RE-EXAMINING TRAVEL CHOICE BEHAVIOR BASED ON A SCOBIT MODEL

Junyi ZHANG[1)*], Zhuo WANG[1)], Lingling WU[1)], Harry TIMMERMANS[2)], and Akimasa FUJIWARA[1)]

[1)] Transportation Engineering Laboratory, Graduate School for International Development and Cooperation, Hiroshima University, Japan

[2)] Urban Planning Group, Eindhoven University of Technology, The Netherlands

ABSTRACT

Aiming at relaxing an unrealistic assumption made in the traditional logit and probit models, this study explores the applicability of an alternative binary choice model, named Scobit model, in the travel behavior analysis. In case of the binary choice situation, the dominant logit and probit models implicitly impose the assumption that individuals who are invariant between the two choice alternatives (i.e., choice probability is 0.5) are most sensitive to changes in the independent variables than people with a clear preference for one of the choice alternatives. This is because both the logistic and normal density functions are symmetric about zero. However, this assumption has not been tested when applying these models. In reality, the probability level at which independent variables have their maximum impact on a change in choice probability is not necessarily 0.5. The Scobit model could relax this assumption by simply introducing a skewness parameter, where allows the model to include the logit model as a special case. With the Scobit model, it is also expected that marginal effects of explanatory variables could be measured in a more proper way. This study confirmed the effectiveness of the Scobit model using several types of travel choice data, including travel mode choice, pre-trip information acquisition behavior, departure time choice behavior, and tourism participation behavior. Especially, the Scobit model is more powerful in representing the heterogeneity in travel choice behavior than the traditional logit model.

Keywords: Travel choice behavior, Scobit, Logit, Marginal effects, Heterogeneity

[*] Corresponding author: +81-82-4246919; zjy@hiroshima-u.ac.jp

INTRODUCTION

In transportation research, discrete choice modes are typically applied to estimate the effects of policy variables (e.g., road pricing and reduction of travel time) on particular facets of travel behavior (e.g., departure time choice, travel mode choice, and route choice). In this context, the calculation of elasticities and corresponding measures such as value of time are useful policy measures which have found ample application and have been widely reported in both the academic and professional literature. It should be realized however that these policy measures do not only depend on the estimated parameters of the discrete choice model, but more fundamentally also on the functional form of the model itself. In case of a binary choice situation, the dominant logit and probit models implicitly impose the assumption that individuals who are invariant between the two choice alternatives (choice probability of 0.5) are most sensitive to changes in the independent variables than people with a clear preference for one of the choice alternatives. This is because both the logistic and normal density functions are symmetric about zero. However, this assumption has not been tested when applying these models.

In reality, the probability level at which independent variables have their maximum impact on a change in choice probability is not necessarily 0.5. If this is the case, elasticities and other model estimations including value of travel time will be biased. The Scobit model, developed by Nagler (1994), allows testing this assumption. This model is based on the Burr-10 distribution (Burr, 1942), which includes a skewness parameter with positive value. When the skewness parameter is equal to 1, the Scobit model returns to the binary logit model. Since marginal effect on the choice probability for a change in an independent variable is the product of probability density function and the corresponding parameter, one can easily imagine that the maximum value of this density function depends on the skewness parameter. Thus, if individuals with initial probability other than 0.5 are those most sensitive to the change, then the logit or probit model would result in a misspecification and consequently biased inferences about the marginal effects of changes of any explanatory variables will be made.

In a previous study on multi-tasking during traveling (Zhang and Timmermans, 2010a) and travel mode choice behavior (Zhang and Timmermans, 2010b), we found evidence for possible misspecification and the potential value of the Scobit model. In this study, we will investigate whether this result was an exception or whether the stipulated problem may be general by looking at several different types of binary travel behavior data that covers departure time choice behavior, travel information acquisition behavior, and tourism participation behavior, where a part of previous results about travel mode choice models will be used for comparison.

The rest of this paper is organized as follows. First, Scobit model will be explained. Second, to compare the model performance in different contexts, several sets of data used in this study will be explained. Third, model estimation and comparison among different behavioral contexts are given. Finally, this study is concluded along with a discussion about future research issues.

A BINARY TRAVEL MODE CHOICE WITH SCOBIT STRUCTURE

This paper only deals with binary choices. We leave the extension to multinomial choices for future research.

Assume that there are two alternatives in a choice set. Then, the utilities of the two alternatives can be defined as,

Alternative 1: $u_{n1} = v_{n1} + e_{n1}$, (1)

Alternative 2: $u_{n2} = v_{n2} + e_{n2}$, (2)

where, *n* indicates a trip maker, u_{n1}, u_{n2} are utility functions, v_{n1}, v_{n2} are deterministic terms of u_{n1}, u_{n2}, and e_{n1}, e_{n2} are error terms of alternatives 1 and 2, respectively.

Then, the probability p_{n1} that trip maker *n* chooses *alternative 1* can be described as,

$$p_{n1} = \Pr(u_{n1} > u_{n2}) = \Pr(e_{n1} - e_{n2} > v_{n2} - v_{n1}). \quad (3)$$

Let us define a new error term $\varepsilon_n = e_{n1} - e_{n2}$ and further assume that it follows a distribution with $F(\varepsilon_n)$. Then the probabilities p_{n1} and p_{n2} can be derived as,

$$p_{n1} = 1 - F(-(v_{n1} - v_{n2})), \quad (4)$$

$$p_{n2} = F(-(v_{n1} - v_{n2})). \quad (5)$$

The deterministic terms v_{n1}, v_{n2} are usually assumed to be a linear function of explanatory variables for each alternative. Then $v_{n1} - v_{n2}$ can be defined as,

$$v_{n1} - v_{n2} = \sum_k \beta_k (x_{n1k} - x_{n2k}), \quad (6)$$

where, x_{n1k}, x_{n2k} are the *k*th variables for alternatives 1 and 2 with parameter β_k, respectively.

