Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 2031 - 2045, 2005

MODELING MULTIPLE SOURCES OF HETEROGENEITY IN MODE CHOICE MODEL

Chih-Wen YANG
Assistant Professor
Department of Logistics Engineering and Management
National Taichung Institute of Technology
No 129, Sanmin Rd, Sec. 3, Taichung,
404 Taiwan.
Fax: +886-4-22196161
E-mail: cwyang@ntit.edu.tw

Abstract: This paper is to compare the difference of heterogeneous effect caused by multiple sources, including unobserved heterogeneity in alternatives, taste variation, and heterogeneous choice sets. We construct several heterogeneous discrete choice models to conduct with various heterogeneities. A stated preference data of intercity travel choice is used as empirical case. The empirical results show that the mixed logit model conducting with taste variation has better explanatory power in single heterogeneity and heterogeneous competing destinations model integrating multiple heterogeneities has the best explanatory power. If we ignore heterogeneity in constructing discrete choice models, the estimation of value of time will be biased. Sensitivity analysis presents that expensive alternatives, like air mode, should adopt cut-price strategy to increase choice share and inexpensive alternatives, like bus and train mode, should improve the level of service in travel time to increase choice share.

Key Words: heterogeneity, taste variation, mixed logit model, latent class model, competing destinations model

1. Introduction

Most mode choice models are based on the frame of multinomial logit (MNL) model to analyze traveler behavior. The main reason is its inherent property of Independence from Irrelevant Alternatives (IIA) and this property is based on the convenient but simplistic assumption of Independent and Identically Distributed (IID) error term. The property of IIA implies the assumption of homogeneity in unobservable components of utility. It will product inconsistent parameter estimation if researcher omits the existence of heterogeneity (Hsiao, 1986). However, the standard MNL model could not incorporate heterogeneity properly. In recent year, many flexible discrete choice models have been developed to conduct different sources of heterogeneity, but there are few literatures to compare the difference of heterogeneous effect across models. In this paper, we review related theory of heterogeneity and construct several heterogeneous discrete choice models to discuss multiple sources of heterogeneity in traveler behavior.

Heterogeneity caused by alternative similarity and individual preference can branch into three major topics: (Baltas and Doyle, 2001) (a) unobserved heterogeneity in alternatives, (b) taste variation, and (c) heterogeneous choice sets. The first branch describes alternative similarity

and can be dealt with different discrete choices models, e.g. generalized extreme value model, multinomial probit model, the heteroscedastic extreme value model, or logit kernel model. The second topic concerns about the variation in tastes across individuals. It can be deal with individual characteristics, segmentation, fixed effects, and random effects model to conduct taste variation. Mixed logit (ML) (McFadden and Train, 2000) and latent class model (LCM) are the popular models. The main difference is on the specification of heterogeneity. The third topic can be dealt with the variation of available alternatives in individual's choice sets. There are two kinds of method to determine individual's choice sets: determinism and probability. However, determinism is influenced likely by researcher and lead to bias estimation. Two-stage discrete choice models (Manrai and Andrews, 1998) use probabilistic formulation to conduct heterogeneity in individual's choice sets.

There are many studies discussing single-resource heterogeneity in transportation and marketing literatures. However, only a few studies describe multiple-resource heterogeneity and compare model differences across heterogeneous models. Gonul and Srinivasan (1993) specified each brand-specific intercept and parameter of explanatory variable vary randomly across households. In addition, they also considered customer's loyalty. But they do not incorporate the heterogeneous effect of individual's choice sets. Munizaga et al., (2000) applied many flexible models, including multinomial logit, hierarchical logit, single element nested logt, heteroskedastic extreme value logit, and multinomial probit models, to analysis heterogeneity between options and observations. This study also involved with data difference on reveal preference and stated preference. However, panel data and inconsistent assumption of pure error term (ex., extreme value distribution, or normal distribution) make the comparison of heterogeneity effect inadequately. Greene and Hensher (2003) compared two heterogeneous models: latent class model and mixed logit model. This study only focused on the variation in tastes. Hence, we attempts to compare model differences between three major heterogeneities with identical assumption of pure error term on cross sectional data.

The purpose of this paper is to compare the difference of heterogeneous effect caused by multiple sources. Hence we construct several flexible models which can involve with all three major types of heterogeneity. Firstly, we estimate MNL as based model and use the same specification on the following models. Then Heterogeneous Logit Kernel (HLK) (Walker, 2001) model is built to conduct the unobserved heterogeneity in alternatives. ML model and LCM are used to deal with the random and systematic heterogeneity in taste variation. Furthermore, we adopt Competing Destinations (CD) (Fotheringham, 1988) model to deal with the variation across individual's choice sets, because this model is less time consuming when alternatives is large. CD model use two-stage choice process to incorporate the variation of choice sets across individuals. Finally, we propose a new synthetic model which is called Heterogeneous Competing Destinations (HCD) (Yang et al., 2003) to incorporate multiple-resource heterogeneity.

