COPULA MODELLING APPROACHES TO JOINTLY REPRESENT TOURISTS' TIME USE AND EXPENDITURE BEHAVIOUR

Masashi KUWANO
Graduate School of Engineering, Kobe University
Rollodai 1-1, Nada-ku, Kobe, 657-8501, Japan
E-mail: kuwano@port.kobe-u.ac.jp / Phone & Fax: +81-78-803-6686

Tatsunori NAKAGAWA
Graduate School of Engineering, Kobe University
Rollodai 1-1, Nada-ku, Kobe, 657-8501, Japan
E-mail: 121t129t@stu.kobe-u.ac.jp / Phone & Fax: +81-78-803-6686

and

Junyi ZHANG
Graduate School for International Development and Cooperation, Hiroshima University,
Kagamiyama 1-5-1, Higashi-Hiroshima, 739-8529, Japan
E-mail: zjy@hiroshima-u.ac.jp / Phone & Fax: +81-82-424-6919

ABSTRACT

Focusing on tourist decisions on time use and expenditure, this study develops a copula-based multivariate survival model to capture the interdependency between these two behaviours. In order to represent heterogeneity in tourist decisions, this study further applies latent class modelling to simultaneously incorporate different distributions and different dependence structures of time use and expenditure. An empirical analysis is carried out based on data obtained in Tottori prefecture, Japan, in 2007. Estimation results confirm the effectiveness of the suggested model from both model performance and applicability viewpoints. The results show that types of dependence structures between time use and expenditure vary among latent classes. Further, the copula parameters reveal that time use and expenditure are significantly related to each other only in the normal copula-type latent class.

Keywords: survival model, copula, interdependency, heterogeneity, latent class modelling

INTRODUCTION

Increased disposable income and leisure time, the development of transport and accommodation, and expanding economies have contributed to an increase in travel. Thus, tourism has become a highly popular leisure activity as well as a powerful and diverse industry directly associated with the growth and economic vitality of most regions. The tourism industry has major economic effects because it prompts wide-ranging travel consumption, including trip preparation, transportation, accommodation, and meal expenses. Actually, according to an estimate by the Japan Travel Bureau Foundation, tour agencies' travel sales amounted to around 6,000 billion yen in 2010. The Ministry of Land, Infrastructure, Transport and Tourism estimated travel consumption at 22.1 trillion yen, the job creation effect at 2.11 million people, and the employment inducement effect at 4.06 million people in fiscal 2009. The term employment inducement effect in this text refers to job creation effect plus ripple effect. These effects produce an economic ripple effect (production ripple effect + value-added inducing effect) amounting to 72.9 trillion yen. Thus, since tourism stimulates a wide range of consumption-related industries, it has the potential to create not only direct but also spillover economic effects, such as the production ripple effect and job creation effect. In addition, because the tourism industry can cash in on foreign demand, it has a great potential to contribute to economic revitalization in certain areas. In particular, rural tourism has attracted increasing attention in the context of multiple objectives, such as sustainable rural development and diversification of the rural economy. A resolvable understanding of tourist behaviour is now required to make effective policies to address these issues.

Meanwhile, people's scope of activity has expanded greatly in space and time as a result of the popularization of private vehicles and improvement of road maintenance. That is to say, the area of people's activities remained limited when the principal means of transit was walking, cycling, or public transportation, severely constrained by time and location. However, with the popularization of the less-constrained private vehicle, people can now expand their range of activities by choosing destinations across a wide space and fixing their schedule within the available time, reflecting individual preferences. The popularization of private vehicles plays an important role in tourism behaviour as well. Specifically, tourists can choose the time and space freely without being bound to public transport schedules and routes. Thus, hours spent and visual experiences savoured in tourist spots increase, leading to tourism demand diversification and, in turn, to tourist behaviour heterogeneity. Many researchers have emphasized the need to understand differences and similarities in individual behaviour and response patterns (Woodside et al., 1987; Cosenza and Davis, 1981; Lawson, 1991; Spoots and Mahoney, 1991; Berdychevsky et al., 2012). One objective of this study is to represent individual heterogeneity in the context of tourist behaviour.

It has long been recognised that the behaviour of tourists includes a number of decisions taken before and after their leaving home (Dellaert et al., 1998, Woodside and Dubelaar, 2002; Mak et al., 2012). These decisions relate to destination choice, travel party, duration, transportation, route choices, accommodation choice, time use, expenditure, attractions, and other holiday activities, and among other factors, affect their subsequent choices and behaviours. These are usually interrelated and show temporal and spatial variations (Spoots and Mahoney, 1991). Such interdependency implies that a tourism policy or marketing activity proposed to influence one aspect of tourist behaviour might intentionally or unintentionally affect other aspects as well. Existing studies usually modelled one aspects of behaviour using other aspects as the explanatory variables. However, it is not able to observe the cause and effect relationship among the aspects. Therefore, existing tourist behaviour studies, which assume false cause and effect relationship among behavioural elements, may cause tourist behaviour to be misunderstood. In other words, tourist decisions should be dealt with in an integrated way for a proper evaluation of tourism policies and marketing activities.