Policy makers or analysts always need to know the marginal effect of a change in $x_{n1} - x_{n2}$. This marginal effect ME_x^p can be derived as,

$$ME_x^p = \frac{\partial p_{n1}}{\partial (x_{n1} - x_{n2})} = f(-\sum_k \beta_k (x_{n1} - x_{n2}))\beta_k, \quad (7)$$

where, $f(\bullet)$ is the probability density function of $F(\varepsilon_n)$.

It is obvious that ME_x^p depends not only on β_k, but also on the value of $x_{n1} - x_{n2}$, and in particular $f(\bullet)$. If a normal or Weibul distribution is assumed, then $f(\bullet)$ will have a maximum at $\sum_k \beta_k (x_{n1} - x_{n2}) = 0$. This means that any given variable $x_{n1} - x_{n2}$ will has its greatest effect on those individuals with $\sum_k \beta_k (x_{n1} - x_{n2})$ being closest to 0, or with p_{n1} being closest to 0.5. However, if individuals with an initial choice probability other than 0.5 are those most sensitive to the change, then the logit or probit model would result in a misspecification and consequently biased inferences about the marginal effect. It is therefore necessary to adopt a more general distribution which allows the highest sensitivity to changes in variables at any initial probability. To meet the above requirement, this study applies the Scobit model (Nagler, 1994), which to the best of our knowledge is not well known in transportation and many other applied sciences. This model can be obtained by assuming the following distribution function $F(\varepsilon_n)$, which is, in fact, a Burr-10 distribution (Burr, 1942). Note that Burr-10 distribution is one of the 12 distributions given by Burr (1942).

$$F(\varepsilon_n) = \frac{1}{(1+\exp(-\varepsilon_n))^{\alpha}} \tag{8}$$

where, α is a parameter used to measure the skewness of Burr-10 distribution.

Having defined the above distribution function $F(\varepsilon_n)$, the probabilities of choosing the two alternatives can be derived as,

$$p_{n1} = 1 - F(-(v_{n1} - v_{n2})) = 1 - \frac{1}{(1+\exp(v_{n1} - v_{n2}))^{\alpha}}, \tag{9}$$

$$p_{n2} = 1 - p_{n1} = \frac{1}{(1+\exp(v_{n1} - v_{n2}))^{\alpha}}, \tag{10}$$

The Burr-10 distribution satisfies the condition that $f(\bullet)$ does not attain a maximum only when $F(\bullet)$=0.5, and it is defined for $-\infty < \varepsilon_t < \infty$. When α is equal to 1, equations (9) and (10) return to the logit model. Thus, the popular logit model is nested within the Scobit model. The Scobit model is also called the *skewed logit* model because it allows a skewed response curve, which is different from the symmetric curve (symmetric about zero) derived from the logit model. The probabilities of alternative 1 for different values of skewness parameter are shown in Figure 1. As seen in Figure 1, the probability curve becomes asymmetric about zero when the skewness parameter is different from 1. Concretely speaking, when the skewness parameter is not equal to 1, the probability becomes less sensitive to gains and more sensitive to losses in the utility of *alternative 1* relative to that of *alternative* 2. This observation is similar to the argument from the prospect theory (Tversky and Khaneman, 1979), which relies on three parameters to explain human decisions: one describes the degree of loss aversion, and the other two parameters explain risk aversion over gains and risk seeking over losses, respectively. The observed sensitivity is further different across the value space of skewness parameter. With the increase in the value of skewness parameter, the probability becomes less sensitive to the change of utility.