The reminder of the paper is organized as follows. In section 2, we illustrate foundation theory and model formulation of the above-mentioned models. Section 3 describes the background of empirical data, including alternatives introduction, customized design of stated preference, attributes and level of service, and survey process. Section 4 presents the calibration process and estimation results of above models and compares explanatory power, parameter significance, and value of time across models. We also design two marketing scenarios to test sensitivity of these models. Finally, we propose our conclusions and suggestions.

2. Model formulation

In this section, we first introduce the mixed logit model. Then we discuss the heterogeneous logit kernel model and the latent class model. Finally, we illustrate the competing destination model and propose a new heterogeneous competing destination model. A brief introduction of model formulation is as follows.

2.1 The mixed logit model

The mixed logit (ML) model (Revelt and Train, 1998; McFadden and Train, 2000; Train, 2003) had used to conduct that the coefficients of explanatory variable vary randomly with a specific probabilistic distribution across individuals initially. In recent years, further progress of calibration in simulation method make ML model apply to many research topics, ex., random coefficients, error components, and taste variation. Following different research topics, ML model is called as random-coefficients logit (Ben-Akiva and Lerman, 1985; Bhat, 1998; Train, 1998), error-components logit (Brownstone and Train, 1999), or probit with logit kernel (Ben-Akiva and Bolduc, 1996; Walker, 2001) model. Model formulation is as follows.

Assuming the utility of alternative i ($i = 1,..., J_t$) for individual t ($t = 1,..., T$) can be expressed as

$$U_{it} = \beta_t' x_{it} + \varepsilon_{it} \tag{1}$$

where X_{it} is the vector of variables associated with alternative i and individual t, β' is the vector of the effects of these variables for individual t , and ε_{it} is the error term.

The conditional probability of individual t choosing alternative i, giving β_t had known, will be (Ben-Akiva and Lerman, 1985)

$$P_t(i \mid \beta_t) = \frac{\exp(\beta_t' x_{it})}{\sum_{j \in C_t} \exp(\beta_t' x_{jt})} \tag{2}$$

where C_t is the set of alternatives available to individual t and ε_{it} 's are independent and identical extreme value distribution of type I.

Assuming β_t is a multivariate Normal density function with $N(0, \mu_t)$, the equation (1) can be rewritten as

$$U_{it} = \beta' x_{it} + \mu_t' x_{it} + \varepsilon_{it} \tag{3}$$

Hence, the probability of individual t choosing alternative i can be expressed as

$$P_t(i) = \int_{\mu_t} \frac{\exp((\beta + \mu_t)' x_{it})}{\sum_{j \in C_t} \exp((\beta + \mu_t)' x_{jt})} f(\mu_t) d\mu_t \tag{4}$$

2.2 The heterogeneous logit kernel model

The logit kernel model can conduct with the similarity and heterogeneity across alternatives. Moreover, it also avoids that parameters associated with covariance matrix increase rapidly when alternatives is large. In the logit kernel model, the ε_{it} error term (as equation (1)) can be decomposed into two components: a probit-like component with a multivariate distribution, and an i.i.d. extreme value distribution of type I (Walker, 2001). Using the factor analytic structure, the ε_{it} can be rewritten as

$$\varepsilon_t = F_t \xi_t + v_t = F_t T \zeta_t + v_t \tag{5}$$

where ξ_t is an $(M \times 1)$ vector of M multivariate distributed latent factors, F_t is a $(J_t \times M)$ matrix of the factor loadings that map the factors to error vector, and v_t is a $J_t \times 1$ vector of i.i.d. extreme value distribution of type I. ζ_t is a $M \times 1$ vector of standard normal distribution, TT' is the covariance matrix of ζ_t, and T is the lower triangular matrix of Cholesky factorization of TT'.

By the variant specification of F_t and T, the logit kernel model can conduct the similarity and/or heterogeneity across alternatives. In our study, we focus on the heterogeneity across alternatives. Hence, the F_t is specified as unit matrix and T is expressed as diagonal matrix

$$T = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_J \end{bmatrix} \tag{6}$$

where σ_i^2 is the variance of alternative i. U_{it} in equation (1) becomes

$$U_{it} = \beta' x_{it} + T\zeta_t + v_t \tag{7}$$

Hence, the formulation of heterogeneous logit kernel model can be expressed as follows.

$$P_t(i) = \int_{\zeta_t} \frac{\exp(\beta' x_{it} + T\zeta_t)}{\sum_{j \in C_t} \exp(\beta'' x_{jt} + T\zeta_t)} f(\zeta_t) d\zeta_t \tag{8}$$

Walker (2001) had illustrated a detailed description about the identification and normalization of HLK model. We extract three important conclusions as follow (Walker, 2001).

1. If $J \geq 3$, $J-1$ of heterogeneous variances can be identified only;
2. The best method of normalization is fixing the minimum variance of alternative to zero;
3. In practice, there are two ways to standardize HLK model:

a. First method is to estimate J versions of identified HLK model, each with a different heterogeneous term normalized, the model with the best fit is the one with the correct normalization.

b. Second method is to estimate an unidentified HLK model with all J heterogeneous terms. The heterogeneous term with the minimum variance should be specified to zero.

In the empirical process, we will use above methods to identify and normalize HLK model.