The purpose of this study is to develop a new model representing tourist behaviour. The uniqueness of the study is that two behavioural elements are taken into account: interdependency of tourist decisions and individual heterogeneity across the sample set. This study develops a copula-based multivariate survival (CMS) model to consider the interdependency of tourist decisions, especially those on time use and expenditure, which are some of the most important factors for tourism policy making and marketing (Bull, 1995; Dellaert et al., 1998; Decrop and Snelders, 2004). Moreover, because tourism preferences are not homogeneous among individuals, we model heterogeneity according to the latent class approach, assuming that tourist behaviour is substantially heterogeneous, and apply the suggested CMS model. A copula is a function that allows us to combine pre-specified univariate marginal distributions to obtain a multivariate joint distribution with the help of a limited number of dependency parameters in a more flexible way and at low computational cost (Nelsen, 2006). The effectiveness of a copula approach has been recognized and applied in the transportation field for a decade (Spissu et al., 2009; Bhat and Eluru, 2009, Kuwano et al., 2011; Zhang et al., 2012). Spissu et al. (2009) formulated a joint model of vehicle type choice and vehicle usage using copula-based joint discrete-continuous model systems, and found the suggested model is superior to a conventional joint discrete-continuous model. Bhat and Sener (2009) proposed a spatial dependence model based on a copula function to consider the dependence between the logistic error terms of different observational units.

We develop a CMS and a latent class copula-based multivariate survival (LC-CMS) model in the next section. In subsequent sections, the survey procedures used and data collected are presented, the model estimation is described, and the results are discussed. Finally, the conclusions are presented, and future research issues proposed.

MODEL DEVELOPMENT

Survival models, also kwon as duration models or hazard models, have been extensively used in other research fields such as econometrics, biostatistics, medical sciences, and industrial engineering. They are also used in studies to explain tourists' time use (Gokovali et al., 2007; Martinez-Garcia and Raya, 2008; Barros and Machado, 2010; Raya, 2012). For example, Barros and Machado (2010) analyse the length of stay of tourists in the Algarve, on the southern coast of Portugal. Several survival models is used such as the Cox as the semi-parametric survival model, the Weibull survival model and Weibull with heterogeneity. Martinez-Garcia and Raya (2008) applied survival models to analyse low-cost tourists' length of stay in Spain, using two model specifications: the Cox and log-logistics models. Gokvali et al. (2007) analysed the length of stay of tourists in Bodrum, Turkey with the Cox and Weibull survival models.

In this section, we first formulate a univariate survival duration model to analyse time use, t (i.e. total length of stay), and expenditure, m, at tourist spots. Second, we develop a copula-based multivariate survival model to simultaneously examine the variables. Then, we describe an LC-CMS model incorporating heterogeneous distributions.

A Univariate Survival Duration Model

In a duration model, time T is a continuous random variable. It measures the duration of staying in some state. Examples are time use and expenditure in tourist spots, considered in this study.

Suppose that T has a continuous probability density function $f(t)$, where t is the realization of T. The distribution function $F(t)$ gives the probability that failure time is less than or equal to t:

$$F(t) = \int_0^t f(s)ds = \Pr[T \le t]. \tag{1}$$

The hazard function can be written as a function of the distribution function $F(t)$ and the corresponding density function $f(t)$ of the variable t.

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{2}$$

Another important construct in hazard-based models is the survivor function $S(t)$, which gives the probability of survival up to t. It is related to the distribution function as follows:

$$S(t) = \Pr(T \ge t) = 1 - \Pr(T \le t) = 1 - F(t). \tag{3}$$

Since $f(t) = -dS(t)/dt$, the hazard can also be written as

$$h(t) = -\frac{d(\log S(t))}{dt}. \tag{4}$$

If the hazard is known, the survivor function can be found through

$$S(t) = \exp(-\int_0^t h(u)du). \tag{5}$$

Then, the density function of t is expressed by

$$f(t) = h(t)\exp(-\int_0^t h(v)dv). \tag{6}$$

If the distribution $f(t)$ is known, then $S(t)$ and $h(t)$ can be uniquely derived. A number of distributions have been proposed and examined in previous studies—the Exponential, Weibull, Gamma, Log-logistic, and Log-normal, for example. This study applies the Weibull distribution, the density function of which is shown as follows:

$$f(t) = \gamma t^{r-1} \exp(-\gamma\beta X)\exp\{-t^{\gamma} \exp(-\gamma\beta X)\} \quad (7)$$

where β and γ are unknown parameters that can be estimated by using the maximum likelihood method, and X is the vector of covariates (independent variables).

The Copula-Based Multivariate Survival Duration Model

Copulas

Let T and M be two random variables with marginal distribution functions $F^T(t)$ and $F^M(m)$, respectively, and let C_θ be a bi-dimensional copula. Then, the function $C_\theta(F^T(t), F^M(m))$ is a cumulative distribution function. Thus, copula functions are used to re-define joint distributions using the given margins. For any pair of scalar random variables (T, M) with a distribution function F, there exists a copula function C_θ such that

$$\begin{aligned} F(t,m) &= \Pr[T \le t, M \le m] \\ &= C_\theta[F^T(t), F^M(m)] \end{aligned}. \quad (8)$$

The copula function C_θ is unique if the marginal distribution functions $F^T(t)$ and $F^M(m)$ are continuous. Here, θ is a dependency parameter, which simultaneously characterizes the dependency between $F^T(t)$ and $F^M(m)$.