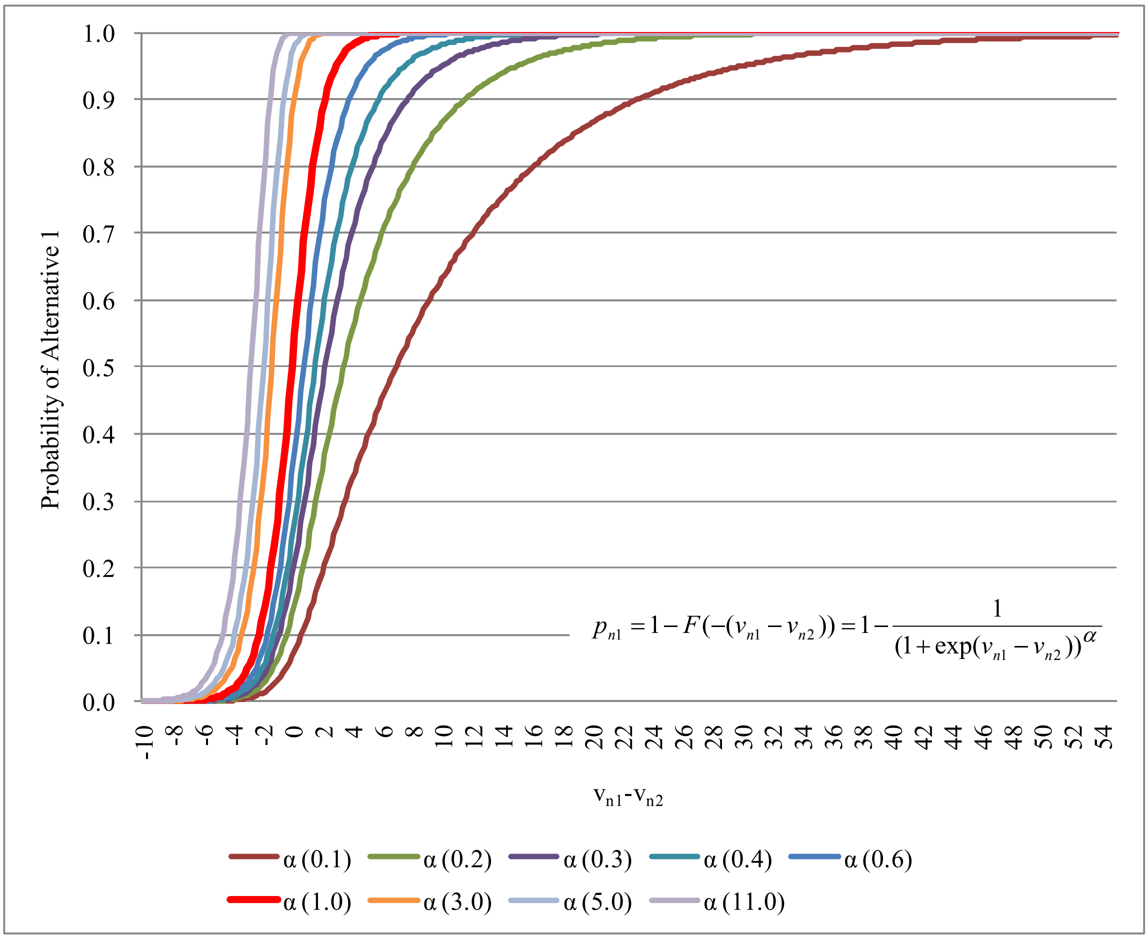

Figure 1. Probabilities of Alternative 1 by Skewness Parameter

With all the above-mentioned equations, the log-likelihood function for the Scobit model is given as follows:

$$LogL = \sum_{n=1}^{N} \ln({p_{n1}}^{\delta_{n1}} {p_{n2}}^{\delta_{n2}}) \tag{11}$$

Here, *N* indicates the total number of samples, and δ_{n1}, δ_{n2} are dummy variables that are equal to 1 when alternatives 1 and 2 are chosen respectively, otherwise 0. The resulting Scobit model can be estimated using standard maximum likelihood estimation method.

DATA

Here, four types of choice behaviors will be modeled and compared to clarify the performance of the Scobit model. These choice behaviors include departure time choice (peak hours vs. off-peak hours), pre-trip information acquisition behavior (refer to information or not), travel mode choice behavior (car or bus), and tourism participation behavior (participation in tourism or not). Needless to say, other behavioral aspects should be examined in a more systematical way. Using the above sets of data is because these were available at the time of writing. Even though existing studies (in political science) have examined the performance of the Scobit model using Monte Carlo experiment (Nagler, 1994), analysis based on actual data could provide more useful insights into the understanding of human behavior.

Departure time choice

The data was selected from a stated preference (SP) survey about departure time and route choice behavior in Beijing in May 2008. It was assumed that drivers' vehicles were equipped with a personal navigation device, which could provide drivers with real-time and dynamic traffic information. In total, four joint choice alternatives are assumed: trunk road in off-peak hours, ring road, trunk road, and branch road in peak hours. In this study, the four alternatives are grouped into two alternatives: peak hours and off-peak hours, for estimating the Scobit model. The assumed attributes and their levels for choice are: travel purpose (business and recreation), error of dynamic travel information prediction (big: 30%, small: 10%), timing constraint for arrival time (whether being late is allowed or not), travel distance (long-, medium- and short-distance), travel time for each road type in peak hours (long and short time), probability of arrival time delay when using each road type in peak hours (low: 20%, high: 60%). Note that probability for arrival time delay in off-peak hours is set at 0% (i.e., early arrival) and travel time for trunk road in off-peak hours is fixed. In addition, in-home activity time before departure was selected as the choice context variable. Drivers were told if they chose peak hours for departure, they would have 2 hours to stay at home; on the contrary, they would only have 30 minutes to stay at home if they chose departure during off-peak hours. Based on the orthogonal experimental design method, 16 SP profiles were obtained. These profiles were further grouped into 4 balanced blocks, which were randomly assigned to each respondent, who was asked to choose one alternative from the four alternatives in each profile. The SP survey was implemented at four major areas in Beijing, and as a result, 624 drivers participated in the survey. Details of the survey refer to Wang et al. (2009). In this study, 1,872 SP responses from the 624 drivers were used by excluding irrelevant samples, and 56% of samples selected to depart during peak hours.

Pre-trip information acquisition behavior

The data was selected from an SP survey about the pre-trip information acquisition behavior in Hiroshima City, Japan in 2002. In the survey, respondents were also asked to answer the questions about the use of information acquisition devices and travel mode choices, which are however not used in this study. The targeted travel information includes length of road traffic congestion shown either in print or diagrammatically, timetables for transit systems (bus and a new transit system), and travel time for all travel modes. Based on the above-mentioned orthogonal experimental design, 25 profiles were constructed after excluding the unrealistic ones (note that attributes related to travel modes and information devices were also combined in the SP design). The 25 profiles were grouped into 5 balanced blocks. Each respondent received only one block of 5 profiles. As a result, 565 local residents returned the SP questionnaires. Details of the survey refer to Zhang and Fujiwara (2004). In this study, 1,952 SP responses were selected, where 70% answered to refer to travel information before departure.

Travel mode choice behavior

The data was selected from a revealed preference (RP) survey about travel mode choices of residents living in Hiroshima City, Japan, where only car and bus were the alternative modes for commuting at the time of survey. This is a four-wave panel survey conducted in 1987, 1990, 1993, and 1994, respectively. As a result, 226 respondents reported their travel mode choice behaviors. It is observed that the shares of bus usage for commuting were 56% in 1987, 58% in 1990, 59% in 1993, and 59% in 1994. In the survey, travel service levels such as travel time and cost were reported as well as household/individual attributes. In fact, along with the RP survey, an SP panel survey was also simultaneously implemented to investigate how people would like to choose a new transit system. Since SP data will not be used in this study, details refer to our previous study (Zhang et al., 2001).

Tourism participation behavior

The data used in this study comes from a survey conducted in Japan based on a telephone interview in 2002. Respondents were randomly selected from telephone directories across the whole country. Different from the above three surveys that were conducted by the authors' laboratory, this survey was conducted with the help of a professional survey company, which tried to collect the samples to reflect the characteristics of the whole population in Japan at the time of survey. The survey collected information about individuals' tourism participation (whether participated in tourism or not, where and how often to visit, and travel expenditure) in a year period, respondents' general tourism preferences and subjective evaluations about the attractiveness of several major destinations, as well as individual/household characteristics. Here we only focus on the participation, i.e., whether participated in tourism or not in a year period. The valid sample size is 1,000 individuals, and 65.7% of respondents participated in tourism. Details of the survey refer to Wu et al. (2009).