2.3 The latent class model

The latent class model adapts discrete finite segment to deal with the heterogeneity across individuals instead of continuous probabilistic distribution. LCM has convenience on parameter calibration and can also analysis the market structure across alternatives (Kamakura and Russell, 1989; Bhat, 1997). LCM assumes that each individual at the same segment has similar choice behavior. Hence, each segment has its own parameters to represent associated choice preference.

Assuming the conditional probability of individual t choosing alternative i in segment s ($s = 1, \cdots, S$) can be expressed as the logit form.

$$P_t(i) \mid s = \frac{\exp(\beta_s^{'} x_{it})}{\sum_{j \in C_t} \exp(\beta_s^{'} x_{jt})} \tag{9}$$

P_{ts} is the marginal probability of individual t belonging to segment s and it is composed of alternative attributes and individual's socioeconomic characters. The mathematical form is as follows.

$$P_{ts} = \frac{\exp(\gamma_s' z_{it})}{\sum_l \exp(\gamma_l' z_{it})} \tag{10}$$

where z_{it} is the vector of alternative attributes and individual's socioeconomic characters and γ_s is the vector associated with these variables in segment s.

The parameter vector of Sth segment (γ_S) has to normalize to zero ensuring the model is identified. In the special case, z_{it} can be specified as the constant term (e.g., 1) and the latent class probabilities would be simply to a constant which should sum to one.

The probability of individual t choosing alternative i will be

$$P_t(i) = \sum_{s=1}^{S} P_{ts} \times \left[P_t(i) \middle| s \right] \tag{11}$$

BIC (Bayesian Information Criterion, or called Schwarz Criterion) has used to determine the appropriate segments (Chintagunta, 1994; Bhat, 1997; Chintagunta, 1999). The mathematic

form is as

$$BIC = -LL(\beta) + 0.5 \cdot k \cdot \ln(N) \tag{12}$$

where k is the number of parameter and N is the sample size. The most appropriate segment has the minimum BIC value.

2.4 CD model

Two-stage discrete choice models are often used to conduct the heterogeneity in individual's choice sets. We adopt the competing destinations (CD) model (Fotheringham, 1988), because the calibration on CD model would not be time consuming when alternative is large. The probability of individual t choosing alternative i is as

$$P_t(i) = \frac{\exp(\beta' x_{it}) \cdot p(i \in C_t)}{\sum_{j=1}^{J} \exp(\beta' x_{jt}) \cdot p(j \in C_t)} \tag{13}$$

where $P(i \in C_t)$ is the probability that alternative i is in individual t's choice set C_t and J is the number of alternatives in the full choice sets. $P(i \in C_t)$ is specified as the cumulative distribution function of normal distribution.

$$P(i \in C_t) = \Phi(\gamma' Y_{it}) \tag{14}$$

where Y_{it} is the vector of attributes of individual t for alternative i and those associated parameters γ.

We use competing destinations model of Fotheringham (1988) as the base two-stage choice model and put unobserved heterogeneity in alternatives and taste variation across individuals into this model. There are two utility functions in CD model, one is on choice set generation stage and the other is on alternative choice stage. To explain explicitly, we only specified the heterogeneity on the alternative choice stage. In practice, researcher can specify the heterogeneity both or either utility functions.

Incorporating the specification of HLK and ML into CD model, the utility of alternative i for individual t in the alternative choice stage is specified as

$$U_{it} = (\beta + \mu_t)' X_{it} + T\zeta_t + \varepsilon_{it} \tag{15}$$

where μ_t represents the taste variation across individuals, T means unobserved heterogeneity in alternatives.

Assuming ε_{it}'s are identically and independently distributed with extreme value distribution of type I, μ_t, ζ_t follow normal distribution, and $\varepsilon_{it}'s, \mu_t, \zeta_t$ are independent of each other. The probability of individual t choosing alternative i is

$$P_t(i) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\Phi(\gamma' Y_{it}) \cdot \exp((\beta + \mu_t)' x_{it} + T\zeta_t)}{\sum_{j=1}^{J} \Phi((\gamma' Y_{jt}) \cdot \exp((\beta + \mu_t)' x_{jt} + T\zeta_t)} d\mu_t d\zeta_t \tag{16}$$

This model is called heterogeneous competing destinations (HCD) model henceforth. The HCD model can conduct unobserved heterogeneity in alternatives, taste variation across individuals, and heterogeneity in individual's choice sets simultaneously.

3. Data

This paper studies the mode and company choice behavior of public transportation passengers traveling between Taipei and Tainan of Taiwan, R.O.C. The metropolitan population for these two cities were about 3 millions and 1 million respectively in 2002. The travel distance is about 300 kilometers. The public transportation serving this route includes three modes, i.e., air, bus, and rail. Air and bus transportation are all privately owned while the monopolistic Taiwan Rail is government owned and operated. Air transportation included Far Eastern Air Transport (FAT air), Trans Asia Airways (TNA air), and UNI Air. Bus transportation included Kuo-Kuang Motor Transport (KK bus), United Highway Bus (UN bus), and Ho-Hsin Bus (HH bus). FAT air is the oldest operator while UNI air is the newest operator in this route.