One can also define copula densities in the same way as one defines probability densities. Let the distribution of (T, M) be continuous. The differentiated form of Sklar's theorem splits the joint density of T and M, $f(t,m)$, into the product of marginal densities $f^T(t)$ and $f^M(m)$, and the copula density $c_\theta(u,v) \equiv \partial^2 C_\theta(u,v)/\partial u \partial v$ becomes

$$f(t,m) = f^T(t) f^M(m) c_\theta[F^T(t), F^M(m)]. \quad (9)$$

Because $F^T(t)$ and $F^M(m)$ have uniform distributions, $c_\theta(u,v)$ is the density of $(F^T(t), F^M(m))$ at (u,v) as well as the conditional density of $F^D(d)$ at point v given $F^t(T) = u$. To estimate unknown parameters, the following log-likelihood function is adopted:

$$\ln L(\alpha,\beta,\theta) = \ln f^T(t;\alpha) + \ln f^M(m;\beta) + \ln(c_\theta[F^T(t;\alpha), F^M(m;\beta);\theta]) \quad (10)$$

Copulas themselves can be generated in different ways, including the method of inversion, the geometric method, and the algebraic method. A rich set of copula functions have been generated using these methods. In this paper, we consider one of the simplest forms of these copulas, the Archimedean copula, which is useful for bivariate data. Archimedean copulas have been widely used because of their mathematical tractability. The Archimedean class is rich and, as a result, does not seem to be very restrictive. Nelsen (2006) gives a detailed description of copula models. From the Archimedean class especially, we will pick up Gumbel, Clayton, and Frank copulas, which present several desired properties. Moreover, the normal copula, which is identical with the multivariate normal distribution and has been used in many previous studies, is also used as a candidate distribution. We will then select the best copula based on a goodness-of-fit index.

1) Normal copula (multivariate normal distribution):

$$C_\theta(u,v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{11a}$$

2) Gumbel copula:

$$C_\theta(u,v) = \exp(-(-\ln(u))^\theta + (-\ln(v))^\theta)^{-1/\theta} \tag{11b}$$

3) Clayton copula:

$$C_\theta(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \tag{11c}$$

4) Frank copula:

$$C_\theta(u,v) = -\theta^{-1} \ln\left(1 - \frac{[1-\exp(-\theta u)][1-\exp(-\theta v)]}{1-\exp(-\theta)}\right) \tag{11d}$$

where θ represents the dependence parameter, Φ indicates the standard normal distribution function, Φ^{-1} is the functional inverse of Φ, and Φ_2 is the bivariate standardized normal distribution function with correlation θ.

Multivariate Survival Duration Model

In this paper, copula-based models are used to capture and explore the dependencies between time use and expenditure in tourist spots. For a pair (T, M) (T: time use, M: expenditure) with a joint distribution function F, the joint survival function is given by $S(t,m) = P[T > t, M > m]$. The margins of S are $S(t,-\infty)$ and $S(-\infty, m)$, which are univariate survival functions of $S^T(t)$ and $S^M(m)$, respectively.
Here, Let X and Y be continuous random variables with copula C_{TM}. Let both α and β be strictly decreasing on Ran T and Ran M, respectively. Then,

$$C_{\alpha(T)\beta(M)}(u,v) = u + v - 1 + C_{TM}(1-u, 1-v)\,. \tag{12}$$

Using equation (12) and the copula function (9), the joint survival function $S(t,m)$ is given as follows:

$$
\begin{aligned}
S(t,m) &= 1 - F^T(t) - F^M(m) + F(t,m) \\
&= S^T(t) + S^M(m) - 1 + C_\theta(F^T(t), F^M(m)) \\
&= S^T(t) + S^M(m) - 1 + C_\theta(1 - S^T(t), 1 - S^M(m)) \\
&= \hat{C}_\theta(S^T(t), S^M(m))
\end{aligned}
\tag{13}
$$

where function $\hat{C}$ is a survival copula of T and M .
The joint distribution formed shown above is used to derive the joint time use and expenditure data.

Applying Latent Class Modelling

We model heterogeneity in individual tourist behaviour using the latent class approach, which is known to provide an attractive platform for modelling unobserved heterogeneity. Thus, we were able to find two more groups that show differences in the sensitivities to choice probability. In the latent class approach, Q_{ik}, the latent class membership probability that individual *i* belongs to latent class k ($k = 1, 2, ..., K$), with a specific tourist behaviour mechanism, is first specified. The likelihood of individual behaviour contingent upon different tourist behaviour mechanisms is weighted by membership probabilities.
Membership probability may be influenced by various factors. Generally speaking, socio-demographic variables are adopted to represent such probabilities in the existing literature (Bucklin and Gupta 1992; Gupta and Chintagunta 1994). Moreover, various functional forms have been proposed to represent membership probability. This study adopts the most convenient form, the multinomial logit-type function (Zenor and Srivastava 1993; Swait and Sweeney 2000) and defines the latent class membership probability in equation (14).

$$
Q_{ik} = \frac{\exp(\sum_n \theta_n D_n)}{\sum_{k=1}^{K} \exp(\sum_q \theta_{kq} D_{kq})}
\tag{14}
$$

where θ_{kq} is the parameter, D_{kq} represents the explanatory variables of individuals and households, and q denotes an item in the explanatory variables.
In order to estimate the probabilities, it is necessary to set the parameters θ_{kq} to zero for a pre-specified latent class K (a reference latent class), which can be any candidate latent class. Under the latent class modelling framework, each class has the same form of likelihood for the copula-based multivariate survival model as shown in equation (10). To differentiate from equation (10), the likelihood $L_{i|k}$ of latent class k is rewritten in equation (15). Then, the individual logarithm likelihood function can be defined by summing up the products of class membership probability (Q_{ik}) and class likelihood for the copula-based multivariate survival model over all classes. This is summarized in equation (16).

