Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen - Bolzano

Fakultät für Wirtschaftswissenschaften | Facoltà di Economia | School of Economics and Management

BEMPS –
Bozen Economics & Management Paper Series

NO 10 / 2013

A Multivariate Nonlinear Analysis of Tourism Expenditures

Marta Disegna, Fabrizio Durante, Enrico Foscolo

A multivariate nonlinear analysis of tourism expenditures

Marta Disegna, **Fabrizio Durante**, **Enrico Foscolo**

{marta.disegna,fabrizio.durante,enrico.foscolo}@unibz.it
Free University of Bozen–Bolzano
School of Economics and Management
39100 Bozen–Bolzano, Italy

Abstract

Independence among different tourism expenditure categories is the most convenient hypothesis for modeling decision–making processes. Nevertheless, the best-suited framework would require dependence among expenditures in order to face individual budget and ordered choices. To this end we provide a new multivariate copula-based logit model with explanatory variables. We applied our tools to the expenditures of the foreign tourists visiting South–Tyrol (Northern Italy), and we underlined the need to go beyond usual independence assumption in order to get more realistic results. The obtained findings are useful for policy makers, marketing experts, and local government in order to know how visitors allocate their travel budget; moreover, they can been exploited to improve the touristic supply by means of ad–hoc promotions, advertising, touristic packages, and attractions.

Keywords: Tourism expenditures; Copula; Foreign tourism; Spending behavior.
JEL codes: C14, C30, D12, L83

1 Introduction

Understanding and measuring the impact of the determinants on the level of tourism expenditures is essential for those countries whose economies refer to tourism as the major source of income in many countries (Hung et al., 2012). In case of Italy nonresident spending on holidays grew by 3.8% in 2012 compared with 2011, covering around 32 billion euro (Banca d'Italia, 2011); the Italian regions that mostly attract tourists have been Trentino–South Tyrol, Veneto (both in the North–East of Italy), Lazio, and Toscana (both in the Center of Italy) (ISTAT, 2012).

In general, studies on tourism demand can be divided into two categories: micro– and macroeconomic based approaches (Wang and Davidson, 2010). While several studies examine Macroeconomic dynamics, we recognize a lack of contributions at the micro–level (Wang and Davidson, 2010; Fredman, 2008; Crouch, 1994). Usually, at the macro level the aggregate data are used to describe the tourist flows to a particular destination and to identify the factors that influence it. Based such data, many different methodologies are used in order to model tourism demand, such as single equation and system of equations, like Almost Ideal Demand System (AIDS) and Time Varying Parameter (TVP) (for comprehensive reviews see Song and Li, 2008; Li et al., 2005; Lim, 1997; Crouch, 1994).

At the micro level, while most studies have focused on the overall tourism expenditures, a limited attention was paid to the analysis of single tourism expenditure categories, such as shopping, food and beverages, accommodation, transportation, entertainment, and their interactions (Brida et al.; Brida et al., 2012; Wang et al., 2006; Bilgic et al., 2008; Hong et al., 2005; Lee, 2002; Lee, 2001; Cai et al., 1995).

Generally, an important characteristic of expenditure is that the amount spent by an individual is the result of two sequential decision processes. First of all, each individual decides if he/she wants to spend (the so–called selection stage), and then he/she decides how much money to spend (the so–called outcome stage). Probit, Tobit, two–stage Tobit and their generalizations are the most common models to deal with this kind of process (Brida and

Scuderi, 2013; Kim et al., 2010; Jang and Ham, 2009).

Furthermore, the microeconomic studies usually assume independence between categories, without considering that consumer preferences are weakly separable and, consequently, different categories may be correlated. Relevant exceptions are represented by Bilgic et al. (2008), in which the bivariate Tobit model (Amemiya, 1974) has been applied, and by Divisekera (2010), where an AIDS model is estimated (Deaton and Muellbauer, 1980).

In order to fill this gap in the literature, this study proposes a novel nonlinear approach to analyse simultaneously the willingness to pay on different dependent tourism expenditure categories, by considering also a set of explanatory variables. The proposed methodology is grounded on the concept of copula function (Jaworski et al., 2013, 2010). Copulas are valid alternatives to several classical multivariate distribution functions such as Normal, Student's t, and Pareto distributions, especially when a different distribution is needed for each univariate margin. To the best of our knowledge, this is the first study in tourism science where copulas are used to build the joint probability distribution of tourism expenditure categories by considering their dependence in a nonlinear way.

The dataset analyzed here consists of $n = 548$ EU nonresident holiday trips in one of the cities in the South–Tyrol, one of the most visited Italian region, in 2011. The data comes from the national survey on "International Tourism in Italy", annually conducted by the Bank of Italy.

The paper is organized as follows. In Section 2 we review the most important contributions in tourism expenditure, with particular attention to the models adopted. In section 3 we give a short introduction on copula theory, and a detailed description of the proposed model for our dependent variables is reported. Section 4 introduces the data analyzed in this paper, while in Section 4.3 the empirical results are summarized. Section 5 concludes, focusing on practical implications and future perspectives.

2 Literature review

Following the consumer behavior theory, each decision-making process to purchase can be described as a two-step stage, or decision, process: 1) the decision to spend or not; and 2) if spending, how much money to spend.