Kuo-Kuang bus was privatized in 2001. In the time under the name Taiwan Motor Transport, it had monopolized this route for nearly twenty years. UN bus is the first legal private bus company operating in this route. It adopts low fare and high frequency policy. HH bus offers luxury service with the highest fare. Each seat in its bus is equipped with a personal LCD screen with six channels and several computer games.

Taiwan Rail has two classes of service, the express train Tze-Chiang (TC train) and semi-express train Chu-Kuang (CK train). It has another slower train Fu-Hsing. But only a few long distance travelers take that service so it is not included in this study.

This paper adopted the stated preference approach to avoid the variation problem of revealed preference data. A customized computer questionnaire that could accommodate individual traveler's travel choice situation was designed to reduce the response bias. The questionnaire first asked each traveler's trip status quo that included his/her social-economic characteristics (sex, age, employment status, personal and household income), the chosen alternative and its departure time and date, the date he/she decided to make the trip, the type of ticket (one way, return, or advance purchase), and the access travel time and waiting time for each alternative. The above information is used to design the choice scenarios. In the pilot survey we found that travelers generally had difficulty to report the egress time so it was discarded in this research.

It is difficult for travelers to choose among eight alternatives in an experiment. So we used a two-step approach to reconstruct an eight alternative choice situation into four simpler choice situations. In the first step, eight alternatives were randomly assigned to three choice scenarios so each scenario would include only two or three alternatives. The alternatives in each scenario should consist of alternatives from different modes. That means alternatives from the same mode will not appear in the same scenario. The actual chosen alternative and its actual attribute levels are always included in the first scenario.

Each scenario includes four attributes. They are access and waiting time, in-vehicle travel time, fare, and frequency. The attribute levels of access and waiting time for different alternatives were obtained from the reported data of individual traveler. The levels of fare were obtained from a pre-collected database according to each individual traveler's personal characteristics. Age, time and date of travel, advance purchase, and type of ticket will all affect fare. There are three levels (high, medium, and low) for in-vehicle travel time, fare, and frequency for each alternative where the medium levels are set by the status quo.

The attribute levels of in-vehicle travel time and fare in the competing alternatives are designed according the tradeoff concept. For example, if the preferred alternative is air in the first scenario, then the in-vehicle travel time and fare for rail and bus alternatives in the second and third scenarios will be decreased to make them more attractive. Computer does this adjustment automatically according to preset criteria. The purpose is to find more precise tradeoff relationships between attributes.

In the second step, we obtain the alternatives and their attribute levels directly from three preferred alternatives in the previous three scenarios to create the fourth scenario. The preferred alternative in this scenario will be the most preferred alternative.

We used choice-based sampling to get the survey sample. Rail and bus travelers were surveyed onboard and air travelers were surveyed in the airport by face-to-face interview method. The survey results are directly recorded into the notebook computer. We successfully surveyed 455 travelers that were almost evenly distributed among eight alternatives in March 2002.

4. Empirical results

In this empirical study, we first construct a MNL model as the base for utility specification. Secondly, we illustrate the estimating results of HLK, ML, LCM, CD, and HCD models, including calibration process, explanatory power, and significance of parameters. Thirdly, the comparison between above heterogeneous models has been implementing to discuss model performance and value of time. Finally, we propose two scenarios to test sensitivity of these models.

The best specification and estimating results of MNL model are shown in Table 1. All explanatory variables can be classified into four groups. A brief description is as follows.

1. Alternative specification constants: FAT is specified as base alternative.

2. Inertia variables: considering individual preferred the same mode with reveal preference, we specified air, bus, and train inertia variable.

3. Level of service attributes: travel cost and travel time (including in-vehicle time and out-vehicle time) are specified to common variables. Frequency (every day) variable is specified to three air alternatives. Frequent flier card represents the reward strategy of airlines for loyal customer. Planning days is specified to bus alternatives to catch the flexible of frequency modulation.

4. Socio-economic variables: according to the travel cost, all alternatives classify into three groups, as three air alternatives, TC train and HH bus (most expensive alternative among respective mode), and the others. Then, we specified individual's personal income on first two groups. Employment status is specified to air alternatives to represent that business trip have higher probabilities of choosing air alternatives. Age variable is specified to UN bus and HH bus because those are new bus companies.

We can see that MNL model has relatively good explanatory power and all explanatory variables have correct signs and are quite significant. The signs of travel cost and travel time variables are negative as expected. The positive signs of two flight specific variables show that travelers have higher probability of choosing airlines when more flights are supplied and frequent flier cardholders have higher probabilities of choosing airlines. We also find that the effects of frequent flier cards on three airlines are not significantly different. Planning days variable shows how many days earlier the travelers planning their trips. The negative sign of this variable shows that travelers planning their trips early will have less probability of choosing bus.

The positive signs of two personal income variable and their coefficients' magnitudes show that higher income travelers have the highest probabilities of choosing three airlines, followed by TC train and HH bus which are more expensive than other train and bus alternatives. The positive sign of employment status variable shows that employed travelers have higher probabilities of choosing air alternatives. The negative sign of age variable demonstrates that elder travelers have less probabilities of choosing UN bus and HH bus, which are two new bus companies.