$$\ln L_{i|k}(\alpha_k, \beta_k, \theta_k) = \ln f^T(t_{i|k}; \alpha_k) + \ln f^M(m_{i|k}; \beta_k) + \ln(c_{\theta_k}[F^T(t_{i|k}; \alpha_k), F^M(d_{i|k}; \beta_k); \theta_k]) \quad (15)$$

$$\ln L = \sum_i \ln\left\{\sum_k \left(Q_{ik} L_{i|k}\right)\right\} \quad (16)$$

where $L_{i|k}$ is the likelihood of households in latent class k.

DATA

In order to examine the effectiveness of the proposed model, we applied a questionnaire survey data set collected in Tottori prefecture, Japan, in 2007. Tottori prefecture, located along the Sea of Japan, is the least populous prefecture in Japan. It is famous for its sand dunes, beautiful beaches, a variety of seasonal scenery, hot springs, and seafood. The number of tourists was 17.545 million in 2011, a 6.9% decline from 2010.

The questionnaire was designed to collect detailed touring activity information from tourists. We asked participants to respond to questions on tourist sites visited, departure and arrival time, money spent at each site, travel mode, and so on. In total, 6,585 questionnaires were randomly distributed among tourists at major attractions and tourist information offices in the four seasons of 2007. Only 761 respondents returned the questionnaires with valid answers. Survey results showed that 56% of the tourists were day trippers, while 44% stayed for one or more nights in Tottori prefecture. In this study, we focused only on day trippers. To use data from those who were not day trippers requires another behavioural choice: whether to stay another night or not. This is beyond the scope of this study. With the above data screening, we were left with valid data from 349 day trippers for this analysis. The questionnaire asked respondents to report time use (unit: minutes) and expenditure (unit: Japanese yen) at each single tourist site. To estimate the models, the information collected with respect to each tourist site had to be aggregated, which means the stay time and expenditure were obtained by summing the time and money spent at all sites visited on a tour.

Figure 1 shows sample distributions related to individual characteristics of participants. Some respondents are under 29 years, and both male and female participants are in almost the same proportion. Most are residents of Tottori prefecture, followed by Hyogo, Okayama, and Osaka prefectures. This means 43% of respondents represent tourists from inside the prefecture. Figure 2 illustrates the sample distribution of behavioural attributes. About 78% travel by car, and more than 90% travel with somebody. The average number of visiting sites is 7.21, and more than 20% of the respondents visit 10 or more sites in a day trip.

Figures 3 and 4 show the sample distribution of time use and expenditure, respectively. Average time use is 3.44 hours, and most time use durations are 3 hours and less than 30 minutes. On the other hand, as shown in figure 5, average expenditure is 14,346 yen, and around 40% of respondents spend from 6,001 to 12,000 yen.

Figure 5 shows a scatter diagram of time use and expenditure. The correlation coefficient between time use and expenditure is 0.224. The results show a weak correlation between the two behavioural elements for the entire sample.

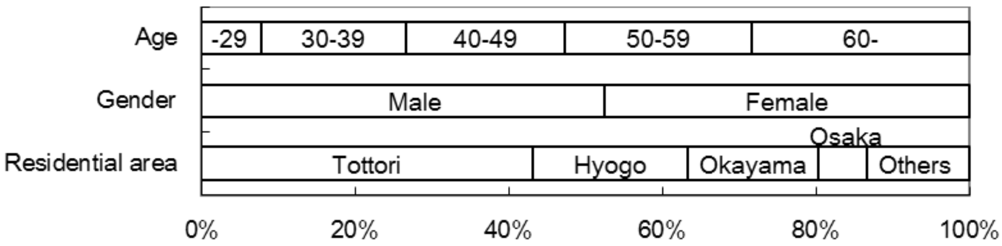

Figure 1 – Sample distribution (Individual characteristics)

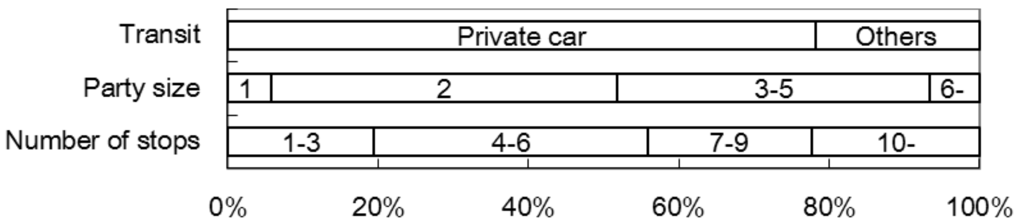

Figure 2 – Sample distribution (behavioural attributes)

