transportation and less at destination.

For comparative purposes, the analysis is rerun with the common modelling of all log absolute expenditure, in Equation (1), by budget parts as dependent variables (Tables 3 and 4). Pillai's trace tests in Table 3 are identical to those in Table 1. The first variable $\ln(x_1)$ in Table 3 is identical to the absolute expenditure in Table 1. For the remaining variables, the results of the tests change, because in Table 3 all tested hypotheses refer to budget volume. Differences in budget allocation are not testable.

Table 4 leads us to conclude that many tourist characteristics affect absolute expenditure on all budget parts in the same direction. Tourists travelling alone, tourists coming from Germany, Scandinavia, Austria, Switzerland and Liechenstein, and high income earners spend more in all budget parts. Similarly, unemployed, low-level employees and students spend less in all budget parts. The results in Table 4 hide the significant effects that all these categories had on at least one budget allocation log-ratio in Table 2.

Conclusion and discussion

Expenditure allocation and expenditure volume are of interest to both tourism scholars and managers. The objective of this article is twofold: first, to provide new methodological tools to combine the analysis of budget share and absolute expenditure; and, second, to determine how traveller characteristics affect both absolute and relative importance of transportation expenditure for tourists arriving to Spain by LCAs.

Regarding the first objective, the approach suggested in this article avoids confounding how tourists distribute their expenditure and how much they spend. This approach shows which variables affect expenditure allocation and which affect absolute expenditure. On the contrary, the usual approach of analysing Equation (1) makes it impossible to draw any conclusion about expenditure allocation. This article has developed a method with the specific

Table 3. MANOVA hypothesis test *p*-values when modelling all log absolute expenditure.

Variable	Multivariate Pillai's trace test	Univariate *F*-tests		
		$\ln(x_1)$ transportation absolute expenditure	$\ln(x_2)$ accommodation absolute expenditure	$\ln(x_3)$ activities absolute expenditure
Travel group	0.000	0.000	0.000	0.000
Country of residence	0.000	0.000	0.000	0.000
Education	0.000	0.000	0.000	0.110
Reported income category	0.000	0.000	0.000	0.000
Repeat visitor	0.000	0.000	0.000	0.052
Gender	0.190	0.268	0.564	0.098
Age	0.000	0.000	0.000	0.000
Professional status	0.000	0.000	0.000	0.000

Table 4. MANOVA parameter estimates when modelling all log absolute expenditure.

Variable	Parameter	$\ln(x_1)$ transportation absolute expenditure	$\ln(x_2)$ accommodation absolute expenditure	$\ln(x_3)$ activities absolute expenditure
	Intercept	4.914**	5.711**	4.171**
Travel group	Alone	0.120**	0.232**	0.196**
	With family	0.088**	−0.116**	0.040
	With friends	0.018	−0.049**	0.287**
	With partner	0	0	0
Country of residence	Germany	0.079**	0.151**	0.317**
	Italy	−0.368**	−0.185**	0.097*
	France	−0.166**	−0.135**	0.189**
	Benelux	−0.022	−0.051	0.137**
	Scandinavia	0.358**	0.206**	0.520**
	Austria, Switzerland, Liechtenstein	0.169**	0.081*	0.487**
	Other	−0.224**	−0.159**	0.208**
	UK and Ireland	0	0	0
Education	Up to high school	0.118**	0.109**	−0.041
	University degree	0	0	0
Reported income category	High income	0.139**	0.270**	0.248**
	Low income	0.120**	0.067	0.169*
	Medium income	0	0	0
Repeat visitor	Yes	−0.081**	−0.103**	−0.051
	No	0	0	0

Gender	Female	0.010	−0.006	−0.036
	Male	0	0	0
Age	+45 retired	0.107**	0.329**	0.057
	15–24	0.014	0.028	0.269**
	+45 not retired	0.084**	0.134**	0.053
	25–44	0	0	0
Professional status	Homemaker	0.127**	0.108*	−0.309**
	Unemployed	−0.167**	−0.101*	−0.334**
	Low-level employee	−0.083**	−0.185**	−0.519**
	Student	−0.157**	−0.256**	−0.453**
	High-level employee	0.115**	0.118**	−0.010
	Self-employed	0.026	0.011	−0.016
	Mid-level employee	0	0	0

Notes: The last (most frequent) category in each variable acts as reference. $^{*}t$-test p-value < 0.01; $^{**}t$-test p-value < 0.001.

needs of tourist and transportation research in mind, and with the aim of making it useful for studying tourist expenditure.