Different methods have been proposed to estimate tourism expenditure, overall or for specific tourism categories such as shopping, food and beverages, accommodation, transportation, and entertainment. Most of the tourism demand studies use Ordinary Least Square (OLS) estimation in linear regression models (Wu et al., 2013, Brida and Scuderi, 2013, Wang and Davidson, 2010). However, this method may lead to inconsistent and bias parameter estimates since the Normal assumption of the dependent variable is often unrealistic due to the fact that the expenditure is assumed to be non–negative (i.e. left censored). In such a case, the Tobit regression model (Tobin, 1958) has been adopted in order to provide a suitable estimation for censored dependent variable (Barquet et al., 2011; Kim et al., 2011; Lee, 2001; Leones et al., 1998). Most recently, the quantile regression model has also been applied in place of the traditional OLS estimation (Saayman and Saayman, 2012; Chen and Chang, 2012), but even in this case an appropriate censored quantile regression model would be preferable. Nevertheless, all these models estimate the effects of a set of explanatory variables on the dependent variable only regarding the second stage of the decision-making process (i.e. how much to spend).

In order to extend these approaches by take into account both stages, the double-hurdle model (Cragg, 1971), was then adopted (Jang and Ham, 2009; Hong et al., 1999). Further developments include: the Heien and Wessells two-step estimator (Heien and Wessels, 1990), that, in contrast with Heckman (1979), allows the use of all the observations at each stage (Brida et al.; Brida et al., 012b; Kim et al., 2010; Jang and Ham, 2009; Jang et al., 2007); the integrated Scobit-based discrete-continuos choice model that allows to estimate the two steps simultaneously (Wu et al., 2013; Zhang et al., 2012b).

Another important economic theory that must be taken into account by modeling the expenditures is the neoclassical economic theory of consumer behavior: the consumers are rational and they want maximize their utility function by choosing among a set of available alternatives. Consumers are assumed to be able to rank goods and services, so that they select the particular combinations for which their utility function takes the largest value, at a certain level of income (budget). Furthermore, the consumer's utility function is "separable". The separability, and in particular the assumption of weak separability, assumes independence only among groups of commodities instead of among individual commodities. This implies that the budgeting procedure by which individuals allocate their incomes among different goods and services is composed by two stages (Deaton and Muellbauer, 1980): first, the individual decides in which broad commodity groups (like, food, tourism, housing, clothes, etc.) want to allocate its income; second, the individual decides which goods and services he/she want to buy within each group without any reference to the expenditure in the other groups. Syriopoulos and Sinclair (1993) applied this approach to the field of tourism suggesting a three-stage budgeting process. In the first stage, visitor allocates its budget between total tourism expenditure and consumption of other goods and services. In the second stage, a visitor allocates his/her tourism budget among different destinations, including the home country. Third, a visitor chooses how to allocate its tourism budget among various goods and services offered by the selected destination(s). Until now, little attention was paid to the analysis of the third stage of this budget process, whereas existing demand studies correspond to the second stage. The studies of Bilgic et al. (2008) and Divisekera (2010) represent two important exception in the tourism demand literature modeling the third stage of this budget process, i.e. considering the dependence among different tourism expenditure categories. The first study adopted the bivariate Tobit model (Amemiya, 1974) to simultaneously estimate the share expenditure on two broad classes of leisure activities (hunting and fishing), while the second estimated an AIDS model (Deaton and Muellbauer, 1980) to study the way by which tourists choose to spend their money among various goods and services they need during the trip. In each model, the dependent variables are expenditure shares, and both models have the drawback that the estimation is done under three restrictive economic conditions, seldom satisfied: adding-up, homogeneity, and symmetry.

Furthermore, in the case of the bivariate Tobit model, "the procedure is computationally cumbersome for systems of more than two share equations" (Bilgic et al., 2008).

As we will see in the next, the approach that we propose allows to simultaneously model, in a nonlinear way, the willingness to pay on different tourist expenditure categories (selection stage of the decision-making process) considering both the hidden structure of dependency (third stage of the budgeting process) and the effect produced by a set of explanatory variables. Unlike the research of Bilgic et al. (2008) and Divisekera (2010), in this work the dependent variables are assumed to be dichotomous. Moreover, no restrictive conditions or a maximum number of expenditure categories are imposed.

3 The methodology

Let us assume that Y_j $(j = 1, \dots, J)$ is the tourism expenditure on the j-th category, as could be shopping, food and beverages, accommodation, transportation, or entertainment. Since it is unrealistic to suppose independence among all these tourism expenditures categories, a suitable dependence model is necessary to be considered.

Classical statistical models for such situations involve the formulation of a parametric (e.g. Normal) model that could describe the links among the involved variables. One of the possible drawbacks of such models is that neither they do allow flexibility in the choice of the univariate margins nor they describe tail dependencies in a convenient way (e.g., see McNeil et al., 2005). An alternative to such models is given by the use of *copulas*, which have been employed in a number of different applications (see Jaworski et al., 2010, 2013), including marketing and tourism science (e.g., Trivedi and Zimmer, 2007; Danaher and Smith, 2011; Zhang et al., 2012a and the references therein).

Copulas are multivariate probability distribution functions (shortly, dfs) whose univariate margins are uniformly distributed on $[0, 1]$. They play an important role in building statistical models since it has been proved that the multivariate df $F_{\boldsymbol{Y}}$ of $\boldsymbol{Y}$, i.e. for all $\boldsymbol{x} \in \mathbb{R}^d$

$$F_{\boldsymbol{Y}}(y_1, \ldots, y_J) = \mathbb{P}(Y_1 \leq y_1, \ldots, Y_J \leq y_J)$$

can be described in the form

$$F_{\boldsymbol{Y}}(y_1, \ldots, y_J) = C\left(F_1(y_1), \ldots, F_J(y_J)\right) \tag{1}$$

where $F_j(x_j) = \mathbb{P}(Y_j \leq y_j)$ and C is a suitable copula. A *copula-based statistical model* is any model for the df F of $\boldsymbol{Y}$ created by (1) where C and $F_1, \ldots, F_J$ belong to specific (parametric or semi–parametric) families of copulas and univariate df's, respectively.

