TOURISM GENERATION ANALYSIS BASED ON A SCOBIT MODEL*

Lingling, WU**, Junyi ZHANG***, and Akimasa FUJIWARA****

1. Introduction

Tourism generation (or participation) is one of the most important aspects in tourism demand forecasting. Research concerning non-participation behavior offers a means of assessing latent demand for tourism, which is essential to tourism forecasting, policymaking, and planning. In addition, a better understanding of the constraints that tourists confront and how these constraints influence their choice behavior is useful to implement policies to encourage tourism participation. This study attempts to analyze individual's decision on whether to go on vacation or not. Since choice of participation in tourism activity can be treated as a binary choice, the binary logit model has been widely applied. However, the logit model assumes that the sensitivity of individuals' choice probabilities to changes in explanatory variables is highest for those who have indifferent preferences over participation and non-participation. This sensitivity is because the logistic density functions are symmetric about zero. This is also true for the probit model. However, this assumption has not been widely tested when applying these models (Zhang and Timmermans[1),2)], 2010a,b). In this study individual's choice of tourism participation is studied based on a Scobit model, which includes a skewness parameter to relax such kind of assumption. The empirical application is carried out using the data stemmed from a survey conducted in Japan based on a telephone interview in 2002.

2. Review

In the last decade, a growing body of research has emerged regarding tourism participation. Existing approaches include: constraint models and microeconomic models.

Constraint models define constraints as factors that are assumed to prohibit participation in tourism (Jackson[3)] 1991). In these models, constraints are classified into three categories: intrapersonal, interpersonal, and structural. These constraints are ordered sequentially so that each level of a constraint must either not exist or be overcome before going on to the next level (Crawford et al.[4)], 1991).

A different approach to participation is microeconomic model. These are utility maximizing choice models in which tourists' choice of participation is influenced by several factors (Fleischer and Seiler[5)], 2002; Stemerding, Oppewal and Timmermans[6)], 1999; Nicolau and Mas[7)], 2005; Mergoupis and Steuer[8)], 2003; Hellstrom[9)], 2006; Melenberg and Soest[10)], 1996; Alegre and Mateo[11)]; 2010). In these studies, factors such as income, age, gender, education level, marital status, health condition, number of children, household size, residential area, traffic condition are found to be influential to tourism participation. However, most of these researches adopted binary logit model or probit model to deal with participation choice under the assumption that the sensitivity of individuals to changes in explanatory variables is highest for those who have indifferent preferences over participation and non-participation.

3. Methodology

The utility of participation U_{ij} can be described as:

$$U_{ij} = V_{ij} - \varepsilon_{ij} = \sum_s \gamma_{js} Z_{ijs} - \varepsilon_{ij}$$
$$Y_{ij} = \begin{cases} 1 & U_{ij} \geq 0 \\ 0 & otherwise \end{cases}$$

Where, Y_{ij} indicates choice of participation (1: participation; 0: non-participation), V_{ij} refers to the deterministic term, Z_{ijs} indicates the sth explanatory variable, γ_{js} is the parameter of Z_{ijs}, and ε_{ij} is an error term.

Then, the probability that individual chooses to participate in tourism is:

$$P_{ij}(Y_{ij}=1) = P(V_{ij} > \varepsilon_{ij}) = F(V_{ij})$$

Here, F indicates the distribution function of error term ε_{ij}. Let f be probability density function of ε_{ij}. Marginal effect of Z_{ijs} on the participation probability P_{ij} is:

*Keywords: Tourism and leisure behavior, Activity participation, Scobit model, marginal effects
Graduate School for International Development and Cooperation, Hiroshima University, 1-5-1, Kagamiyama, Higashi-Hiroshima City, 739-8529, Japan, Phone & Fax: +81-82-424-6919; **wuling1017@gmail.com; ***zjy@hiroshima-u.ac.jp; ****afujiw@hiroshima-u.ac.jp)

$$\partial P_{ij}(Y_{ij}=1)/\partial Z_{ijs}=f(\sum_s \gamma_{js}Z_{ijs})\gamma_{js}$$

In existing research, ε_{ij} is assumed to follow either a normal distribution or a Weibul distribution. In both case, marginal effect will reach a maximum when V_{ij} is equal to zero. This implies that the change of variable of Z_{ijs} will have its greatest effect on individuals with the value of V_{ij} equal to zero, or the probability P_{ij} equal to 0.5.