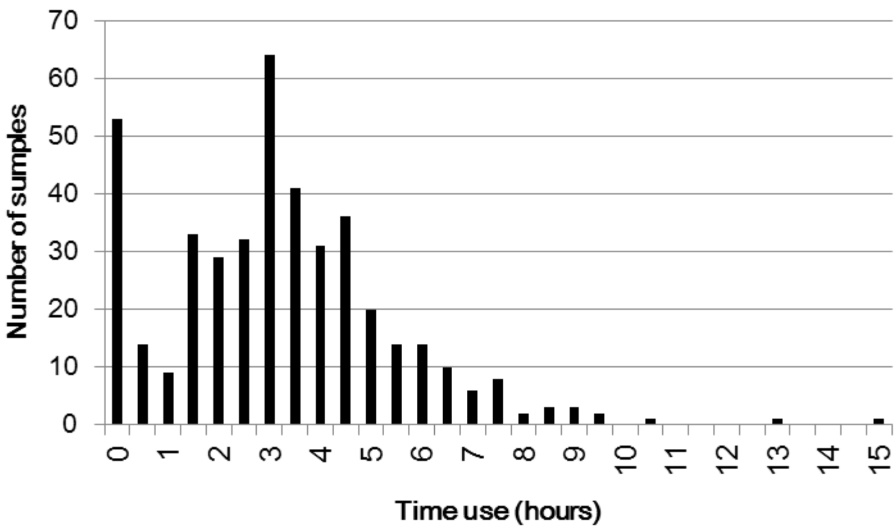

Figure 3 – Distribution of time use

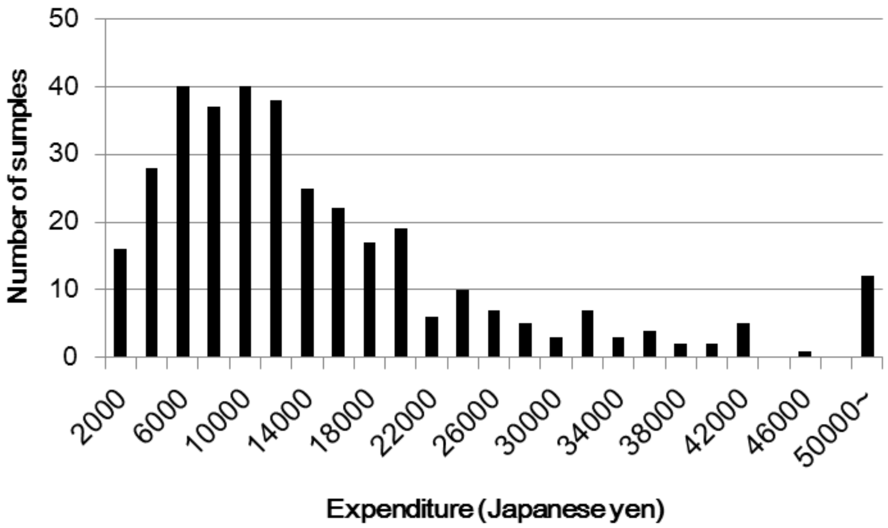

Figure 4 – Distribution of expenditure

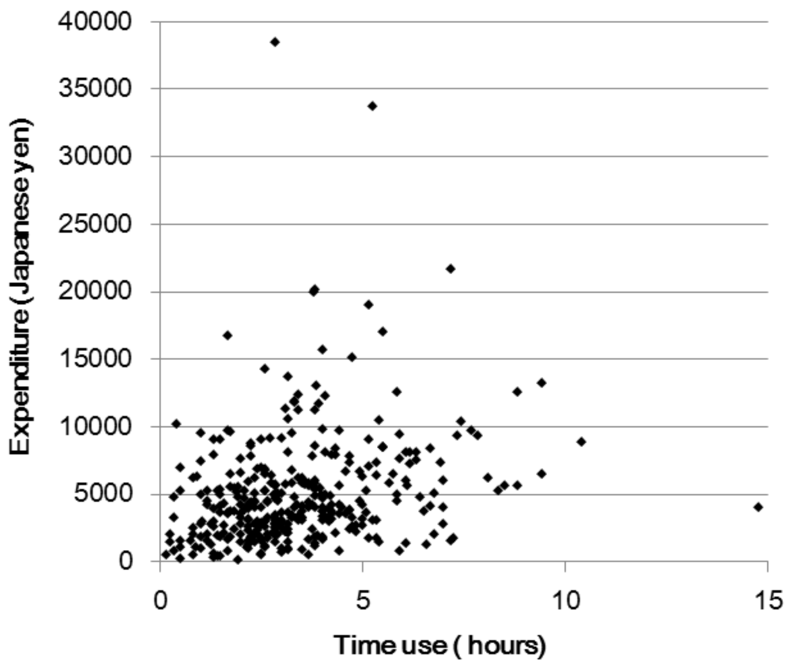

Figure 5 – Scatter diagram of time use and expenditure

MODEL ESTIMATION

Performance of the Suggested Model

Our first step was to identify a suitable model that can factor in both interdependency and independence between time use and expenditure. The Bayesian information criterion (BIC) index, which is one of the goodness-of-fit indices of a statistical model, was used to select candidate models. The BIC index includes a penalty term for the number of parameters in the model, which is larger than that of the Akaike information criterion (AIC). Therefore, BIC prefers a model with fewer parameters than AIC does. The BIC is defined as follows:

$$BIC = -\ln(Lc) + 0.5 \times p \times \ln(N) \tag{17}$$

where $\ln(Lc)$ is the log-likelihood of estimation at convergence, p is the number of parameters, and N is the number of observations.
This study examined four types of copula functions: normal, Gumbel, Clayton, and Frank. BIC values are tabulated for both an ordinary model and the proposed model (see table 1). The BIC values of the proposed CMS models are lower than that of the ordinary model, which assumed independence between time use and expenditure. The results show the effectiveness of the proposed model. Moreover, a comparison of the CMS models' BIC values indicated that the Frank copula model provides the best goodness of fit, followed by the Gumbel, normal, and Clayton copula models.