4.1 Estimation

Maximum simulated likelihood (Train, 2003) and Gauss programming language (Aptech Systems, 1995) are used to calibrate following models involving the evaluation of multi-dimensional integral. We apply the Halton sequences (Bhat, 2001) to increase the efficiency of parameter calibration. 75, 100, and 150 random draws have been specified in accordance with 3, 5, and 8 dimensional integral. All estimation result is present as Table 1 and Table 2.

HLK model is built to conduct the unobserved heterogeneity in alternatives. The specification is adding a standard deviation term to each alternative respectively. For identification and normalization, it can only specify 7 (J-1) standard deviation terms. There are two ways to normalize HLK model (Walker, 2001). In our calibration process, we find that although estimating eight times of identified HLK model is more time consuming, but the estimating results is more stable. Hence, we adopt above method to normalize HLK model. The results show that specifying the standard deviation of FAT air constant to zero can have the best log-likelihood value. It represents that the heterogeneity of FAT air is the minimum. UN bus has the maximum heterogeneity. For consistent comparison, we still retain insignificant standard deviation terms.

In transportation and marketing literatures, there are two types of discrete choice model to deal with taste variation across individuals. The main difference depends on the assumption of taste variation. ML model adopts continuous probabilistic distribution to involve taste variation in contrast to discrete finite segments of LCM. Hence, we construct both ML and LCM model to compare which model has better explanatory power.

Table 1. The Coefficients of MNL, HLK, ML, and LCM Models (t Values in Parentheses)

Model	MNL	HLK	ML	LCM	
Explanatory variable				Segment 1	Segment 2
FAT air	0 (-)	0 (-)	0 (-)	0 (-)	0 (-)
standard deviation	-	-	-	-	-
TNA air	-0.15 (-0.7)	-0.18 (-1.0)	-0.22 (-0.8)	0.01 (0.03)	-10.0 (-4.0)
standard deviation	-	0.01 (0.02)	-	-	-
UNI air	-0.02 (-0.0)	-0.34 (-1.1)	-0.09 (-0.2)	0.11 (0.3)	-0.89 (-0.7)
standard deviation	-	0.98 (1.9)	-	-	-
KK bus	2.26 (1.4)	1.40 (0.9)	6.60 (2.5)	4.27 (2.5)	-5.28 (-4.3)
standard deviation	-	0.88 (0.8)	-	-	-
UN bus	5.53 (3.2)	5.32 (3.5)	10.8 (3.8)	7.52 (4.2)	8.64 (3.7)
standard deviation	-	1.10 (2.2)	-	-	-
HH bus	4.44 (2.8)	2.26 (2.1)	9.80 (3.7)	5.83 (3.4)	9.44 (4.3)
standard deviation	-	3.43 (6.2)	-	-	-
TC train	3.84 (2.9)	3.41 (3.7)	8.31 (4.1)	4.93 (3.5)	8.96 (4.2)
standard deviation	-	1.41 (2.7)	-	-	-
CK train	2.89 (1.9)	2.34 (1.9)	8.17 (3.3)	4.62(2.9)	6.20 (2.7)
standard deviation	-	1.07 (1.6)	-	-	-
Air inertia	2.67 (7.1)	2.98 (6.9)	17.7 (2.6)	10.6 (4.2)	5.31 (2.9)
standard deviation	-	-	12.5 (2.4)	-	
Bus inertia	1.78 (5.0)	2.58 (7.8)	3.18 (4.2)	1.68 (3.0)	
Train inertia	0.99 (2.8)	1.38 (3.4)	0.47 (0.8)	1.15 (2.5)	
Travel cost (1000 NT$)	-3.17 (-5.1)	-3.65 (-5.6)	-9.75 (-5.0)	-5.62 (-5.5)	
Travel time (1000 minutes)	-11.2 (-3.0)	-13.5 (-2.9)	-30.6 (-3.8)	-20.7 (-5.7)	-5.29 (-3.7)
standard deviation	-	-	20.6 (3.5)	-	
Flight frequency	0.21 (2.0)	0.18 (3.5)	0.21 (1.7)	0.23 (2.1)	
Frequent flier card	1.12 (4.8)	1.27 (5.4)	1.22 (4.0)	1.34 (5.5)	
Planning days (bus)	-0.11 (-2.4)	-0.13 (-2.3)	-0.26 (-2.6)	-0.11 (-2.6)	
Personal income (air) (1000 NT$)	0.05 (4.0)	0.06 (3.8)	0.15 (3.7)	0.09 (4.1)	
Personal income (HH bus,TC train)	0.03 (3.0)	0.03 (2.5)	0.05 (4.0)	0.02 (2.3)	
Employment status (air)	1.91 (3.5)	2.16 (3.7)	6.11 (3.5)	3.00 (3.2)	
Age (UN bus, HH bus)	-0.04 (-2.7)	-0.05 (-3.0)	-0.07 (-3.3)	-0.04 (-3.3)	
Segment parameter	-	-	-	-0.85 (-4.9)	-1.97 (-11)
Sample size	455	455	455	455	
Number of parameters K	19	26	21	30	
$LL(0)$	-946.13	-946.13	-946.13	-946.13	
$LL(\hat{\beta})$	-648.28	-646.92	-633.68	-636.96	
ρ^2	0.315	0.316	0.330	0.327	
$\bar{\rho}^2$	0.295	0.289	0.308	0.295	

Note: $\bar{\rho}^2 = 1-(LL(\hat{\beta})-K)/LL(0)$

ML model is used to conduct that parameters across individuals present variation caused by taste variation. Hence, we specify heterogeneous parameters to the form of random coefficient, and use ML model to construct random coefficient logit model. After various specifications, we find that coefficients of air inertia and travel time show standard normal distribution. Both coefficients of standard deviation are significantly and means of both variables are different to MNL model.