However, since there might be heterogeneous initial probability among individuals, in reality, the logit or probit model would result in a misspecification. To overcome this problem, this study adopts an alternative distribution function.

$$F(\varepsilon_{ij};\alpha)=\frac{1}{(1+\exp(-\varepsilon_{ij}))^{\alpha}}$$

Then, the probability that an individual chooses to participate in tourism can be derived as follows.

$$P_{ij}(Y_{ij}=1)=F(V_{ij};\alpha)=\frac{1}{(1+\exp(-V_{ij}))^{\alpha}}$$

This model is called as Scobit model, named by Nagler[12] (1994). Here α is skewness parameter. When α is equal to 1, the model will become logit model. In the Scobit model, marginal effect of Z_{ijs} on the participation probability P_{ij} is:

$$\partial P_{ij}(Y_{ij}=1)/\partial Z_{ijs}=\alpha\exp(-\sum_s \gamma_{js}Z_{ijs})(1+\exp(-\sum_s \gamma_{js}Z_{ijs}))^{-\alpha-1}\gamma_{js}$$

It can be noticed that the marginal effect of Z_{ijs} is influenced by α, which does not assume that the sensitivity of choice probabilities to changes in explanatory variables is highest for those individuals with initial probability of 0.5. Since the initial probability may be different across individuals, α can be further defined using some individual attributes z_{iq}.

$$\alpha_i=\exp(\sum_q \theta_q z_{iq})$$

Here, θ_q is the parameter of variable z_{iq}. The exponential function is adopted to meet the requirement that α>0. To estimate the Scobit model, the following log-likelihood function can be obtained.

$$LogL=\sum_{i=1}^{N} In(p_{i1}^{\delta_{i1}}\ p_{i2}^{(1-\delta_{i1})})$$

Where, N indicates the total number of individuals, and δ_{n1} is dummy variable, which is equal to 1 when individual chooses to participate in tourism, otherwise 0.

4. Model Estimation and Results

(1) Data

The data used in this study comes from a survey conducted in Japan based on a telephone interview in 2002. The survey collected information about individuals' tourism participation in a year period, individual/household characteristics and tourism preference. The valid sample size is 1000 individuals, 65.7% of them participate in tourism activities in one year period. Individuals' attributes are summarized in Table 1.

(2) Explanatory Variables

Individual's age, employment status, annual income, household size, vacation system are adopted as explanatory variables. In the Scobit model with heterogeneous α, age, household size are used to explain skewness parameter α.

(3) Model Estimation Results

To compare the difference of logit and Scobit model, we first estimated the Scobit model with homogeneous α and the binary logit model. However, the results do not show much difference between the two models. For the Scobit model with homogeneous α, it is still expected that different people may have different value of α. Therefore, the whole sample is divided into two segmentations based on residential area in this study. Segmentation 1 is individuals from large cities and segmentation 2 is the rest ones. The Scobit model with homogeneous α and the binary logit model are estimated for each segmentation, respectively. The results of the estimation are shown in Table 3. It can be seen that parameters of individual's age, employment status, vacation length are statistically significant at 95% or 99% level. The positive parameters of employment status and vacation length indicate that individuals who are employed and who have longer vacation are more likely to participate in tourism. The negative parameter of age means that younger people are more likely to participate in tourism. To test whether Scobit model outperforms logit model or not, χ^2 statistic are calculated for each segmentation. It is found that for segmentation 1, the accuracy of the Scobit model is higher than that of the logit model at 90% level.