Copula–based models have a number of attractive features. First, they permit the combination of any univariate marginal distributions that need not come from the same family. Second, they are very general, encompassing a number of existing multivariate models, and provide a framework for generating many more. Third, they allow to model discrete as well as continuous random variables (although the former case needs to be treated with specific methods as illustrated, for instance, by Joe, 1997). These advantages give copula models a greater potential for use in empirical analysis than existing probability models used.

In our analysis, we construct statistical models based on the following families of copulas that are common in the literature (for a formal definition of these classes see Durante and Sempi, 2010; Hofert and Maechler, 2011):

- Normal copula, which can be derived from the multivariate Normal distribution;
- Clayton copula, which belongs to Archimedean class and may have a non-zero lower tail dependence coefficient;
- Gumbel and Joe copula, which belongs to Archimedean class and may have a non–zero upper tail dependence coefficient;

- Frank copula, which belongs to the Archimedean class and does not exhibit any typical behavior in the tails.

In this work, we consider a random vector $\mathbf{Y}$ whose components are related to the different tourism expenditure categories. Specifically, we assume that each Y_j is a dichotomous variable describing the decision to spend (Y_j equal to 1) or not (Y_j equal to 0) in the j-th expenditure category. Moreover, we assume that the binary random variables are observed together with a vector $\mathbf{x} \in \mathbb{R}^p$ of explanatory random variables. Our aim is to estimate the distribution of $(Y_1, \ldots, Y_J)$ given a $\mathbf{x}$, i.e. to describe how expenditure choices are related to a specific sets of explanatory variables. In order to do this, following a copula approach, we assume that the multivariate df of $\mathbf{Y}$ may be expressed for every $y_j \in \{0, 1\}$ in the form

$$\mathbb{P}(Y_1 \leq y_1, \ldots, Y_J \leq y_J \mid \mathbf{x}) = C(F_1(Y_1 \leq y_1 \mid \mathbf{x}), \ldots, F_J(Y_J \leq y_J \mid \mathbf{x})), \quad (2)$$

where the marginal distribution functions $F_1(\cdot|\mathbf{x}), \ldots, F_J(\cdot|\mathbf{x})$ are fitted separately to the explanatory variables, and $C \in (C_\theta)$ is a suitable copula coming from a specific family indexed by θ. Models of type (2) are not new in the literature; for an overview about its recent applications, see (Song, 2007; Nikoloulopoulos and Karlis, 2008; Erhardt and Czado, 2012; Genest et al., 2013) and the references therein. However, to the best of our knowledge, this is the first time that this model is adopted in describing the consumer expenditures.

According to a copula approach, the estimation of the model (2) is done in two steps. First, we fit a univariate model for the marginal distributions. In particular, since each Y_j is a binary random variable and, hence, is completely determined by $\pi_j = \mathbb{P}(Y_j = 1 \mid \mathbf{x})$, we may consider a logistic regression,

$$\pi_j(\mathbf{x}) = \frac{\exp\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}_j\right)}{1 + \exp\left(\boldsymbol{x}_j^\top \boldsymbol{\beta}_j\right)}$$

where $\boldsymbol{x}_j$ stands for the vector of covariates and $\boldsymbol{\beta}_j$ represents a vector of parameters. The estimation of the marginal model is performed by maximum likelihood procedures and the values of the parameters $\widehat{\beta}_1, \ldots, \widehat{\beta}_d$ are obtained.

Secondly, we fit a suitable copula C to the residuals of the marginal models. Specifically, such a C is uniquely determined, for each vector $\mathbf{x}$ of explanatory variables, on the product of the ranges of the univariate marginals, given by

$$\{0, 1 - \pi_1(\mathbf{x}), 1\} \times \cdots \times \{0, 1 - \pi_J(\mathbf{x}), 1\}.$$

Supposed that C belongs to a specific parametric family (C_θ), the value of the parameter is chosen by maximum likelihood procedures adopted to the function

$$\sum_{i=1}^{n} \log(h_{C_\theta}(y_{i1}, \ldots, y_{iJ} \mid \mathbf{x})), \tag{3}$$

where

$$h_{C_\theta}(y_1, \ldots, y_J) = \sum \operatorname{sign}(\mathbf{v}) C(\widehat{F}_1(v_1 \mid \mathbf{x}), \ldots, \widehat{F}_d(v_d \mid \mathbf{x})). \tag{4}$$

Here the sum is taken over all $\mathbf{v} \in \mathbb{R}^J$ such that $v_j \in \{y_j, y_j - 1\}$, $\operatorname{sign}(\mathbf{v}) \in \{-1, 1\}$ and equals 1 if and only if the cardinality of the set of all indices j's such that $v_j = y_j - 1$ is even. Moreover, $\widehat{F}_i$ are the estimated marginal distribution functions of the previous step. For more details about this estimation, see Genest et al. (2013).

Notice that, fitting marginal parametric distributions first and the parametric copula for fixed margins afterwards are known as inference functions for margins or the IFM method (e.g., see Joe, 1997, Section 10.1).