LCM used probabilistic latent class concept to segment all sample into a few finite segments. The segments should be determined in advance. Then LCM can be specified and estimated. Each segment represents respective choice behavior. It means that parameters in different segment may not be identical. Hence, in the process of calibration, we first specify respective parameters in each segment. Then if the result of likelihood ratio test shows that separate specification could not be better than identical specification, it represents parameters are

homogeneous across segments. The results show that coefficients of air inertia and travel time have response heterogeneity. We estimate 2, 3, and 4 segments of LCM and those respective BIC are 728, 764, and 790. We can find that two segments of LCM have the best explanatory power. For comparison consistently, we specify the segment variable to a constant. According to equation (10), we can calculate that probabilities of both segments are 0.75 and 0.25.

Table 2. The Coefficients of CD and HCD Models (t Values in Parentheses)

Model	MNL	CDM	HCDM
Explanatory variable in choice set generation			
Constant	-	2.48 (2.3)	2.50 (2.4)
Travel cost-Low income	-	-5.48 (-4.0)	-7.93 (-5.9)
Travel time-High income	-	-5.04 (-1.9)	-5.33 (-1.5)
Flight frequency	-	1.03 (4.8)	1.67 (6.8)
Planning days (bus)	-	-0.13 (-2.6)	-0.29 (-2.6)
Personal income (HH bus,TC train)	-	0.02 (2.0)	0.03 (3.4)
Air-standard deviation	-	-	1.22 (1.2)
Bus-standard deviation	-	-	3.71 (4.1)
Explanatory variable in alternative choice			
FAT air	0 (-)	0 (-)	0 (-)
TNA air	-0.15 (-0.7)	-0.28 (-3.0)	-0.34 (-4.0)
UNI air	-0.02 (-0.0)	-0.29 (-3.1)	-0.36 (-4.0)
KK bus	2.26 (1.4)	1.94 (2.0)	-0.05 (-2.2)
UN bus	5.53 (3.2)	5.07 (3.3)	5.12 (-0.7)
HH bus	4.44 (2.8)	4.83 (2.9)	5.08 (-0.7)
TC train	3.84 (2.9)	4.79 (3.9)	10.1 (3.4)
CK train	2.89 (1.9)	3.02 (1.9)	7.54 (1.3)
Air inertia	2.67 (7.1)	2.48 (7.3)	9.62 (2.1)
standard deviation	-	-	7.07 (1.8)
Bus inertia	1.78 (5.0)	1.88 (5.6)	15.8 (2.7)
Train inertia	0.99 (2.8)	1.08 (3.2)	1.55 (3.3)
Travel cost (1000 NT$)	-3.17 (-5.1)	-2.55 (-4.2)	-6.87 (-5.5)
Travel time (1000 minutes)	-11.2 (-3.0)	-14.7 (-4.4)	-30.3 (-5.0)
Flight frequency	0.21 (2.0)	-	-
Frequent flier card	1.12 (4.8)	1.13 (5.3)	1.21 (4.9)
Planning days (bus)	-0.11 (-2.4)	-	
Personal income (air) (1000 NT$)	0.05 (4.0)	0.05 (3.6)	0.16 (4.4)
Personal income (HH bus,TC train)	0.03 (3.0)	-	-
Employment status (air)	1.91 (3.5)	1.80 (3.3)	3.71 (3.2)
Age (UN bus, HH bus)	-0.04 (-2.7)	-0.03 (-2.5)	-0.09 (-3.9)
Bus-standard deviation	-	-	9.08 (3.0)
Sample size	455	455	455
Number of parameters K	19	22	26
$LL(0)$	-946.13	-946.13	-946.13
$LL(\hat{\beta})$	-648.28	-637.22	-615.13
ρ^2	0.315	0.326	0.349
$\bar{\rho}^2$	0.295	0.303	0.322

CD model is used to deal with the heterogeneity in choice sets across individuals. There are two utility functions in CD model. Besides those variables specified in MNL model, we add another two explanatory variables: travel cost-low income and travel time-high income. If individual's personal income is less than NT$ 40,000, then the travel cost of each alternative is specified to the former variable respectively; otherwise, this variable is specified to zero. The specification of travel time-high income is similar to above illustration. All variables have been specified to either choice set generation stage or alternative choice stage. The best specification of CD model is shown in Table 2. All explanatory variables have correct signs

and are quite significant.