Table 1. Summary of Data Characteristics

Gender	Male	50.5%
	Female	49.5%
Age	10-20 years old	4.0%
	20-30 years old	16.0%
	30-40 years old	16.0%
	40-50 years old	16.0%
	50-60 years old	16.0%
	60-70 years old	16.0%
	Over 70 years old	16.0%
Employment Status	Employed	52.7%
	Unemployed	47.3%
Number of Household member	1 member	4.5%
	2 members	22.8%
	3 members	20.8%
	4 members	28.2%
	More than 4 members	23.7%
Income	<4 million yen	62.7%
	4-8 million yen	33.8%
	>8 million yen	3.5%

Table 2. Explanatory Variables

Explanatory variables	Description
Individual and Household Socio-demographics	
Age	Age of individual
Employment (dummy	1 if employed, 0 otherwise
Income (million yen)	Annual income (not categorized)
Household size	Number of household members
Vacation system	
Vacation	The longest vacation he can get in a year

Table 3. Model Estimation Results of Scobit and Logit Models (Homogeneous Skewness Parameter)

Explanatory variable	Logit model		Scobit model	
	Parameter	t-value	Parameter	t-value
Segmentation 1				
Income	-0.267	-0.769	-0.293	-1.810
Employment	1.061	2.829	0.859	1.643
Vacation	0.036	2.783	0.027	1.350
Skewness parameter			0.375	1.612(0)
				2.677(1)
Age	-0.102	-2.167	-0.295	-1.810
Household size	0.094	1.076	0.034	0.284
Sample size	290			
Initial log-likelihood	-201.01			
Converged log-likelihood	-174.53		-172.77	
McFadden's Rho-squared	0.132		0.140	
Adjusted McFadden's Rho-squared	0.107		0.111	
χ^2 statistic to test whether Scobit outperforms logit or not	3.52 > 2.71 (critical value: df=1, 90% level)			
Segmentation 2				
Income	0.174	0.837	0.173	0.838
Employment	0.745	3.769	0.763	3.350
Vacation	0.026	3.376	0.026	2.980
Skewness parameter			1.050	3.366(0)
				0.160(1)
Age	-0.071	-2.278	-0.064	-1.226
Household size	0.041	0.761	0.045	0.781
Sample size	710			
Initial log-likelihood	-492.13			
Converged log-likelihood	-448.55		-448.54	
McFadden's Rho-squared	0.089		0.089	
Adjusted McFadden's Rho-squared	0.078		0.076	
χ^2 statistic to test whether Scobit outperforms logit or not	0.02 < 2.71 (critical value: df=1, 90% level)			