Table 1 – Comparisons of BIC values

Copula type	BIC
Normal copula	2199.50
Gumbel copula	2196.54
Clayton copula	2214.36
Frank copula	2192.77
Without copula (conventional model)	2218.89

In the second model estimation step, we examine, using the latent class approach, how effective the model is in considering heterogeneity. In this approach, the number of classes is determined by several trials. The candidate models are estimated under the given number of latent classes; then a suitable model is chosen based on the goodness of fit. In this study, three latent classes were used. However, we could not observe the third latent class, which means two latent classes are good enough for this case study.
We tabulated BIC values with two latent class models. Here, different copula functions were applied for each latent class. The BIC value of the model with the combination of normal copula for latent class 1 and Clayton copula for latent class 2 is the lowest across all trials. Moreover, the BIC value is higher than that of the CMS model without latent class in table 1. It is also found that the interdependency structure of the Frank copula with the homogeneity assumption is represented by a combination of two other copula types.

Table 2 – Comparison of BIC (LC-CMS)

Copula type		Latent class 2			
		Normal	Gumbel	Clayton	Frank
Latent class 1	Normal	2301.18	2312.35	2297.78	2309.60
	Gumbel	—	2297.96	2313.50	2298.24
	Clayton	—	—	2314.46	2308.44
	Frank	—	—	—	2309.88

Table 3 compares expected stay time and expenditure for each model. From the comparison, the model accuracy of LS-CMS, with an observed stay time of 3.44 hours and expenditure of 14,346 yen, is the highest. Based on the above discussion, the estimated model provides empirical evidence for the effectiveness of the LS-CMS model in terms of both goodness of fit and model accuracy. Therefore, the LS-CMS model with the heterogeneity assumption should be considered for tourist behaviour analysis, including interdependency between stay time and expenditure.

Table 3 – Comparison of model accuracy

	Expected time use	Expected expenditure
Observed value	3.44	14,346
Without copula (conventional model)	3.29 (4.3%)	12,697 (11.5%)
CMS model (with Frank copula)	3.31 (3.6%)	12,813 (10.6%)
LS-CMS model (with normal and Clayton copulas)	3.45 (0.36%)	13,442 (6.3%)

Note: Within parentheses are percentages of error from observed values.

Estimation Results of the LC-CMS Model

The estimation results of the LC-CMS model, which provides the highest goodness of fit, are shown in table 4.

Note that one of the membership probabilities should be set to zero for identification purposes. In this study, the membership probability of latent class 2 is set to zero. According to the estimation results of membership probability, the parameters of male dummy and access time within each destination are positive. Therefore, males as well as tourists who visit a number of spots tend to belong to latent class 1.

The estimation results of copula parameters show a positive interdependence between stay time and expenditure, since the normal copula parameter is statistically significant and positive in latent class 1. This result indicates that expenditure increases with stay time, and the value of the interdependency parameter is higher than 0.224, which is the correlation coefficient between time use and expenditure of all samples as mentioned in figure 3. On the other hand, the copula parameter is not statistically significant in latent class 2, which means this class does not have interdependency between stay time and expenditure. As mentioned in the previous subsection, the Frank copula shows the highest goodness of fit in the CMS model without heterogeneity. Dependence structure in the tails of the Frank copula tends to be relatively weak compared to the normal copula, and the strongest dependence is centred in the middle of the distribution, which suggests that the Frank is most appropriate for data that exhibit weak tail dependence. However, with the heterogeneity assumption based on the latent class model, the interdependency structure is represented by a combination of the normal copula with relatively high correlation and the Clayton copula with poor correlation. The results indicate that ignoring heterogeneity leads to a misunderstanding of the interdependency structure and the strength of the association.

In reference to the parameters in the survival function related to time use in latent class 1, the parameters for age and residential location are negative, and the parameters for size of travel party and access time within each destination are positive. Therefore, older individuals as well as visitors from a long distance and longer travel time within destinations are associated with a longer stay at the destination. The parameters for the survival function of expenditure in latent class 1, age, size of travel party, travel model dummy, and access time within each destination, are significantly positive. Therefore, older individuals, visitors with a companion, visitors by train, and visitors who visit a number of spots tend to spend more money than others. However, both parameters in the survival function, stay time and expenditure, are less significant in latent class 2.

Figures 6 and 7 illustrate the probability density projection for time use and expenditure, respectively. The highest probability is observed with a time use of about 3.0 hours and an expenditure of about 1,000 yen, while the expected time use estimated by the model is 3.58 hours and the expenditure 13,949 yen, in latent class 1. On the other hand, in latent class 2, the highest probability is observed at about 2.0 hours and 3,500 yen, while the expected time use and expenditure are 3.26 hours and 12,709 yen, respectively. It is clear that the probability density distributions of time use and expenditure in latent class 1 are significantly different from those in latent class 2. Thus, heterogeneity is particularly influential on time use and expenditure.