In practice, different copula families are fitted to the same set of univariate marginals. A criteria is therefore need in order to select a proper copula among the possible different choices. In order to do this, we perform a test due to Vuong (1989). For more details about this procedure, we refer to Erhardt and Czado (2012).

Finally, a copula C_θ is estimated together with a set of marginals $\widehat{F}_1, \ldots, \widehat{F}_J$, and Eq. (2) can be used in order to describe the dependence among different expenditure choices given a predefined set of explanatory variables, as will be illustrated in the following.

4 The empirical study

4.1 The survey

The case study we are going to illustrate deals with the annual survey conducted by the Bank of Italy entitled “International Tourism in Italy” in 2011 (Banca d’Italia, 2011). The main aim of this survey is to monitor travel expenditure and length of stay of inbound and outbound visitors from/to Italy. Travel expenditure includes the total consumption of goods and services made in the country visited divided into the following categories: transportation, accommodation, food and beverages, shopping, and other services. This survey is conducted each year since 1996. The inbound–outbound frontier survey is the technique adopted for the collection of the data used in order to create the touristic balance. The stratified sampling method was applied (using different type of stratified variables per each type of frontier), and face–to–face interviews are made at national border (including highways, railway, airports, and harbors) . The sampling is carried out independently at each type of frontier. Tourists are interviewed at the end of the trip, when they are returning to their place of habitual residence. Interviews are conducted in different moment of the day, during both working day and holiday and month by month in the considered year with a fixed number of interviews per each period of survey. The questionnaires are anonymous, and written in 14 languages.

The questionnaire can be divided into four main sections: 1. socio–demographic characteristics of the interviewees (gender, age, occupation, and residence); 2. information on the trip (number of nights spent during the trip, type of accommodation, cities visited, means of transportation, motives of the trip, organization of the trip); 3. information on the expenditure (total expenditure divided into different category of expenditure, means of payment); 4. level of satisfaction on different aspects of the trip (hospitality and friendliness of the people; cities and works of art; landscape and natural environment; hotels and other accommodation; food and beverage; price and cost of living; quality and variety of products offered in stores; information and tourist services; safety of tourists) and on overall with the destination.

In 2011, 145,000 valid questionnaires have been randomly collected among foreign visitors through face–to–face interviews made using more than 200 interviewers located in more than 80 national border points.

We focus our case study on the 548 foreign visitors who visited any city in the province of Bolzano, in the South–Tyrol region (Northern Italy) in 2011 and whose main purpose was "tourism, holiday, leisure".

4.2 The dataset

The survey on "International Tourism in Italy" collects the actual expenditure (in Euro) for each of the five main categories under which the travel expenditure can be divided; i.e., transportation, accommodation, food and beverages, shopping, and other services. For the purpose of our study, each category is transformed into a dichotomous variable sets to 1 when a positive expenditure is observed, 0 otherwise.

Table 1 reports the estimates of the tetrachoric correlation coefficients with their significance levels. We observe the food and beverages category is not significantly correlated with the transportation and shopping categories. Moreover, the estimated correlation between the other services and shopping categories is very weak. For sake of illustration, we concentrate hence on expenditures on accommodation, transportation, and shopping that exhibit a stronger positive dependence.

After this preliminary selection, Chi–squared contingency table test statistics are performed for testing pairwise independence between dependent variables. As Table 2 shows, no case of independence among the $J = 3$ variables is recognized.

Thus, in the sample we observe that 86% of visitors are tourists (i.e. people who spend at least one night in the province of Bolzano), 76% purchase tickets and/or transportation fuel within Italy, and 69% buy souvenir, gift, item of clothing, or other things for personal usage (see table 3).

In order to prevent possible collinearity, we select a suitable set of explanatory variables (including destination attributes, other travel expenditures, and tourists' individual attributes). For destination attributes, the cognitive evaluation of the visited destination is selected to describe how much satisfied the visitors are with the visit. It is measured by asking respondents to report their level of satisfaction with the destination on a 10–point Likert–type scale ranged from [1] *Very unsatisfied* to [10] *Very satisfied*. The amount of money (expanded for the population and weighted for the number of nights and municipalities viewed during the trip) spent on food and beverage and on other services is included as explanatory variables in order to consider the whole expenditure behavior of the tourist. The selected individual attributes include composition of the travel party (1 = alone; 0 = otherwise), employment status (1 = employee; 0 = otherwise), and residential location (1 = from Germany or Austria; 0 = otherwise). In the samples, 12% of visitors travel alone, 72% are employed, and 72% are German–speaking.

Table 3 reports the complete list of the explanatory variables used in the application, with a brief description and some descriptive statistics.

4.3 Model results

As describe in Section 3, three univariate logistic regression models for the expenditures on accommodation, transportation, and shopping (Y_j, $j = 1, 2, 3$) are estimated by using White's robust standard variance-covariance matrix (White, 1980), which corrects for possible heteroskedasticity of the error term. A number $k = 6$ of explanatory variables are involved in the model (see Table 3 for a description). The backward results are presented in Table 4

Table 5 reports the estimated copula parameters obtained by adapting eq. (3), by including the previously estimated marginal models. Specifically, we fit the following family of copula: Normal, and the Archimedean copulas by Clayton, Gumbel, Frank, and Joe (see Section 3). As can be interpreted from these estimates, overall the pairwise fitted dependencies are positive. Notice that, while Archimedean copulas include only one parameter to describe dependence among the three expenditure categories, Normal copula is described by three parameters (related to pairwise correlation coefficients).