MODEL ESTIMATION AND DISCUSSION

Model accuracy and skewness parameter

To compare the performances of the Logit and Scobit models, we first tried to find the best set of explanatory variables for the Logit model with respect to each dataset based on a preliminary study, and then estimated the Scobit model using the same set of variables. In total, we estimated seven models: a departure time choice model, a pre-trip information acquisition model, a tourism participation model, and four travel mode choice models (using a four-wave panel data). Indicators of model accuracy and the estimated skewness parameters are shown in Tables 1 and 2.

Remember that when the skewness parameter is equal to 1, the Scobit model returns to the Logit model. We conducted two types of t-test: one against 0 and the other against 1. It is obvious that all the skewness parameters are statistically different from 0 at the 95% significance level, implying that introducing the skewness parameter is meaningful in a statistical sense. However, looking at the t-score against “1”, in only three out of the seven models, the skewness parameters are different from “1” at the 95% level: the departure time choice model and the travel mode choice models in 1990 and 1994, suggesting that those three Scobit models should be applied to replace the Logit model. Focusing on the CHISQ test results shown in Table 2, it is found that in three out of the seven datasets, the Scobit model is estimated to be superior to the Logit model. This is consistent with the observation in Table 1. These estimation results at least suggest that the Logit model is not always suitable to represent the binary travel choice behavior.

Looking at the values of skewness parameters, except for the value in the pre-trip information acquisition model (which is very close to 1), they are quite different from 1. Using these values to calculate the probabilities, as shown in Figure 1, one can expect a very different shape of the probability curve from that from the Logit model. For the departure time choice model and travel mode choice models which skewness parameters range between 0.6609 and 0.9717, one can observe changes in the probabilities across a wider range of utility space, comparing to the Logit model. In contrast, for the tourism participation model, the skewness parameter is 3.5520, suggesting the relevant utility space becomes much narrower.

To figure out how the Scobit and Logit models are different from each other, we calculated the choice probabilities of all the samples for each dataset. The results are shown in Figure 2. Almost no difference between the Scobit and Logit models is observed with respect to the pre-trip information acquisition model and tourism participation model. We expected a large difference between the Scobit and Logit models for at least for the tourism participation model, because its skewness parameter is 3.5520, which is much larger than 1. However, the observed difference is ignorable. Such indifference might be because of two reasons: one is because both models do not have a significant skewness parameter, and the other is because especially for the tourism participation model, the parameters of explanatory variables in the Scobit and Logit models are quite different. For the travel mode choice models, the models in 1987 and 1993 show indifferent between the Scobit and Logit models, while the models in 1990 and 1993 show some discrepancies. This might be because the estimated skewness parameters are statistically different from “1”.

Comparing parameters of explanatory variables

Since the Scobit model introduces a skewness parameter, it is difficult to directly compare the parameters of explanatory variables. But we can compare them in an indirect way. In Table 3, which shows the estimated parameters of explanatory variables and t-score, there is a column to show the relative influence of each variable. The relative influence of variable *x* is defined as the ratio between the parameter of *x* and the parameter of an explanatory variable, which is arbitrarily selected as a reference. For example, in the departure time choice model, "activity time at home divided by travel time" is taken as a reference and the variable is named as the reference variable. Other explanatory variables are named as comparison variables.

Table 1. Model Estimation Results (1): Skewness Parameters

Behavioral aspects	Sample Size	Year of Survey	Number of Explanatory Variables	Skewness Parameter		
				Value	t-score (0)	t-score (1)
Peak-Offpeak choice	1872	2008	6	0.6609	5.545	***-2.845***
Information acquisition	1952	2002	6	1.0668	11.263	0.705
Tourism generation	1000	2002	11	3.5520	0.173	0.125
Travel mode choice (1987)	226	1987	2	0.8665	8.233	-1.268
Travel mode choice (1990)	226	1990	2	0.8121	8.978	***-2.078***
Travel mode choice (1993)	226	1993	2	0.9736	7.971	-0.216
Travel mode choice (1994)	226	1994	2	0.7822	7.044	***-1.961***

Table 2. Model Estimation Results (2): Model Accuracy

Behavioral aspects	Initial Log-Likelihood	Final Log-Likelihood		Adjusted McFadden's Rho-square		Chisq Test
		Logit	Scobit	Logit	Scobit	
Peak-Offpeak choice	-1297.572	-1245.75	-1242.66	0.0369	0.0387	***-6.18***
Information acquisition	-1353.023	-1145.12	-1144.86	0.1510	0.1508	-0.52
Tourism generation	-693.147	-596.69	-596.69	0.1296	0.1287	0.00
Travel mode choice (1987)	-156.651	-101.701	-100.932	0.3450	0.3470	-1.54
Travel mode choice (1990)	-156.651	-122.492	-120.654	0.2111	0.2194	***-3.68***
Travel mode choice (1993)	-156.651	-82.5264	-82.5033	0.4685	0.4662	-0.05
Travel mode choice (1994)	-156.651	-105.093	-103.057	0.3231	0.3333	***-4.07***

The relative influence of each explanatory variable changes between the Scobit and Logit models. The changes are further different across different behavioral contexts. For the departure time choice behavior and pre-trip information acquisition behavior, the relative influences of comparison variables estimated by the Scobit model become larger; while those of tourism participation behavior and travel mode choice behavior show an opposite trend.

Focusing on temporal changes of the relative influences, this study estimated the travel model choice model for four time points using a panel data. Even though the relative influences of comparison variables become smaller in the Scobit model, the changes are much smaller, in comparison to other behavioral contexts.