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. The CODA methodology offers the potential to construct tailor-made log-ratios which are intuitive to interpret and suit one's own research questions, as far as budget allocation is concerned. A partition tree of components is a clear and useful tool in this respect. The CODA methodology has been rightly criticized for ignoring volume even when it is available. Drawing from Pawlowsky-Glahn *et al* (2015a) this article shows alternative ways to include the absolute expenditure in the analysis, which can also be tailored to the research questions at hand. In research oriented towards the impact of expenditure on transportation industries, absolute expenditure related to transportation is an attractive choice, as shown in this article. The research objectives may also lie in other expenditure components or even in all of them. The flexibility of this approach adapts to any kind of study on tourist expenditure.

The suggested approach is basically a rotation of the common approach by means of a change of basis. As such it yields identical multivariate tests, while leading to more interpretable univariate tests and estimates. It also shares its statistical assumptions, which have to be checked for the particular data set at hand by means of the usual diagnostics in MANOVA and MANCOVA models. Non-normality, for instance, will be an issue both for the common approach, and the approach in this article when sample sizes are small. Robust methods have recently become available for MANOVA of compositions (Mateu-Figueras *et al,* 2015), but they have not yet been developed for models including a total.

Regarding the second objective, LCA users' characteristics which lead them to spend more on transportation in absolute terms and those which lead them to spend more in relative terms are mostly different. Tourists spending the most in relative terms are those travelling with family, from the UK and Ireland, with up to high-school education, with low and medium income, middle-aged, low-level employees, and homemakers. Whereas, tourists spending the most in absolute terms are those travelling alone, from Scandinavian countries, with up to high-school education, with either high or low income, first-time visitors, over 45 years' old, homemakers and high-level employees.

When comparing the results with the literature on absolute expenditure, this article confirms previous findings regarding travel group size, age, income and professional status. Looking at the results on relative expenditure, it replicates previous findings on education, income, country of residence, travel group and professional status.

As far as practical implications are concerned, airlines' pricing strategies can benefit from knowing which tourists spend more in absolute terms according to, for instance, age, travelling with children or being a student. Looking at the economic implications for destinations, they might concentrate on attracting tourists who spend relatively less on transportation or, in other words, who allocate more of their trip budgets to at-destination expenses (negative coefficients in the y_1 column in Table 2).

This study is of course limited by the available variables in the EGATUR survey. Further research on the expenditure decisions relevant to the transpor-

tation industry is needed with a comprehensive list of predictors (Thrane, 2014). It also has to be taken into account that the results are not generalizable beyond the Spanish scope.

As regards methodological further research, zero expenditure has deserved wide attention both in the econometric literature and in the CODA literature, but there has been hardly any cross-fertilization. Drawing from both traditions, additional work is needed about the analysis and replacement of zeros when compositions are combined with absolute information.

Finally, as regards applied further research, researchers are encouraged to use the method introduced in this article for other tourism expenditure studies, by adapting log-ratios and absolute expenditure to their own objectives – for example, by analysing the geometric average of absolute at-destination expenses in economic impact studies or by focusing on absolute expenditure on activities in active tourism research. Of course, this method can also be used to study other means of transport, and even to compare them by means of moderating effects.

References

Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol 44, No 2, pp 139–177.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2000), 'Logratio analysis and compositional distances', *Mathematical Geology*, Vol 32, No 3, pp 271–275.

Barquet, A., Brida, J.G., Osti, L., and Schubert, S. (2011), 'An analysis of tourists' expenditure on winter sports events through the Tobit censorate model', *Tourism Economics*, Vol 17, No 6, pp 1197–1217.

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.

Brida, J.G., Bukstein, D., Garrido, N., and Tealde, E. (2012), 'Cruise passengers' expenditure in the Caribbean port of call of Cartagena de Indias: a cross-section data analysis', *Tourism Economics*, Vol 18, No 2, pp 431–447.

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

Chang, K.-L., Chen, C.-M., and Meyer, T. (2013), 'A comparison study of travel expenditure and consumption choices between first-time and repeat visitors', *Tourism Management*, Vol 35, pp 275–277.

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.

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

Dobruszkes, F. (2013), 'The geography of European low-cost airline networks: a contemporary analysis', *Journal of Transport Geography*, Vol 28, pp 75–88.

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.

Ferrer-Rosell, B., Coenders, G., and Martínez-Garcia, E. (2015), 'Determinants in tourist expenditure composition: the role of airline types', *Tourism Economics*, Vol 21, No 1, pp 9–32.

Ferrer-Rosell, B., Coenders, G., and Martínez-Garcia, E. (2016), 'Segmentation by tourist expenditure composition. An approach with compositional data analysis and latent classes', *Tourism Analysis*, Vol 21, No 5.

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.

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.

Graham, A. (2013), 'Understanding the low cost carrier and airport relationship: a critical analysis of the salient issues', *Tourism Management*, Vol 36, pp 66–76.