Table 4 – Estimation results of the LC-CMS model

	Latent class 1		Latent class 2	
Explanatory variables	Parameter		Parameter	
Effects on latent membership probability				
Male dummy	0.527	*		
Size of travel party (no. of members)	-0.065			
Access time from home to first destination (minutes)	-0.203			
Access time within each destination (minutes)	0.647	**		
Constant term	-1.001	*		
Interdependency between stay time and expenditure				
Copula type	Normal copula		Clayton copula	
Copula parameter	0.320	**	0.082	
Stay time				
Geometry parameter	0.067	**	0.284	*
Scale parameter	1.998	**	1.575	**
Age	-0.090	+	0.001	
Size of travel party (no. of members)	0.125	**	0.080	
Residential location dummy (Tottori:1, other: 0)	-0.253	+	-0.019	
Access time from home to first destination (minutes)	-0.072		-0.012	
Access time within each destination (minutes)	0.067	**	0.192	*
Expenditure				
Geometry parameter	6.399	**	7.140	
Scale parameter	1.572	**	1.219	**
Age	0.176	**	0.135	
Size of travel party (no. of member)	0.379	**	0.304	**
Travel mode dummy (rail: 1, other: 0)	1.325	**	1.366	
Access time within each destination (minutes)	0.127	**	0.167	*
Residential location dummy (Tottori:1, other: 0)	0.053		0.349	
Probability of belonging to the class	0.591		0.409	
Converged log-likelihood L(β)	-1046.43			
BIC	2297.78			

Note: **significant at the 1% level, *significant at the 10% level

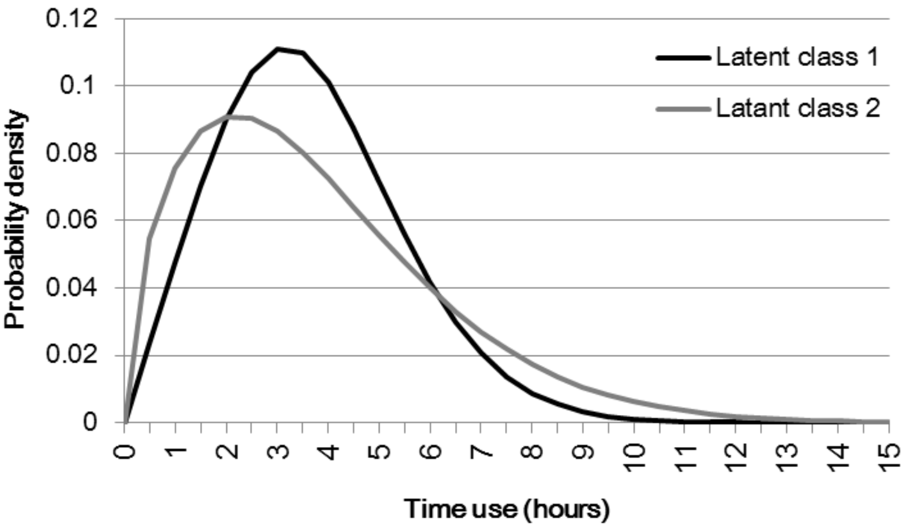

Figure 6 – Scatter diagram of time use and expenditure

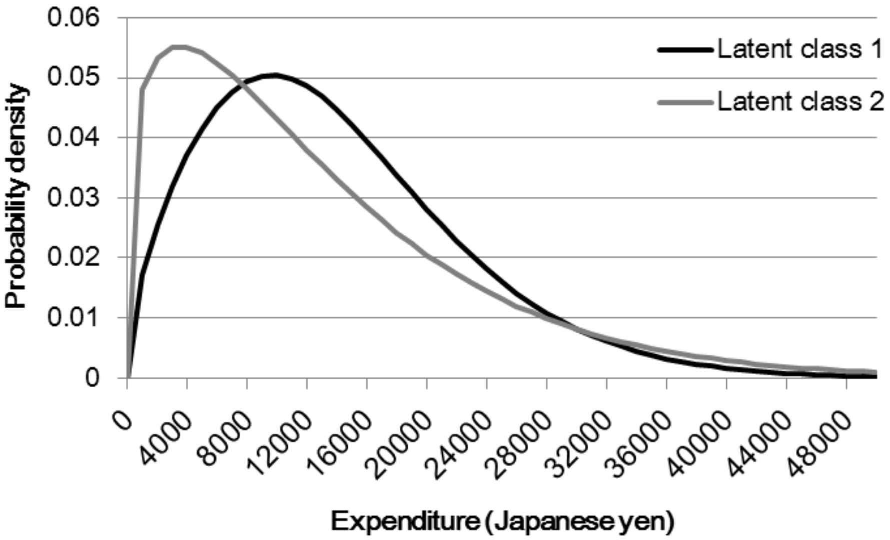

Figure 7 – Scatter diagram of time use and expenditure

CONCLUSIONS

Considering that tourist decisions jointly influence various behaviours, finding modelling methods that can deal with them in a consistent way has become necessary. Moreover, with the spread of private vehicles, the behaviour of tourists in destinations has become increasingly diverse since they can modify their activities to suit their preferences. Therefore, developing an analytical approach to jointly consider diverse tourist behaviours and individual heterogeneity could have various implications for credible and balanced policy-making decisions. We developed an integrated model to jointly represent time use and expenditure behaviour at destinations as well as individual heterogeneity. Our most important contribution to the tourism literature is that the copula and latent class approaches are jointly introduced and empirically examined. Copulas are functions that integrate multivariate distribution functions with their one-dimensional margins by introducing a limited number of dependency parameters, where the one-dimensional margins are not necessarily the same. Furthermore, because tourists are expected to have different preferences, this study applied the latent class approach to explain heterogeneity.