As such, it has a greater flexibility also to model different strength of dependence in the pairs (for instance, between accommodation and transportation expenditures in this case). We are aware the choice of copula families may be enlarged by using other classes from the literature.

As a term of comparison, in the following the Independence copula is considered as well since it represents an usual assumption for the touristic research framework.

The test decisions on the right copula model applied to our data are given in Table 6. We firstly observe that the Independence copula is not preferred over any other copula according to the Vuong test. Only the Normal copula is preferred among the remaining two classes of copulas. Among the Archimedean class members no clear evaluation of the preferred model can be made, since no Archimedean copula outperforms the others. We thus conclude that Normal copula may represent the best choice for $(Y_1, Y_2, Y_3)^\top$ according to our sample.

The distribution of the eight combinations for $(Y_1, Y_2, Y_3)^\top$ is listed in Table 7 for fitted Normal copula and Independence–based probability mass function (pmf) by setting covariates at their median values $\tilde{\boldsymbol{x}}_j$. Notice that the median tourist is "satisfied on overall with South Tyrol as destination", has a positive expenditure on food and beverage while has a null expenditure on other service, he/she is employee, comes from Germany or Austria, and makes the visit with someone.

Looking at table 7, we may appreciate the advantage to refuse the a priori independence assumption. In fact, in extreme scenarios, where visitors decide to spend in all categories or to not spend in anyone, the dependence scenario among $(Y_1, Y_2, Y_3)^\top$ shows higher values of the estimated probabilities.

Adopting the independence assumption implies hence an underestimation of the purchase in all the considered expenditure categories, in contrast with the more realistic dependence assumptions does. Moreover, under the independence assumption, a high percentage (about 11%) of tourists who have the stronger impact on the regional economy is excluded by any economical analysis that a local government can do.

For further underlying the advantages of the proposed approach, the probability distribution of the eight combinations of $(Y_1, Y_2, Y_3)^\top$ are calculated under dependence (Normal copula) and independence assumption, by varying the values of the dichotomous explanatory variables, keeping the other explanatory variables fixed to the median value. Tables 8, 9, and 10 report the estimated probability. All these findings show that the independence assumption underestimate the probability of the extreme combinations (spend in all the categories and not spend in anyone). On the basis of this analysis, the higher probability to spend both in accommodation, transportation, and shopping can be attributed to a tourist satisfied on overall with South Tyrol, who spends also on food and beverage but not on other service, who comes from Germany or Austria, makes the visit with someone, and who is not employee (Table 8). It is necessary to underline that the retiree is the main category among the not employee (78%). Therefore, we can say that the retirees have the higher probability to spend in all the tourism expenditure categories, held the other variables be fixed at their median values. On the other hand, tourists who make the visit alone generally have the higher probability to spend only for shopping or for other expenditure, not considered in this study (Table 9). Finally, we can observe that, keeping fixed at the median the other variables, tourists who do not come from Germany or Austria have a lower probability to spend in all the three categories and in accommodation and transportation than German–speaking ones, while they present a higher probability to spend in the other six combinations of $(Y_1, Y_2, Y_3)^\top$ (Table 10).

5 Conclusions

We have presented a multivariate Logit model based on the copula function useful in order to study dependence in the expenditure behavior. As a byproduct, given a set of explanatory variables associated with the tourist profile, we are able to predict the probability that the tourist is likely to spend in two or more tourism expenditure categories. This model is more flexible respect to the more common AIDS and bivariate Tobit model so far used for modelling different tourism expenditure simultaneously, since no restrictive conditions on its use are imposed.

To illustrate the main features of the proposed method in the tourism field, we analyzed the data collected by the Bank of Italy in 2011 throughout the survey entitled "International Tourism in Italy". In particular, we have considered the sub-group of 548 foreign visitors who visited the South-Tyrol region, in the Northern Italy, for "tourism, holiday, or leisure". The dependent variables adopted in this study are dichotomous and describe the individual decision-making process that lead to spend or not in a particular tourism expenditure categories, i.e. the first (selection) stage of the decision-making process. The tourism expenditure categories taken into consideration are accommodation, transportation, and shopping. Nevertheless, we stress that this model can be extended by using an higher number of dependent variables, and overcome the limitation encountered in using the bivariate Tobit model (Bilgic et al., 2008).

The main finding is that the independence among expenditures provides an underestimation of the probability of spending in all categories. Furthermore, we have observed that the tourist who presents the highest probability to spend in all the three categories is satisfied on overall with South Tyrol; spends also on food and beverage but not on other service; comes from Germany or Austria; makes the visit with someone; and is retired.

These results are of potential interest in tourism management and for local governments, e.g. in order to know how and with which probability visitors allocate their travel budget into different expenditure categories. Managing this information is fundamental for policy makers and marketing experts in order to improve the touristic supply by means of ad–hoc promotions, advertising, touristic packages, and attractions, and to implement specific consumer-based marketing campaign that offer a combination of different services (meals, lodging, shopping, etc.) according to tourist preferences.

Finally, notice that the proposed model cannot be adopted to estimate also the amount of money spend by the tourist in different categories (the outcome stage of the decision-making process). This will a matter of future investigation.

References

Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42(6):999–1012.

Banca d'Italia (2011). *Turismo Internazionale dell'Italia*. Banca d'Italia. http://www.bancaditalia.it/statistiche/rapp_estero/turismo-int.