Hadjikakou, M., Chenoweth, J., Miller, G., Druckman, A., and Li, G. (2014), 'Rethinking the economic contribution of tourism: case study from a Mediterranean Island', *Journal of Travel Research*, Vol 53, No 5, pp 610–624.

Hong, G-S., Morrison, A.M., and Cai, L.A. (1996), 'Household expenditure patterns for tourism products and services', *Journal of Travel and Tourism Marketing*, Vol 4, No 4, pp 15–39.

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

ITE (2013a), *Movimientos Turísticos en Fronteras (Frontur) y Encuesta de Gasto Turístico (Egatur) 2012*, Instituto de Turismo de España, Madrid (http://estadisticas.tourspain.es/es-ES/estadisticas/egatur/ Anuales/Movimientos Turísticos en Fronteras (Frontur) y Encuesta de Gasto Turístico (Egatur) 2012.pdf, accessed 26 November 2015).

ITE (2013b), *Turismo, Tráfico Aéreo y Compañías Aéreas de Bajo Coste en el Año 2012*, Instituto de Turismo de España, Madrid (http://estadisticas.tourspain.es/es-ES/estadisticas/otrasestadisticas/ companiabajocoste/anuales/Informe anual de CBC. Año 2012.pdf, accessed 26 November 2015).

Kim, W.G., Kim, T.T., Gazzoli, G., Park, Y., Kim, S.H., and Park, S.S. (2011), 'Factors affecting the travel expenditure of visitors to Macau, China', *Tourism Economics*, Vol 17, No 4, pp 857–883.

Kim, S.S., Prideaux, B., and Chon, K. (2010), 'A comparison of results of three statistical methods to understand the determinants of festival participants' expenditures', *International Journal of Hospitality Management*, Vol 29, No 2, pp 297–307.

Lee, S., Jee, W., Funk, D., and Jordan, J. (2015), 'Analysis of attendees' expenditure patterns to recurring annual events: examining the joint effects of repeat attendance and travel distance', *Tourism Management*, Vol 46, pp 177–186.

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.

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., 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.

Martín-Fernández, J.A., Buxeda-i-Garrigós, J., and Pawlowsky-Glahn, V. (2015a), 'Logratio analysis in archeometry: principles and methods', in Barceló, J.A., and Bogdanovic, I., eds, *Mathematics and Archaeology*, CRC press, Boca Raton FL, pp 178–189.

Martín-Fernández, J.A., Daunis-i-Estadella, J., and Mateu-Figueras, G. (2015b), 'On the interpretation of differences between groups for compositional data', *SORT- Statistics and Operations Research Transactions*, Vol 39, No 2, pp 231–252.

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

Mason, K.J., and Alamdari, F. (2007), 'EU network carriers, low cost carriers and consumer behaviour: A Delphi study of future trends', *Journal of Air Transport Management*, Vol 13, pp 299–310.

Mateu-Figueras, G., Pawlowsky-Glahn, V., and Egozcue, J.J. (2011), 'The principle of working on coordinates', in Pawlowsky-Glahn, V., and Buccianti, A., eds, *Compositional Data Analysis. Theory and Applications*, Wiley, New York, pp 31–42.

Mateu-Figueras, G., Daunis-i-Estadella, J., and Martín-Fernández, J.A. (2015), 'Manova for

compositional data, non-parametric first approach', paper presented at the Workshop on Compositional Data Analysis (CoDaWork 2015), L'Escala, Spain.

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.

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.

Pawlowsky-Glahn, V., Egozcue, J.J., and Lovell, D. (2015a), 'Tools for compositional data with a total', *Statistical Modelling*, Vol 15, No 2, pp 175–190.

Pawlowsky-Glahn, V., Egozcue, J.J., and Tolosana-Delgado, R. (2015b), *Modeling and Analysis of Compositional Data*, Wiley, Chichester.

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

Thrane, C. (2014), 'Modelling micro-level tourism expenditure: recommendations on the choice of independent variables, functional form and estimation technique', *Tourism Economics*, Vol 20, No 1, pp 51–60.

Thrane, C. (in press), 'Analysing related choices in tourism research: an intuitive and non-technical primer', *Tourism Economics*, DOI: 10.5367/te.2014.0453.

Van Der Boogaart, K.G., and Tolosana-Delgado, R. (2006), 'Compositional data analysis with "R" and the package "compositions"', *Geological Society, London, Special Publications*, Vol 264, pp 119–127.

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.

Zheng, B., and Zhang, Y. (2013), 'Household expenditures for leisure tourism in the USA, 1996 and 2006', *International Journal of Tourism Research*, Vol 15, No 2, pp 197–208.