Data collected from tourists (one-day trippers: 349 respondents) visiting Tottori prefecture, Japan, in 2007 were used to confirm the effectiveness of the LC-CMS model developed. The effectiveness of the suggested model in this study has been confirmed empirically from the perspectives of both model goodness of fit and accuracy. The estimation results of membership probabilities clarified the existence of two groups with different time use and expenditure types. The first group of tourists, mostly male, visit more tourist sites and spend relatively longer time and more money at the destination. Furthermore, this group shows interdependency between time use and expenditure, which means the higher the expenditure, the greater the time use. Moreover, their time use and expenditure behaviour could be described by explanatory variables defined in this study since they are statistically significant for time use and expenditure survival functions. The other group spends relatively shorter time and less money in tourist sites. This group shows poor correlation between time use and expenditure, which means they decide their behaviour independently. Further, representing their behaviour with the explanatory variables we used is difficult. These results empirically confirm that ignoring interdependency and heterogeneity could lead to biased estimates of tourism behaviour. Moreover, from the perspective of an analytical method, while this study only focuses on two behavioural elements and two latent classes, it shows that the suggested model can flexibly represent multi-dimensional decision-making mechanisms and multi-peak distributions.

There are still some unresolved issues. First, although this study proposed an LS-CMS model, the explanatory variables were constrained by limited information from the questionnaire survey. Therefore, we failed to describe the behaviour of latent class 2 fully, which means the variables are less significant. Second, policy variables were not included in the model as explanatory variables. We need to continue our investigation and collect further data to obtain a closer estimate of tourist behaviour in order to assist policy making in a more comprehensive way. Third, we need to clarify how to extend the developed model to simultaneously represent more behavioural aspects, such as destination choice, travel mode choice, and activity participation choice. Since tourist behaviour usually involves group decisions, it is worth applying the copula modelling approach to represent intra-participant interaction. Because these behaviours are multi-dimensional, copula models could be expected to play increasingly important roles in explaining complicated decisions.

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

Berdychevsky, L., Poria, Y., and N. Uriely (2012). Sexual behaviour in women's tourist experiences: Motivations, behaviours, and meanings, Tourism Management, (in press).

Bhat, C.R., and N. Eluru (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling, Transportation Research Part B: Methodological, 43(7), 749-765.

Bucklin, R.E., and S. Gupta (1992). Brand choice, purchase incidence, and segmentation: an integrated modeling approach. Journal of Marketing Research, 29(2), 201-215.

Bull, A. (1995). The economics of travel and tourism (2nd ed.), New York, NY:Longman.

Cosenza, R.M., and D.L. Davis (1981). Family vacation decision making over the family life cycle: A decision and influence structure analysis, Journal of Travel Research, 20(2), 17-23.

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

Dellaert, B.G.C., Ettema, D.F., and C. Lindh (1998). Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists' sequential choices of travel components. Tourism Management, 19(4), 313-320.

Gokovali, U., Bahar, O., and M. Kozak (2007). Determinants of length of stay: a practical use of survival analysis, Tourism Management, 28(3), 736-746.

Gupta, S., and P.K. Chintagunta (1994). On using demographic variables to determine segment membership in logit mixture models, Journal of Marketing Research, 31(1), 128-139.

Japan Travel Bureau Foundation (2012). http://www.jtb.or.jp/index_english.html.

Kuwano, M., Fujiwara, A., Zhang, J., and M. Tsukai (2011). Joint modeling of vehicle holding duration and traveling distance with a copula-based multivariate survival model, Journal of Asian Transport Studies, 1(3), 303-317.

Lawson, R. (1991). Patterns of tourist expenditure and types of vacation across the family life cycle, Journal of Travel research, 29(4), 12-18.

Mak, A.H.N., Lumbers, M., Eves, A., and R.C.Y. Chang (2012). Factors influencing tourist food consumption, International Journal of Hospitality Management, 31(3), 928-936.

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

Ministry of Land, Infrastructure, Transport and Tourism (2012). http://www.mlit.go.jp/en/index.html.

Nelsen, R.B. (2006). An introduction to copulas (2an ed.) New York: Springer-Verlag.

Raya, J.M. (2012). Length of stay for triathlon participants in the challenge Maresme-Barcelona: A survival approach, Journal of Sport & Social Issues, 36, 89-105.

Spissu, E., Pinjari, A., Pendyala, R., and Bhat C.R. (2009). A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel, Transportation, 36, 403–422.

Spotts, D.M., and E.M. Mahoney (1991). Segmenting visitors to a destination region based on the volume of their expenditures. Journal of Travel Research, 29, 24-31.

Swait, J., and J.C. Sweeney (2000). Perceived value and its impact on choice behavior in a retail setting, Journal of Retailing and Consumer Services, 7(2), 77-88.

Woodside, A.G., and C. Dubelaar (2002). A general theory of tourism consumption systems: a conceptual framework and an empirical exploration. Journal of Travel Research, 41(2), 120-132.

Woodside, A.G, Cook, V.J., and W. Mindak (1987). Profiling the heavy traveller segment, Journal of Travel Research, 25, 9-14.

Zenor, M. J., and R. Srivastava (1993). Inferring market structure with aggregate data: a latent segment logit approach, Journal of Marketing Research, 30, 369-379.

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