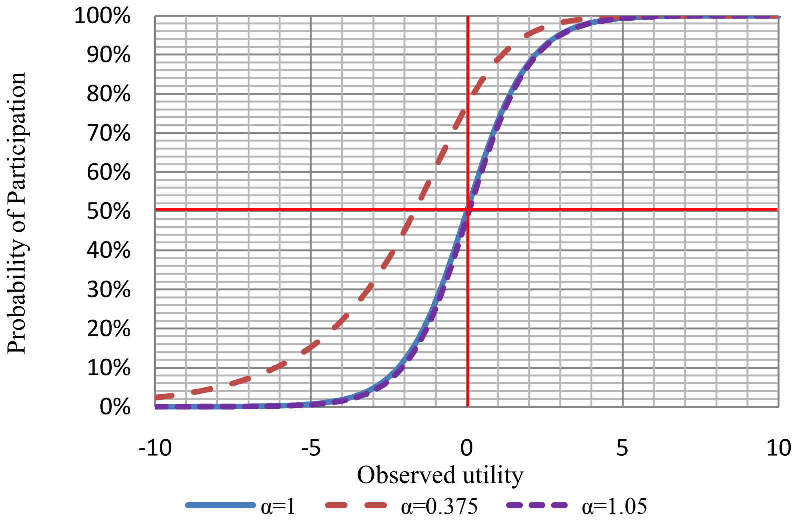

Figure 1. Probability of Participation under different value of α

In terms of skewness parameter, when it is equal to 1, the Scobit model becomes the logit model. Two types of t-test are conducted: one corresponds to the null hypothesis α=0 and the other to α=1. From the result, we can see that for segmentation 1 α (=0.375) is not different from 0 but significant different from 1 at 99% level, while for segmentation 2 it is different from 0 but not significant different from 1. Figure 1 shows the probability of participation under these two values of α. It can be noticed that when the value of α is 0.375, the participation probabilities have a very different curve from the logit curve. In this case, the individuals who are most sensitive to the utility change are those who have initial probability of 0.62. While, the curve for segmentation 2 is almost the same as the logit curve. From these curves we can also see that if we want to increase individual's participation probabilities in segmentation 1 from a certain level below 50% to 50%, greater change of observed utility should be made compared with segmentation 2, which means that it is more difficult to encourage people in segmentation 1 to participation in tourism by increasing the observed utility.

Table 4. Model Estimation Results of Scobit and Logit Models (Heterogeneous Skewness Parameter)

Explanatory variable	Logit model		Scobit model	
	Parameter	t-value	Parameter	t-value
Segmentation 1				
Income	-0.267	-0.769	-0.285	-0.838
Employment	1.061	2.829	0.984	2.837
Vacation	0.036	2.783	0.033	2.797
			(Skewness parameter)	
Age	-0.102	-2.167	-0.069	-2.022
Household size	0.094	1.076	0.084	1.169
Sample size	290			
Initial log-likelihood	-201.01			
Converged log-likelihood	-174.53		-174.91	
McFadden's Rho-squared	0.132		0.130	
Adjusted McFadden's Rho-squared	0.107		0.107	
Segmentation 2				
Income	0.174	0.837	0.168	0.827
Employment	0.745	3.769	0.726	3.906
Vacation	0.026	3.376	0.024	3.501
			(Skewness parameter)	
Age	-0.071	-2.278	-0.050	-2.192
Household size	0.041	0.761	0.030	0.691
Sample size	710			
Initial log-likelihood	-492.13			
Converged log-likelihood	-448.55		-448.82	
McFadden's Rho-squared	0.089		0.088	
Adjusted McFadden's Rho-squared	0.078		0.078	

Next, we estimated the Scobit model with heterogeneous skewness parameter, which is defined as an exponential function of age and household size (see Table 4). The Adjusted McFadden's Rho-squared values are same for the two models. The parameters of individual's age, employment status, vacation length are statistically significant at 99% level.