We have construct HLK, ML, LCM, and CD models to conduct respectively three major heterogeneities which are derived from alternative, taste variation, and choice sets. Now, we propose a synthetic model, HCD, to conduct simultaneously all three major heterogeneity involved in random utility model. The best specification of HCD model is shown in Table 2. Most variables are similar to CD model except of standard deviation of mode specific constant and air inertia. In this intercity travel choice of combing transport modes and operating companies (or service class), the heterogeneity in alternatives is representing as the type of transport mode. The result shows that mode heterogeneity in alternatives exists on air and bus alternatives in choice set generation stage, and also exists on bus alternatives in alternative choice stage. Furthermore, taste variation across individuals exists on air inertia variable.

4.2 Comparisons

In this section, we will compare above heterogeneous models from explanatory power, parameter difference, and sensitivity. Firstly, we use nested and non-nested likelihood ratio test (Ben-Akiva and Lerman, 1985) to judge which model has the best explanatory power. All heterogeneous models are significantly better than MNL model except HLK model. In the heterogeneity of taste variation, ML model adopts the specification of continuous probabilistic distribution and LCM used the specification of discrete finite segments. The test result shows that ML model is significantly better than LCM. It means the specification of probabilistic distribution has better explanatory in this empirical study.

Among four single-resource heterogeneous models (HLK, ML, LCM, and CD model), ML model have a fewer explanatory variables and better log-likelihood value. We have known that ML model has better explanatory power and the magnitude of heterogeneity in taste variation is greater than the others. Furthermore, we compare HCD model incorporating multiple-resource heterogeneity and ML model (single-resource). The result of non-nested likelihood ratio, $\Pr(\bar{\rho}^2_{HCD} - \bar{\rho}^2_{ML} > 0.014) \le 1.01E-8$, shows that the probability of such a difference would be exceed for a sample of 455 observations and 8 alternatives is less than 1.01E-8. So HCD model is significantly better than ML model. This result is shown that considering multiple-heterogeneity completely can improve model's explanatory power significantly.

Evaluating the absolute parameter across models is not informative because of scale difference. However contrasts of parameter in relative ratio are very informative. We summarize estimated values of time (VOT) across above models in Table 3. Both VOT of two segments in LCM are 221 and 56. Segment 1 is similar to MNL and HLK model, but segment 2 is only one-forth of MNL. We use segment probability as weights to calculate the mean of VOT in LCM. The result is shown in Table 3. VOT of CD model is separated to twp components: low-income and high income. The relative ratio of observations between low-income and high-income is used to be weight to calculate the mean of VOT. HCD model is used similar method to calculate its mean of VOT. The VOT of two most explanatory models (HCD, ML) are quite close (187, 188 NT$/hr). The VOT of other less explanatory models are quite different. This result shows that ignoring heterogeneity in constructing discrete choice model would lead to bias estimation on value of time.

Table 3. Estimated Values of Time in Discrete Choice Models

	MNL	HLK	ML	LCM		CD		HCD	
				Segment 1	Segment 2	Low Income	High Income	Low Income	High Income
Value of time (NT$/hr)	212	222	188	221	56	176	354	149	274
				180		230		187	

Finally, we use two best explanatory models, ML and HCD, to implement sensitivity analysis. We compare the change in absolute choice shares in response to a change in the level of attributes across individuals. Two scenarios are proposed to discuss the change in absolute choice shares.

1. Scenario 1 (S1): a 5% decrease in travel cost and a 5% increase in travel time;

2. Scenario 2 (S2): a 5% decrease in travel time and a 5% increase in travel cost.

S1 and S2 are mixed scenarios, composed of changes in travel cost and travel time. S1 strategy use longer travel time to gain lower travel cost and S2 strategy use higher travel cost to shorten travel time. These two scenarios are most common strategies in transportation. Hence, we propose these two scenarios to analysis the change in choice share across alternatives. The results are summary in Table 4.

The sensitivity of S1 in ML model and HCD model is similar except HH bus, as well as S2. In the S1 scenario, choice share of air alternatives are raised and the others are reduced. On the contrary, choice share of air alternatives are reduced and the others are raised in the S2 scenario. Absolute choice share of FAT air have greater increment than other alternatives in S1 scenario. In S2 scenario, absolute choice share of UN bus have greater increment than other alternatives, except HH bus in HCD model, and CK train have less increment than other alternatives. This sensitivity analysis shows that expensive alternatives, like air mode, should adopt cut-price strategy to increase choice share and inexpensive alternatives, like bus and train mode, should improve the level of service in travel time to increase choice share.

Table 4. Comparison of Sensitivity in Four Scenarios

	ML		HCD	
	S1	S2	S1	S2
FAT air	7.02	-6.13	4.96	-4.53
TNA air	1.88	-2.26	1.28	-1.43
UNI air	0.04	-0.03	0.06	-0.07
KK bus	-0.23	0.48	-0.77	1.03
UN bus	-1.76	2.14	-2.38	2.37
HH bus	-0.53	0.54	0.23	-0.26
TC train	-0.56	0.31	-1.90	1.79
CK train	-0.24	0.24	-0.57	0.64

Note: Absolute change in choice share.