Table 3. Model Estimation Results (3): Parameters of Explanatory Variables

Explanatory variable	Logit Model				Scobit Model			
	Parameter	Relative value	t-statistic		Parameter	Relative value	t-statistic	
(1) Departure time choice model								
Activity time at home divided by travel time	0.1979	*1.0000*	4.018	**	0.1573	*1.0000*	2.685	**
Trip purpose (1: work; 0: Leisure)	-0.5043	*-2.5483*	-5.067	**	-0.5865	*-3.7293*	-4.717	**
Permission of being late (1: Yes; 0: No)	0.6733	*3.4024*	7.178	**	0.7053	*4.4846*	6.470	**
Ownership of car navigation system (1: Yes; 0: No)	-0.2851	*-1.4408*	-2.122	*	-0.3474	*-2.2086*	-2.166	*
Age	-0.0776	*-0.3919*	-1.792		-0.2073	*-1.3181*	-2.605	**
Gender (1: Male; 0: Female)	-0.0010	*-0.0048*	-0.008		-0.0661	*-0.4200*	-0.491	
(2) Information acquisition model								
Car information								
Congestion shown in print (1: Yes; 0: No)	0.2539	*1.0000*	2.257	*	0.2032	*1.0000*	1.576	
Congestion shown diagrammatically (1: Yes; 0: No)	0.4197	*1.6531*	3.951	**	0.3689	*1.8154*	2.981	**
Astramline information								
Timetable (1: Yes; 0: No)	0.8844	*3.4828*	7.646	**	0.8189	*4.0298*	5.748	**
Travel time (1: Yes; 0: No)	0.5967	*2.3499*	5.594	**	0.5441	*2.6774*	4.329	**
Bus information								
Timetable (1: Yes; 0: No)	0.0962	*0.3790*	0.923		0.0522	*0.2566*	0.443	
Travel time (1: Yes; 0: No)	0.4877	*1.9205*	4.237	**	0.4353	*2.1422*	3.303	**
(3) Tourism generation model								
Income (million yen)	0.0080	*1.0000*	0.045		0.1080	*1.0000*	0.295	
Employment (1: Employed; 0: Others)	0.6570	*82.1250*	2.737	**	1.4020	*12.9815*	0.522	
Vacation: The longest vacation that one can get in a year	0.0240	*3.0000*	2.725	**	0.0490	*0.4537*	0.507	
Age	-0.0480	*-6.0000*	-0.970		-0.1720	*-1.5926*	-0.579	
Household size: Number of household members	0.0240	*3.0000*	0.461		0.0910	*0.8426*	0.509	
(4)-1 Travel model choice model: 1987								
Travel time difference (car-bus) (minute)	0.0260	*1.0000*	1.384		0.0301	*1.0000*	1.434	
Travel cost difference (car-bus) (yen)	-0.0064	*-0.2448*	-7.704	**	-0.0067	*-0.2231*	-6.632	**
(4)-2 Travel model choice model: 1990								
Travel time difference (car-bus) (minute)	-0.0169	*1.0000*	-1.414		-0.0208	*1.0000*	-1.629	
Travel cost difference (car-bus) (yen)	-0.0036	*0.2149*	-6.648	**	-0.0038	*0.1814*	-6.316	**
(4)-3 Travel model choice model: 1993								
Travel time difference (car-bus) (minute)	-0.0975	*1.0000*	-4.805	**	-0.0985	*1.0000*	-4.470	**
Travel cost difference (car-bus) (yen)	-0.0040	*0.0407*	-8.074	**	-0.0040	*0.0404*	-7.771	**
(4)-4 Travel model choice model: 1994								
Travel time difference (car-bus) (minute)	-0.0401	*1.0000*	-3.274	**	-0.0510	*1.0000*	-2.881	**
Travel cost difference (car-bus) (yen)	-0.0035	*0.0878*	-7.196	**	-0.0037	*0.0726*	-5.946	**

Influence of market segmentation

Here, taking the tourism participation behavior as an example, which skewness parameter is estimated to be insignificantly different from “1”, we re-estimated the tourism model for two market segments: one for those living large cities and the other for those living small cities/towns. This segmentation was arbitrarily selected. The estimation results are shown in Tables 4 and 5, respectively. It is observed that the skewness parameter for large cities is statistically different from “1” at the 95% significance level. In contrast, the parameter for small cities/towns is indifferent from “1”. These results suggests that even though the tourism participation model using the whole sample suggests the indifference between the Scobit and Logit models, such indifference may not be true with respect to all market segments. Considering the increasing importance of dealing with different population groups, development of choice models should not ignore the influences of different population groups.

Focusing on the relative influence as defined previously, one can see that for most of the explanatory variables, the model for large cities derives clearly smaller relative influences of comparison variables in the Scobit model, while the relative influences of comparison variables in the model for small cities/towns do not show clear differences.