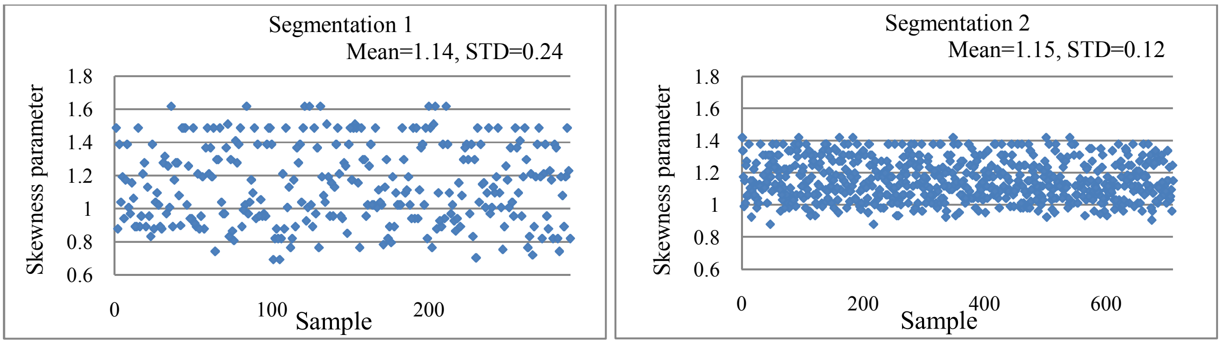

Figure 2. Distribution of skewness parameter

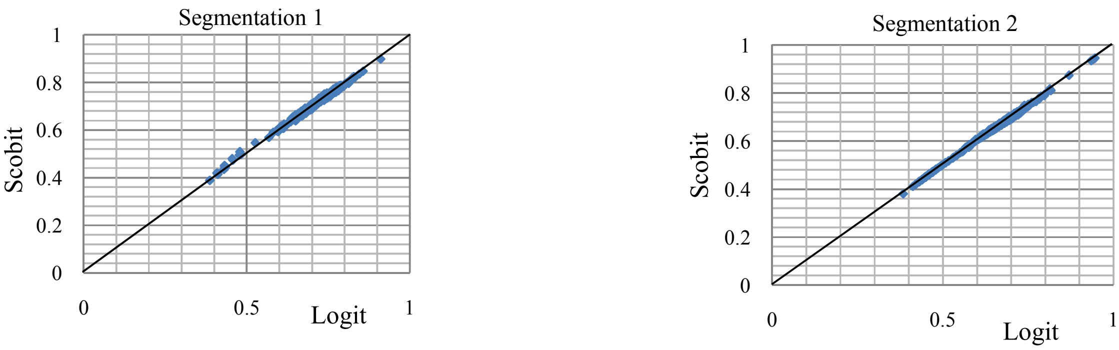

Figure 3. Comparison of Choice Probability of Participation

The skewness parameter across individuals is shown in Figure 2. We can see that the skewness parameters are different across individuals and the average values are 1.14, 1.15 with the standard deviation 0.24, 0.12 for the two segmentations, respectively, which are not statistically different from 1. To further examine the difference of the two models, the calculated choice probabilities of participating in tourism from these two models are illustrated in Figure 3. One can see that the results of two models are almost the same. The Scobit model estimates a little higher choice probabilities of participation when choice probability is over 50% for segmentation 1.

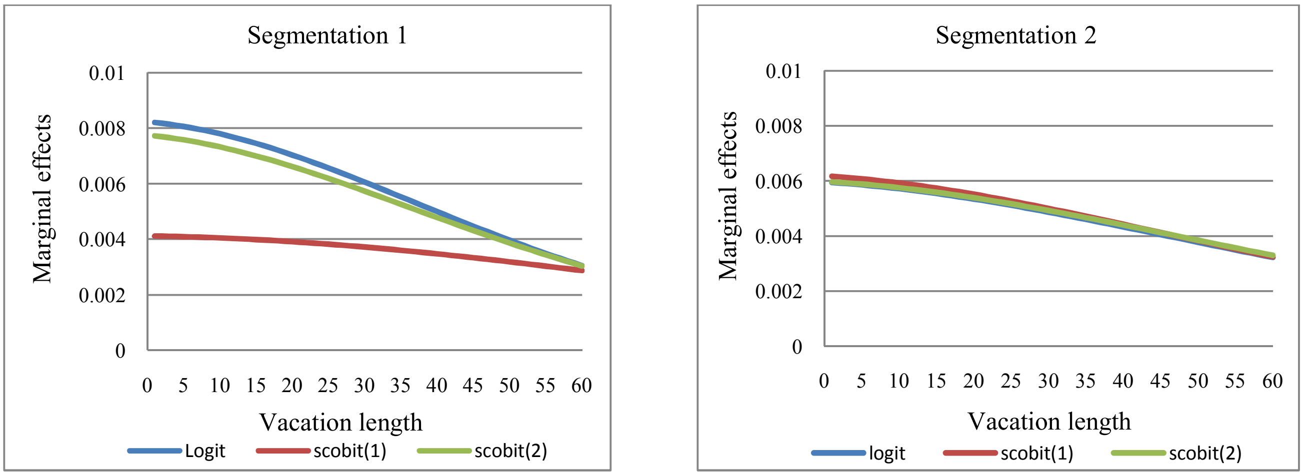

Figure 4. Marginal Effects of Vacation Length on Participation Probability

(4) Marginal Effects

Here, we calculated the marginal effects of vacation length on participation probability based on the logit model, the Scobit model with homogeneous skewness parameter (Scobit 1) and the Scobit model with heterogeneous skewness parameter (Scobit 2). In segmentation 1, the marginal effects from logit model and the Scobit model with heterogeneous skewness