5. Conclusions

This paper is to compare the difference of heterogeneous effect caused by multiple sources, including unobserved heterogeneity in alternatives, taste variation, and heterogeneous choice sets. We review related theory of heterogeneity and construct several heterogeneous discrete choice models to deal with various heterogeneities. Furthermore, we also propose a new synthetic model, heterogeneous competing destinations model, to integrate multiple-resource heterogeneity. A stated preference data of intercity travel choice is used as empirical case. There are eight alternatives composed of transport modes and operating companies (or service class).

The empirical results show that all heterogeneous models are significant than MNL model except HLK model. In single-resource heterogeneity, ML model conducting with taste variation has better explanatory power. HCD model is successful integrating multiple heterogeneities simultaneously and its explanatory power is significantly better than the other heterogeneous models. The estimation results show that mode heterogeneity exists on air and bus alternatives in choice set generation stage, and also exists on bus alternatives in alternative choice stage. Furthermore, taste variation across individuals exists on air inertia variable.

In the parameter comparison, we calculate the value of time of estimated models to discuss the heterogeneous effect on the estimation of VOT. The result shows that ignoring heterogeneity in constructing discrete choice model would lead to bias estimation on value of time. In the sensitive analysis, we propose two mixed marketing strategies to analysis the change in choice share across alternatives. We find that expensive alternatives, like air mode, should adopt cut-price strategy to increase choice share and inexpensive alternatives, like bus and train mode, should improve the level of service in travel time to increase choice share.

In this study, we assume the pure error term is distributed of extreme value. So all the interpretation is base on the formulation of logit form to discuss heterogeneity. Although McFadden and Train (2000) had pointed that mixed logit can approximate to any random utility model in specific conditions, we still have more interests if heterogeneous effect is unchanged while the pure error term is specified by other distributions. This issue of interaction between pure error term and heterogeneity is worth further attention.

REFERENCES

Aptech Systems (1995) **Gauss Applications: Maximum Likelihood**. Aptech Systems Inc., Maple Valley, WA.

Baltas, G. and Doyle, P. (2001) Random utility models in marketing research: A survey, **Journal of Business Research, Vol. 51**, 115-125.

Ben-Akiva, M. and Bolduc, D. (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, working paper, Dept. of Civil Engineering, MIT.

Ben-Akiva, M. and Lerman, S.R. (1985) **Discrete Choice Analysis: Theory and Application to Travel Demand**. The MIT Press, Cambridge, MA.

Bhat, C.R. (1997) An endogenous segmentation mode choice model with an application to

intercity travel, **Transportation Science, Vol. 31, No. 1**, 34-48.

Bhat, C.R. (1998) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling, **Transportation Research A, Vol. 32**, 495-507.

Bhat, C.R. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, **Transportation Research B, Vol. 35**, 677-693.

Brownstone, D. and Train, K. (1996) Forecasting new product penetration with flexible substitution patterns, **Journal of Econometrics, Vol. 89, No. 1**, 109-129.

Chintagunta, P.K. (1994) Heterogeneous logit model implication for brands positioning, **Journal of Marketing Research, Vol. 31**, 304-311.

Chintagunta, P.K. (1999) Measuring the effects of new brand introduction on inter-brand strategic interaction, **European Journal of Operational Research, Vol. 118**, 315-331.

Fotheringham, A.S. (1988) Consumer store choice and choice set definition, **Marketing Science, Vol. 6**, 299-310.

Greene, W.H. and Hensher, D.A. (2003) A latent class model for discrete choice analysis: Contrasts with mixed logit, **Transportation Research B, Vol. 37**, 681-698.

Gonul, F. and Srinivasan, K. (1993), Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues, **Marketing Science, Vol. 12 , No. 3**, 213-229.

Hsiao, C. (1986) **Analysis of Panel Data**. Cambridge University Press, Cambridge.

Kamakura, W.A. and Russell, G.J. (1989) A probabilistic choice model for market segmentation and elasticity structure, **Journal of Marketing Research, Vol. 26**, 379-390.

Manrai, A.K. and Andrews, R.L. (1998) Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions, **European Journal of Operational Research, Vol. 111**, 193-215.

McFadden, D. and Train, K. (2000) Mixed MNL models for discrete response, **Journal of Applied Econometrics, Vol. 15, No. 5**, 447-470.

Munizaga, M.A., Heydecker, B.G. and Ortuzar, J.de D. (2000) Representation of heteroskedasticity in discrete choice models, **Transportation Research B, Vol. 34**, 219-240.

Revelt, D. and Train, K. (1998) Mixed logit with repeated choices: Households' choices of appliance efficiency level, **Review of Economics and Statistics, Vol. 80, No. 4**, 647-657.

Train, K. (1998) Recreation demand models with taste variation over people, **Land Economics, Vol. 74, No. 2**, 230-239.

Train, K. (2003) **Discrete Choice Methods with Simulation.** Cambridge University Press, Cambridge.

Walker, J. (2001) Extend discrete choice models: Integrated framework, flexible error structures, and latent variables, Ph. D. Dissertation, Department of Civil and Engineer, MIT.

Yang, C.W., Duann, L.S., Chen, Y.C. and Fang, W.P. (2003) Intercity bus choice models with choice set generation and heterogeneity, **Journal of Eastern Asia Society for Transportation Studies, Vol. 5**, 1518-1533.