Barquet, A., Brida, J., and Osti, L., S.-S. (2011). An analysis of tourists' expenditure on winter sports events through the Tobit censorate model. *Tourism Economics*, 17(6):1197–1217.

Bilgic, A., Florkowski, W., Yoder, J., and Schreiner, D. (2008). Estimating fishing and hunting leisure spending shares in the united states. *Tourism Management*, 29:771–782.

Brida, J., Disegna, M., and Osti, L. Visitors' expenditure behaviour at cultural events: the case of christmas markets. *Tourism Economics*. DOI: 10.5367/te.2013.0237.

Brida, J., Disegna, M., and Osti, L. (2012). The effect of authenticity on visitors' expenditure at cultural events. *Current Issues in Tourism*. DOI: 10.1080/13683500.2012.674105.

Brida, J., Disegna, M., and Scuderi, R. (2012b). The visitors' perception of authenticity at the museums: archaeology versus modern art. *Current Issues in Tourism*. DOI: 10.1080/13683500.2012.742042.

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

Cai, L., Hong, G.-S., and Morrison, A. (1995). Household expenditure patterns for tourism products and services. *Journal of Travel and Tourism Marketing*, 4(4):15–40.

Chen, C. and Chang, K. (2012). The influence of travel agents on travel expenditures. *Annals of Tourism Research*, 39(2):1258–1263.

Cragg, J. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39:829–844.

Crouch, G. (1994). The study of international tourism demand: A survey of practice. *Journal of Travel Research*, 33(4):41–55.

Danaher, P. and Smith, M. (2011). Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30:4–21.

Deaton, A. and Muellbauer, J. (1980). *Economics and consumer behavior*. Cambridge: Cambridge university Press.

Divisekera, S. (2010). Economics of tourist's consumption behavior: Some evidence from australia. *Tourism Management*, 31:629–636.

Durante, F. and Sempi, C. (2010). Copula theory: An introduction. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Workshop on Copula Theory and its Applications*, volume 198 of *Lecture Notes in Statistics*, chapter 1, pages 3–31. Springer Berlin Heidelberg.

Erhardt, V. and Czado, C. (2012). Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuariual Journal*, 2012(2):106–129.

Fredman, P. (2008). Determinants of visitor expenditures in mountain tourism. *Tourism Economics*, 14(2):297311.

Genest, C., Nikoloulopoulos, A., Rivest, L., and Fortin, M. (2013). Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Brazilian Journal of Probability and Statistics*, 27(3):265–284.

Heckman, J. (1979). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4):475–492.

Heien, D. and Wessels, C. (1990). Demand system estimation with micro data: a censored regression approach. *Journal of Business & Economic Statistics*, 8(3):356–371.

Hofert, M. and Maechler, M. (2011). Nested Archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, 39(9):1–20.

Hong, G.-S., Fan, J., Palmer, L., and Bhargava, V. (2005). Leisure travel expenditure patterns by family life cycle stages. *Journal of Travel and Tourism Marketing*, 18(2):15–30.

Hong, G. S., Kim, S. Y., and Lee, J. (1999). Travel expenditure patterns of elderly households in the US. *Tourism Recreation Journal*, 24(1):43–52.

Hung, W.-T., Shang, J.-K., and Wang, F.-C. (2012). Another look at the determinants of tourism expenditure. *Annals of Tourism Research*, 39(1):495–498.

ISTAT (2012). *Annuario statistico Italiano, 2012*. ISTAT. http://www.ontit.it/opencms/opencms/ont/it/documenti/02947.

Jang, S. and Ham, S. (2009). A double–hurdle analysis of travel expenditure: Baby boomer seniors versus older seniors. *Tourism Management*, 30(3):372–380.

Jang, S., Ham, S., and Hong, G. (2007). Food-away-from-home expenditure of senior household in the United States: a double-hurdle approach. *Journal of Hospitality & Tourism Research*, 31(2):147–167.

Jaworski, P., Durante, F., and Härdle, W. K., editors (2013). *Copulae in Mathematical and Quantitative Finance*, volume 213 of *Lecture Notes in Statistics - Proceedings*. Springer, Berlin Heidelberg.

Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors (2010). *Copula Theory and its Applications*, volume 198 of *Lecture Notes in Statistics - Proceedings*. Springer, Berlin Heidelberg.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Number 73 in Monographs on Statistics and Applied Probability. Chapman and Hall.

Kim, 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*, 29:297–307.

Kim, W., Kim, T., Gazzoli, G., Park, Y., Kim, S., and Park, S. (2011). Factors affecting the travel expenditure of visitors to Macau, China. *Tourism Economics*, 17(4):857–883.

Lee, H.-C. (2001). Determinants of recreational boater expenditure on trips. *Tourism Management*, 22(6):659–667.

Lee, H.-C. (2002). Determinants of visitor expenditures on a local festival: a Tobit analysis. *Journal of Tourism Science*, 26(1):31–46.

Leones, J., Codby, B., and Crandall, K. (1998). Tracking expenditures of the elusive nature tourists of Southeastern Arizona. *Journal of Travel Research*, 36(3):56–6.

Li, G., Song, H., and Witt, S. (2005). Recent development in econometric modeling and forecasting. *Journal of Travel Research*, 44(1):82–99.

Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research*, 24(4):835–849.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management. Concepts, techniques and tools*. Princeton Series in Finance. Princeton University Press, Princeton, NJ.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

Saayman, M. and Saayman, A. (2012). Determinants of spending: An evaluation of three major sporting events. *International Journal of Tourism Research*, 14(2):124–138.