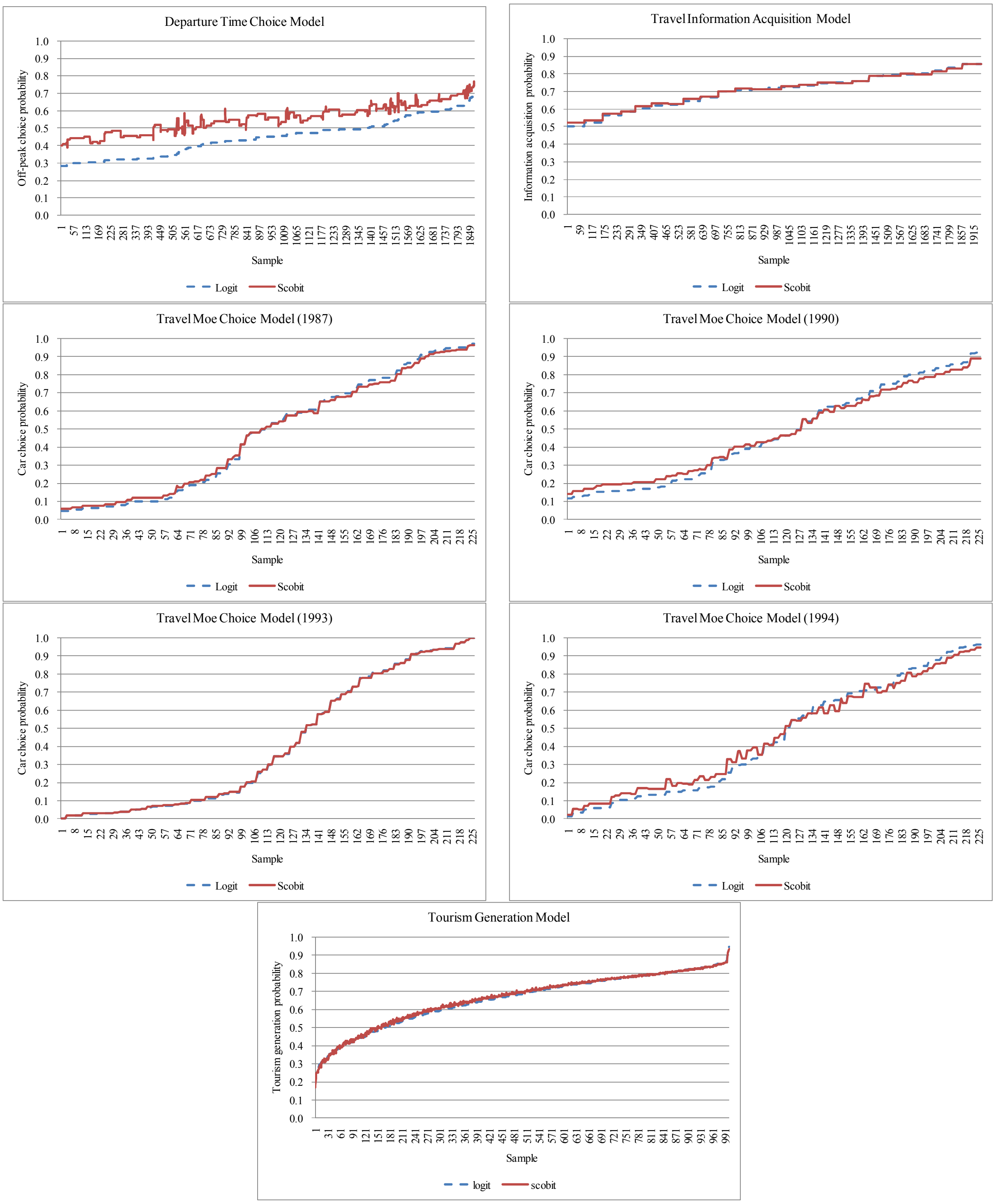

Figure 2. Comparisons of Choice Probabilities between Scobit and Logit Models

Table 4. Tourism Generation Model for Large Cities

Explanatory variable	Logit model				Scobit model			
	Parameter	Relative Influence	t-score		Parameter	Relative Influence	t-score	
Income	-0.267	1.000	-0.769		-0.293	1.000	-1.810	+
Employment	1.061	-3.974	2.829	**	0.859	-2.932	1.643	+
Vacation	0.036	-0.135	2.783	**	0.027	-0.092	1.350	
Skewness parameter					0.375		1.612 (0)	+
							2.677 (1)	**
Age	-0.102	0.382	-2.167	*	-0.295	1.007	-1.810	+
Household size	0.094	-0.352	1.076		0.034	-0.116	0.284	
Sample size	290							
Initial log-likelihood	-201.01							
Converged log-likelihood	-174.530				-172.770			
Adjusted McFadden's Rho-squared	0.107				0.111			

Table 5. Tourism Generation Model for Small Cities and Towns

Explanatory variable	Logit model				Scobit model			
	Parameter	Relative Influence	t-score		Parameter	Relative Influence	t-score	
Income	0.174	1.000	0.837		0.173	1.000	0.838	
Employment	0.745	4.282	3.769	**	0.763	4.410	3.350	**
Vacation	0.026	0.149	3.376	**	0.026	0.150	2.980	**
Skewness parameter					1.050		3.366 (0)	**
							0.160 (1)	
Age	-0.071	-0.408	-2.278	*	-0.064	-0.370	-1.226	
Household size	0.041	0.236	0.761		0.045	0.260	0.781	
Sample size	710							
Initial log-likelihood	-492.130							
Converged log-likelihood	-448.550				-448.540			
Adjusted McFadden's Rho-squared	0.078				0.076			

Heterogeneity of skewness parameter

As shown in Table 1, three out of the seven Scobit models estimated insignificant skewness parameters. As estimated at the previous sub-section, even though the tourism participation model estimated insignificant skewness parameter using the whole sample, when segmenting the sample into two groups, the skewness parameter becomes significant for one of the two groups. This suggests that the skewness parameter might be heterogeneous across the whole population. In other words, some individuals may show the highest sensitivity to change at p_{n1} = 0.5, some at p_{n1} <0.5, and some at p_{n1} >0.5. However, it is difficult for an analyst to figure this out in advance. To accommodate such heterogeneity, this study therefore defines α as a function of some individual attributes (z_{nq}), where θ_q is the parameter of the qth variable z_{nq}. Note that the exponential function is adopted to meet the requirement that $\alpha_n > 0$.

$$\alpha_n = \exp(\sum_q \theta_q z_{nq}) , \qquad (12)$$

Here, taking travel mode choice behavior as an example, since the previous analyses estimated insignificant skewness parameters for the data in 1987 and 1993, we re-estimated the models for these two time points by introducing equation (12) into the Scobit model (equations (9) and (10)). The resulting model is called the heterogeneous Scobit model.

Following equation (12), define the heterogeneous skewness parameter as an exponential function of socio-demographic attributes including sex, age, employment, and number of household members. For the purpose of comparison, the same set of socio-demographic attributes is also introduced into the Logit model as explanatory variables together with travel time and cost variables. Estimation results are shown in Table 6. Looking at the model accuracy, it is demonstrated that introducing socio-demographic attributes of trip makers remarkably improved the model accuracy in the sense that the Adjusted McFadden's Rho-square values in Table 6 are about 20%~40% higher than the corresponding values in Table 2. It is found that most of the socio-demographic attributes in the two models have statistically significant parameters at the 95% significance level.