*Keywords: Tourism and leisure behavior, Activity participation, Scobit model, marginal effects
Graduate School for International Development and Cooperation, Hiroshima University, 1-5-1, Kagamiyama, Higashi-Hiroshima City, 739-8529, Japan, Phone & Fax: +81-82-424-6919; **wuling1017@gmail.com; ***zjy@hiroshima-u.ac.jp; ****afujiw@hiroshima-u.ac.jp)

parameter do not show much difference. While the marginal effects from the Scobit model with homogeneous skewness parameter are much smaller than that of logit model. It indicates that the influence of vacation length change will be smaller derived from the Scobit 1. In segmentation 2, the marginal effects are almost the same for the three models.

5. Conclusions

This study analyzes individual's decision on whether to go on vacation or not based on a Scobit model, which includes a skewness parameter to relax assumption that the sensitivity of individuals to changes in explanatory variables is highest for those who have indifferent preferences over participation and non-participation. The empirical application is carried out using the data stemmed from a survey conducted in Japan based on a telephone interview in 2002. Using this data the impacts of several attributes on participation decisions in tourism are investigated. It is revealed that individual's age, employment status and vacation system have significant influence on their choice of participation in tourism. In terms of model specification, the role of skewness parameter is discussed based on the estimated results. Furthermore, the results derived from Scobit model and binary logit model are compared. In the case of the scobit model with homogeneous skewness parameter, the sample is divided into two segmentations. It is found that for segmentation 1 (individuals from large cities), the accuracy of the Scobit model is higher than that of the logit model and the skewness parameter is significant different from 1. In the case of the Scobit model with heterogeneous skewness parameter, the results are almost the same as logit model. One explanation could be that skewness parameter is influenced by some unobserved variables other than those adopted in this study. Thus, it is necessary to conduct further research to examine the difference between Scobit model and logit model.

References

1) Zhang, J. and Timmermans, H.J.P. (2010a) A Scobit-based panel analysis of public transport users' multitasking behavior, Transportation Research Record (in press)

2) Zhang, J. and Timmermans, H. (2010b) A Scobit-based travel mode choice model, Proceedings of the 2010 Academy of Marketing Science Annual Conference, Portland, Oregon, USA, May 26 ~ 29.

3) Jackson, E. L.: Leisure constraints/constrained leisure: special issue introduction, Journal of Leisure Research, Vol.23, pp.279–285, 1991.

4) Crawford, D., Jackson, E. and Godbey, G.: A hierarchical model of leisure constraints, Leisure Sciences, Vol.13, pp.309–320, 1991.

5) Fleischer, A. and Seiler, E.: Determinants of vacation travel among Israeli seniors: theory and evidence, Applied Economics, Vol.34, pp.421-430, 2002.

6) Stemerding, M., Oppewal, H. and Timmermans, H.: A constraints-induced model of park choice, Leisure Sciences, Vol.21, pp.145-158, 1999.

7) Nicolau, J. L. and Mas, F. J.: Stochastic modeling: A three-stage tourist choice process, Annals of Tourism Research, Vol.32, pp.49-69, 2005.

8) Mergoupis, T. and Steuer, M.: Holiday taking and income, Applied Economics, Vol.35, pp.269-284, 2003.

9) Hellstrom, J.: A bivariate count data model for household tourism demand, Journal of Applied Econometrics, Vol.21, pp.213-226, 2006.

10) Melenberg, B. and Soest, A. V.: Parametric and semi-parame'1ric modelling of vacation expenditures, Journal of Applied Econometrics, Vol.11, pp.59-76, 1996.

11) Alegre, J., Mateo, S. and Pou, L.: An analysis of households' appraisal of their budget constraints for potential participation in tourism, Tourism Management, Vol.31, pp.45-56, 2010.

12) Nagler, J.: Scobit: An alternative estimator to logit and probit, American Journal of Political Science, Vol.38, pp.230-255, 1994.