Song, H. and Li, G. (2008). Tourism demand modeling and forecasting – a review of recent research. *Tourism Management*, 29(2):203–220.

Song, P. X.-K. (2007). *Correlated data analysis: modeling, analytics, an applications*. Springer, New York, NY.

Syriopoulos, T. and Sinclair, M. (1993). An econometric study of tourism demand: the aids model of us and european tourism in mediterranean countries. *Applied Economics*, 25:1541–1552.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):24–36.

Trivedi, P. K. and Zimmer, D. M. (2007). *Copula Modeling: An Introduction for Practitioners*. Now Publishers, Hanover, Mass.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

Wang, Y. and Davidson, M. (2010). A review of micro–analysis of tourist expenditure. *Current Issues in Tourism*, 13(6):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*, 8(5):333–346.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48:817–838.

Wu, L., Zhang, J., and A., F. (2013). Tourism participation and expenditure behaviour: Analysis using a scobit based discrete–continuous choice model. *Annals of Tourism Research*, 40:1–17.

Zhang, H., Zhang, J., and Kuwano, M. (2012a). An integrated model of tourists' time use and expenditure behaviour with self-selection based on a fully nested Archimedean copula function. *Tourism Management*, 33(6):1562–1573.

Zhang, J., Xu, L., and Fujiwara, A. (2012b). Developing an integrated Scobit-based activity participation and time allocation model to explore influence of childcare on women's time use behavior. *Transportation*, 39(1):125–149.

Table 1: Tetrachoric correlation coefficients.

	Y_1	Y_2	Y_3	Y_4	Y_5
Accommodation (Y_1)	1				
Transportation (Y_2)	0.724***	1			
Shopping (Y_3)	0.338***	0.269***	1		
Food and beverages (Y_4)	0.344***	0.144 (0.107)	-0.027 (0.791)	1	
Other services (Y_5)	0.475***	0.233***	-0.130*	0.236**	1

Notes: ***Significant at $p \leqslant 0.01$, **Significant at $p \leqslant 0.05$, *Significant at $p \leqslant 0.1$.

Table 2: Pearson's Chi–squared test statistics for testing independence among $(Y_1, Y_2, Y_3)^\top$.

		p–value
$(Y_1, Y_2)^\top$	$\chi^2_1 = 114.9235$	***
$(Y_1, Y_3)^\top$	$\chi^2_1 = 20.7766$	***
$(Y_2, Y_3)^\top$	$\chi^2_1 = 12.0671$	***
$(Y_1, Y_2, Y_3)^\top$	$\chi^2_{67} = 84346267$	***

Notes: ***Significant at $p \leqslant 0.01$, **Significant at $p \leqslant 0.05$, *Significant at $p \leqslant 0.1$.

Table 3: Description of the involved variables.

Label	Variable	Description	Mean (Median)
Dependent			
Y_1	Expenditure on accommodation	1 = the expenditure on accommodation is positive; 0 = otherwise	0.86 (1)
Y_2	Expenditure on transportation	1 = the expenditure on internal (i.e. in Italy) transportation is positive; 0 = otherwise	0.76 (1)
Y_3	Expenditure on shopping	1 = the expenditure on shopping is positive; 0 = otherwise	0.69 (1)
Explanatory			
X_1	Overall satisfaction	level of satisfaction with the destination measured on a 10-point Likert-type scale	8.47 (8)
X_2	Food and beverage expenditure	expenditure for food and beverages in Euro weighted for the number of nights and municipalities viewed during trip and expanded to the population (continuous)	$200,404$ (62,674)
X_3	Expenditure for other service	expenditure for other services in Euro weighted for the number of nights and municipalities viewed during trip and expanded to the population (continuous)	$69,675$ (0)
X_4	Alone	1 = the respondent makes the trip alone; 0 = the respondent makes the trip with someone	0.12 (0)
X_5	Employee	1 = the respondent is employee; 0 = the respondent is student, unemployed, housewife, retired, or in other occupation	0.72 (1)
X_6	German-speaking	1 = the respondent comes from Germany or Austria; 0 = the respondent comes from a foreign country, excluding Germany and Austria	0.76 (1)

Notes: In the case of dichotomous variables, the mean value is equal to the proportion of 1.

Table 4: Stepwise Logit regression coefficients.

Independent variables	Accommodation[a] (Y_1)	Transportation[b] (Y_2)	Shopping[c] (Y_3)
(X_1) Overall satisfaction	0.889 (0.16)	0.859 (0.14)	0.052 (0.02)
(X_2) Food and beverage expenditure	$> 0.001 (> 0.01)$	$< -0.001 (> 0.01)$	$> 0.001 (> 0.01)$
(X_3) Expenditure for other service	$< -0.001 (> 0.01)$	$< -0.001 (> 0.01)$	
(X_4) Alone	-1.169 (0.33)	-0.938 (0.28)	-0.417 (0.27)
(X_5) Employee	1.029 (0.28)	0.932 (0.23)	
(X_6) German-speaking	-0.746 (0.39)	-0.887 (0.32)	0.556 (0.21)
Constant	-5.524 (1.36)	-5.419 (1.18)	–

Notes: Significance level $\alpha = 25\%$. Robust standard errors in brackets (White, 1980). "–" Not included among the regressors.