Table 6. Heterogeneous Scobit Model and Logit Model

Explanatory variable	Logit Model		Scobit Model	
	Parameter	t-statistic	Parameter	t-statistic
(1) Model for the year of 1987				
Travel time difference (car-bus) (minute)	0.0428	1.858	0.0525	2.110
Travel cost difference (car-bus) (yen)	-0.0064	-6.790	-0.0070	-5.563
Value of travel time (yen/hour)	-403		-452	
			(Skewness parameter)	
Sex (1: Male; 0: Female)	1.2821	2.224	0.9606	2.330
Age	-0.1025	-3.939	-0.0758	-3.965
Employment (1: employed; 0: unemployed)	4.2383	4.048	3.3121	4.351
Number of household members	-0.3965	-1.661	-0.3824	-2.421
Converged log-likelihood	-84.05		-80.35	
Adjusted McFadden's Rho-squared	0.4488		0.4731	
(3) Model for the year of 1993				
Travel time difference (car-bus) (minute)	-0.1265	-3.378	-0.1189	-3.751
Travel cost difference (car-bus) (yen)	-0.0048	-6.031	-0.0048	-7.120
Value of travel time (yen/hour)	1592		1499	
			(Skewness parameter)	
Sex (1: Male; 0: Female)	2.9801	3.377	1.8141	4.179
Age	-0.1255	-3.932	-0.0723	-4.493
Employment (1: employed; 0: unemployed)	2.1011	1.437	1.1865	1.120
Number of household members	0.8661	2.508	0.5146	2.397
Converged log-likelihood	-65.69		-64.23	
Adjusted McFadden's Rho-squared	0.5692		0.5788	

The skewness parameters across trip makers are shown in Figure 3. The average values of the skewness parameters are 1.2 and 1.5 with the standard deviations 1.1 and 1.3 for the two time points, respectively. These values are substantially higher than those estimated in the Scobit model with homogeneous skewness parameter. Especially, 42%~52% of samples have the skewness parameters larger than 1, and 20~30% of samples even have the skewness parameter larger than 2.

The accuracy of the Scobit model is about 2~5% higher than that of the Logit model. To further understand the difference of the two models, the calculated choice probabilities for the car and the bus from these two models are illustrated in Figure 4. In 1987, larger differences between the two models are observed in the sides of smaller (about 0.1~0.2) and larger (about 0.7~0.9) choice probabilities for both car and bus. Relatively large differences between the two models are observed across the whole choice probability space in 1993.