[a] $N = 548$; Wald $\chi^2(6) = 63.27$; Prob $> \chi^2 = 0$; Log pseudolikelihood= -174.071; McKelvey & Zavoina's R^2=0.379.

[b] $N = 548$; Wald $\chi^2(6) = 82.12$; Prob $> \chi^2 = 0$; Log pseudolikelihood= -238.591; McKelvey & Zavoina's R^2=0.347.

[c] $N = 548$; Wald $\chi^2(4) = 89.14$; Prob $> \chi^2 = 0$; Log pseudolikelihood= -329.196; McKelvey & Zavoina's R^2=0.066

Table 5: Fitted copula parameters for different trivariate copula families with binary margins.

		IFM estimates		
Normal	$(\hat{\rho}_{12}, \hat{\rho}_{13}, \hat{\rho}_{23})^\top$	0.8732 (0.0018)	0.3935 (0.0030)	0.3360 (0.0040)
Clayton	$\hat{\theta}_{123}$	1.0806 (0.1299)	–	–
Gumbel	$\hat{\lambda}_{123}$	1.3888 (0.0515)	–	–
Frank	$\hat{\gamma}_{123}$	3.5562 (0.3513)	–	–
Joe	$\hat{\delta}_{123}$	1.4900 (0.0726)	–	–

Table 6: Vuong (1989)'s test statistics and decisions (bold symbols) for different copula choices modeling the dependence structure of $\boldsymbol{Y}$ at the significance level 10%.

	Independence (I)	Normal (I)	Clayton (I)	Gumbel (I)	Frank (I)	Joe (I)
Independence (II)	–					
Normal (II)	−6.2732 **(II)**	–				
Clayton (II)	−5.7942 **(II)**	4.5667 **(I)**	–			
Gumbel (II)	−5.1899 **(II)**	2.4518 **(I)**	−2.6198 **(II)**	–		
Frank (II)	−5.8141 **(II)**	4.5561 **(I)**	−1.6959 **(II)**	2.7801 **(I)**	–	
Joe (II)	−4.6998 **(II)**	2.8396 **(I)**	−1.8430 **(II)**	3.5756 **(I)**	−1.7854 **(II)**	–

Table 7: Outcomes pmf for $(Y_1, Y_2, Y_3)^\top$ by setting covariates at their median values $\tilde{\boldsymbol{x}}_j$.

y_1	y_2	y_3	$\hat{p}_{\boldsymbol{Y}}^{Normal}(y_1, y_2, y_3 \mid \tilde{\boldsymbol{x}}_j)$	$\hat{p}_{\boldsymbol{Y}}^{Indep}(y_1, y_2, y_3 \mid \tilde{\boldsymbol{x}}_j)$
1	1	1	0.3056	0.1900
1	1	0	0.3027	0.2738
1	0	1	0.0243	0.0846
0	1	1	0.0235	0.0935
1	0	0	0.0377	0.1218
0	1	0	0.0603	0.1347
0	0	1	0.0563	0.0416
0	0	0	0.1897	0.0599

Table 8: Outcomes pmf for $(Y_1, Y_2, Y_3)^\top$ by setting covariates at their median values $\tilde{\boldsymbol{x}}_j$ $(j \neq 4)$ and $X_4 = 1$.

y_1	y_2	y_3	$\hat{p}_{\boldsymbol{W}}^{Normal}(y_1, y_2, y_3 \mid x_4 = 1)$	$\hat{p}_{\boldsymbol{W}}^{Indep}(y_1, y_2, y_3 \mid x_4 = 1)$
1	1	1	0.0034	0.0000
1	1	0	0.0007	0.0000
1	0	1	0.0029	0.0033
0	1	1	0.0039	0.0040
1	0	0	0.0007	0.0047
0	1	0	0.0016	0.0058
0	0	1	0.3992	0.4024
0	0	0	0.5872	0.5798

Table 9: Outcomes pmf for $(Y_1, Y_2, Y_3)^\top$ by setting covariates at their median values $\tilde{\boldsymbol{x}}_j$ $(j \neq 5)$ and $X_5 = 0$.

y_1	y_2	y_3	$\hat{p}_{\boldsymbol{Y}}^{Normal}(y_1, y_2, y_3 \mid x_5 = 0)$	$\hat{p}_{\boldsymbol{Y}}^{Indep}(y_1, y_2, y_3 \mid x_5 = 0)$
1	1	1	0.4574	0.3790
1	1	0	0.3570	0.3598
1	0	1	0.0213	0.0660
0	1	1	0.0106	0.0579
1	0	0	0.0318	0.0627
0	1	0	0.0266	0.0550
0	0	1	0.0238	0.0101
0	0	0	0.0716	0.0096

Table 10: Outcomes pmf for $(Y_1, Y_2, Y_3)^\top$ by setting covariates at their median values $\tilde{\boldsymbol{x}}_j$ $(j \neq 6)$ and $X_6 = 0$.

y_1	y_2	y_3	$\hat{p}_{\mathbf{Y}}^{Normal}(y_1, y_2, y_3 \mid x_6 = 0)$	$\hat{p}_{\mathbf{Y}}^{Indep}(y_1, y_2, y_3 \mid x_6 = 0)$
1	1	1	0.2056	0.0809
1	1	0	0.1569	0.1166
1	0	1	0.0287	0.0915
0	1	1	0.0397	0.1114
1	0	0	0.0296	0.1318
0	1	0	0.0671	0.1605
0	0	1	0.1357	0.1259
0	0	0	0.3367	0.1814