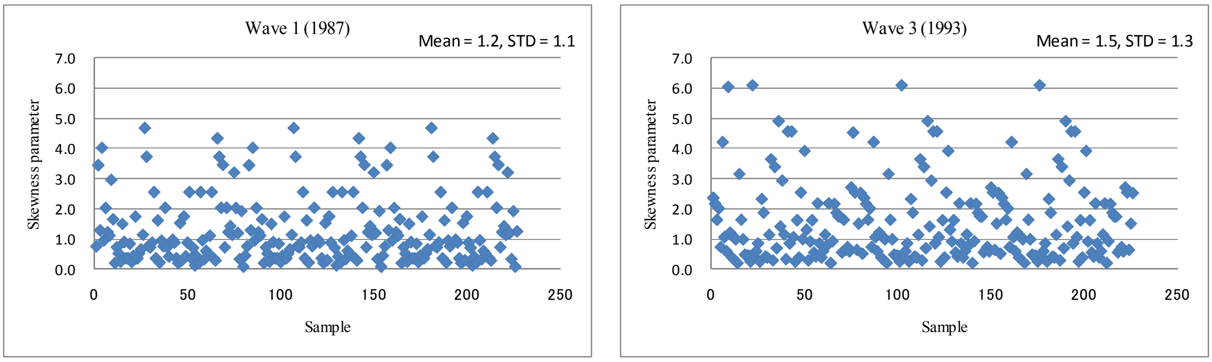

Figure 3. Distribution of Skewness Parameter at the Four Waves

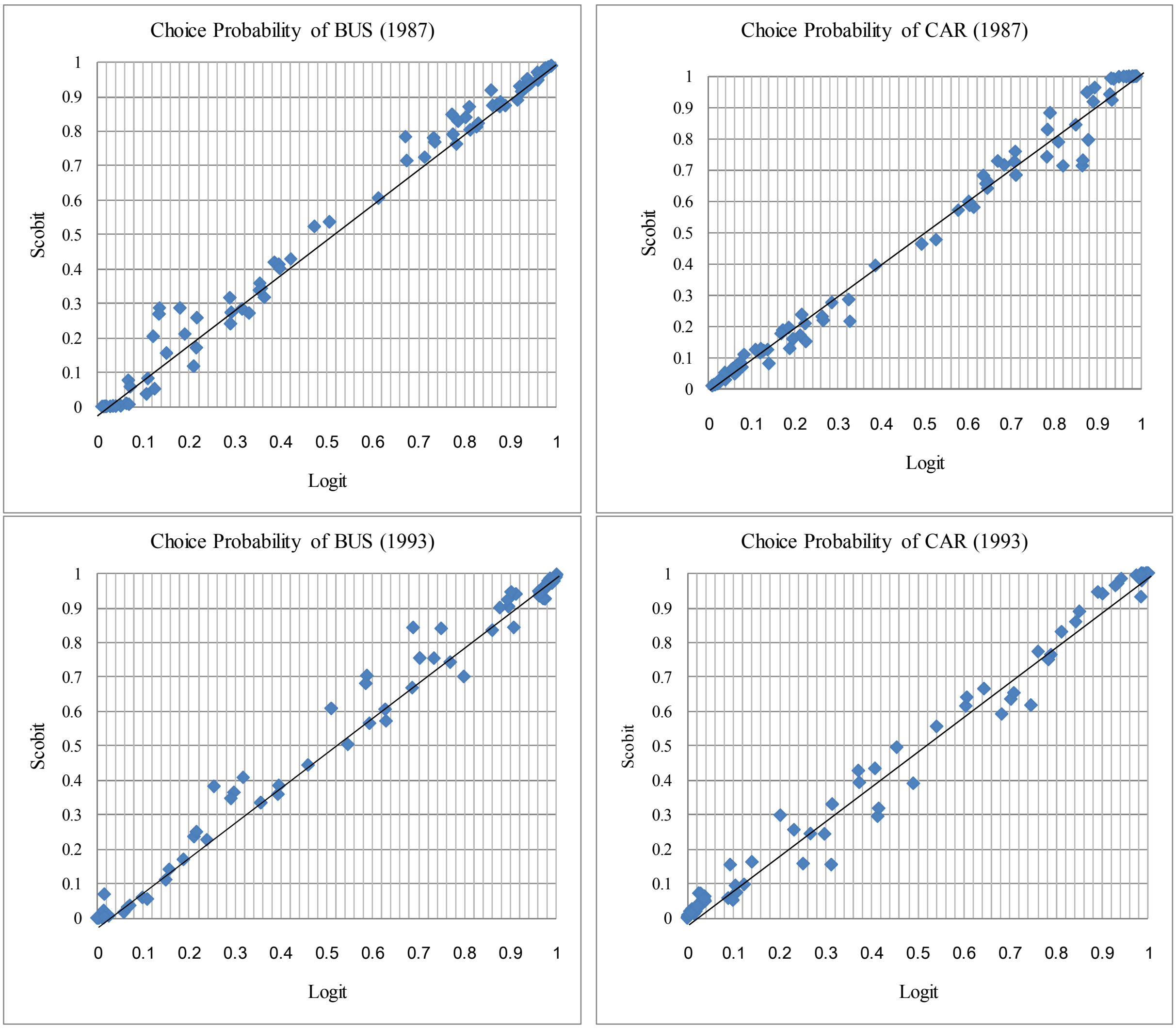

Figure 4. Choice Probabilities from Travel Model Choice Models in 1987 and 1993

CONCLUSIONS

Policy makers in transportation always need to know whether transportation policies could effectively change people's travel behavior or not. For this purpose, the binary logit (or probit) model has been widely applied. Since the choice probability calculated from the logit (or probit) model is symmetric about zero utility, marginal effects of a policy variable show maximal value for those trip makers whose preferences are invariant for two choice alternatives in a choice set, i.e., those trip makers are most sensitive to change in the policy variable. To date, this property of the logit (or probit) model had not been tested in transportation. In this paper, we therefore tested this property by comparing the performance of a binary logit model against a Scobit model, which adds a skewness parameter with positive value. The Scobit model includes the logit model as a special case because when the skewness parameter is equal to 1, the Scobit model returns to the logit model. In this sense, the Scobit model has a more general model structure than the logit model in representing travel choice behavior. The Scobit model can be easily estimated using standard maximum likelihood estimation method. Conceptually, the Scobit clearly outperforms the logit model. This is the first study in transportation to give an extensive analysis about the applicability of the Scobit model using several datasets, which covers some major behavior contexts.

Considering the diversity of travel choice behavior, we selected travel mode choice, departure time choice, pre-trip information acquisition, and tourism participation as examples to empirically clarify the applicability of the Scobit model in the travel behavior analysis. In this sense, we covered daily and non-daily choice behavior in transportation. Except for the travel mode choice model, which was estimated using 226 samples, the other models were estimated using 1,000~2,000 samples. As a result, the skewness parameter was estimated to be significant in the context of departure time choice and travel mode choice models, suggesting that the logit model is not always suitable to represent the binary choice behavior. Statistical tests about the difference between the Scobit and logit models also support this conclusion. It was further found that the Scobit model shows different performance in representing daily behavior than non-daily behavior, for example, the estimated skewness parameter in tourism participation model is 3.5520 (even though not significantly different from 1), which is several times higher than those in the other models. The larger skewness parameter implies that changes in utility across a wider range could only derive the changes in choice probabilities in a much narrower range. In other words, different from daily travel behavior, non-daily behavior could become more captive even when the difference of utilities between two alternatives exceeds a much smaller threshold. On the other hand, it was revealed that clear differences between choice probabilities from the Scobit and Logit models were only observed with respect to those models with statistically significant skewness parameters.

Even though tourism participation model estimated an insignificant skewness parameter against "1" using the whole samples, when re-estimating the model by segmenting the samples into residents living in large cities and small cities/towns, it was found that the skewness parameter becomes statistically different from "1" for large cities. When defining the skewness parameter as a function of some individual attributes (i.e., assuming that the skewness parameter is heterogeneous across samples), we further found that even for those Scobit models estimated to be indifferent from the Logit model, most parameters of the introduced individual attributes are statistically significant and more than 40% of samples have the skewness parameter larger than 1. Choice probabilities calculated from the Scobit and Logit models are also clearly different. These results suggest the existence of heterogeneity in the skewness parameter.

The above findings suggest that the Scobit model could provide a new tool to look at discrete travel choice behaviors from a different angle from traditional models. Extending the Scobit model to represent multinomial choice behavior will be our next research target.

REFERENCES

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

Burr, I.W. (1942), Cumulative frequency functions. *Annals of Mathematical Statistics*, 13, 215-232.

Wang, Z., Zhang, J., and Fujiwara, A. (2009) Evaluating the effects of dynamic travel information on driver's stated departure time and route choice behaviors in Beijing, China, *Proceedings of International Symposium on Urban Planning 2009*, Tainan, Taiwan, August 26-28, 219-228.

Wu, L., Zhang, J., and Fujiwara, A. (2009) A model of heterogeneously nested structure of tourists' destination and travel party choices, *Proceedings of the 15th Annual Conference, Asia Pacific Tourism Association (APTA)*, Incheon, Korea, July 9-12 (CD-ROM).

Kahneman, D., and Tversky, A. (1979) Prospect Theory: An Analysis of Decisions Under Risk, Econometrica, 47 (2), 263-291.

Zhang, J. and Fujiwara, A. (2004) Evaluating the effects of multi-modal travel information based on a discrete choice model with unequal and asymmetric structure, *Proceedings of the 11th World Congress on Intelligent Transport Systems*, Nagoya, Japan, October 18-24 (CD-ROM).

Zhang, J., Sugie, Y. and Fujiwara, A. (2001) A mode choice model separating taste variation and stated preference reporting bias, *Journal of the Eastern Asia Society for Transportation Studies*, 4, No.3, 23-38.

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

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.