DYNAMIC DISCRETE CHOICE MODELS CONSIDERING UNOBSERVED HETEROGENEITY WITH MASS POINT APPROACH

Yoriyasu SUGIE
Professor
Graduate School for IDEC
Hiroshima University
Kagamiyama 1-5-1, Higashi-Hiroshima
739-8529 Japan
Fax: +81-824-24-6919
E-mail: ysugie@ipc.hiroshima-u.ac.jp

Junyi ZHANG
Project Development Division,
Pacific Consultants Co., Ltd.
Sekido 1-7-5, Tama-Shi, Tokyo
206-8550 Japan
Fax: +81-423-72-6393
E-mail: Syunkitsu.Chou@ss.pacific.co.jp

and

Akimasa FUJIWARA
Associate Professor
Graduate School for IDEC
Hiroshima University
Kagamiyama 1-5-1, Higashi-Hiroshima
739-8529 Japan
Fax: +81-824-24-6921
E-mail: afujiw@ipc.hiroshima-u.ac.jp

Abstract: This paper aims to develop new dynamic discrete mode choice models treating unobserved heterogeneity based on the Mass Point approach using short-term panel data. In order to examine the effectiveness of the approach, Laird- and Lindsay-type models are estimated by using three-period individual panel data collected before and after the opening of a new railway station in Hiroshima. It was consequently shown that MP models are superior to that of models without unobserved heterogeneity, while dynamic MP models are superior to static MP models without state dependence with respect to the goodness-of-fit of models.

1. INTRODUCTION

Dynamic discrete choice models have developed remarkably since the latter half of 1980s to capture changes in individual travel behavior. They have tended to use panel data, which is collected repeatedly from the same sample over time. However, these models assume a homogeneous structure among individuals. The individual specific effects resulting from omitted variables greatly influence the estimated parameters in models, so that the assumed homogeneity may lead to erroneous conclusions which are in turn applied to explain transportation phenomena, and to evaluate transportation policies.

Heterogeneity can be defined as differences in travel behavior caused by temporary or individual characteristics. When short-term panel data is used for study, it is difficult to consider the effect of temporary characteristics and only individual heterogeneity can be taken into account. Therefore, in the rest of this paper, the term heterogeneity simply refers to the individual heterogeneity.

Heterogeneity can be classified as the observed and the unobserved. Eqn (1) represents the utility function of a discrete choice model which considers this.

$$U_{ijt} = \sum_{k=1}^{K} \beta_{k,i}\, x_{k,ijt} + \delta_{ij} + \varepsilon_{ijt} \tag{1}$$

where, i, j, t : indexing individual, alternative and time,
U_{ijt} : utility function,

$x_{k,ijt}$: kth explanatory variable,
$\beta_{k,i}$: individual parameter of $x_{k,ijt}$,
δ_{ij} : parameter of the unobserved heterogeneity and
ε_{ijt} : error term following an identical and independent distribution for i, j, t.

$\beta_{k,i}$ corresponds to the observed heterogeneity. For example, individual travel behavior may change according to individual characteristics including sex, age, income, etc. which are measurable. Much research has been done concerning observed heterogeneity based on, for examples, market segmentation approaches, taste variation models and conjoint analysis.

Unobserved heterogeneity means the differences in travel behavior caused by unmeasurable characteristics, such as taste, attitude, motivation, etc., or omitted variables. For example, provided that the measurable characteristics described above are not considered in the model, they cause a problem of unobserved heterogeneity if they are statistically significant. Furthermore, unobserved heterogeneity can also partially explain the effect of state dependence or serial correlation which is not incorporated into the model. Therefore, the unobserved heterogeneity is a very general concept in travel behavior models. Of course, if the measurable characteristics, state dependence and serial correlation affect the result of model estimation, they must obviously be incorporated into the models. However, there exist many complicated individual characteristics affecting travel behavior and not all of them can be measured precisely.

In this paper, dynamic discrete mode choice models which consider the unobserved heterogeneity caused by unmeasurable characteristics, are developed based on the Mass Point (MP) approach when only short-term panel data is available.

2. UNOBSERVED HETEROGENEITY AND DYNAMIC DISCRETE CHOICE MODELS: AN OVERVIEW

2.1 Approaches Treating Unobserved Heterogeneity

Consider the following utility function with the parameter δ_{ij} of unobserved heterogeneity.

$$U_{ijt} = \sum_{k=1}^{K} \beta_k x_{k,ijt} + \delta_{ij} + \varepsilon_{ijt} \tag{2}$$

where, β_k is invariant among individuals.

There exist two approaches to estimate eqn (2) using a maximum likelihood (ML) method. The first one is a fixed-effects approach in which δ_{ij} is assumed to be determinate. The second is a random-effects approach in which δ_{ij} is assumed to be a random variable.

Concerning the fixed-effects approach, the conventional ML method cannot be applied because the number of parameters increases with the number of observations (Neyman and Scott, 1948). Andersen (1970) showed that the conditional maximum likelihood (CML) method should be applied in this case. Chamberlain (1980) specified a logit model with fixed-effects using the CML method. Nevertheless, several limitations exist in this logit model: (a) ordinary maximum likelihood estimation requires sufficient statistic for δ_{ij}, (b) the parameters of time invariant variables cannot be estimated, (c) only individuals whose choice results change temporally, can be handled and (d) δ_{ij} cannot be estimated consistently with short-term panel data.

In contrast to the fixed-effects approach, the random-effects approach does not have such problems. This approach can be further classified into a parametric approach in which continuous distribution of δ_{ij} is assumed, and a MP approach in which discrete distribution of δ_{ij} is assumed.

Probability density function of dependent variable y_{ijt}, conditional on explanatory variables vector $\mathbf{x}_{ijt}$, parameter vector $\boldsymbol{\beta}$ of $\mathbf{x}_{ijt}$ and parameter δ_{ij}, of unobserved heterogeneity, is defined as $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$. The continuous distribution function of δ_{ij} is assumed as $F_i(\delta_{ij}|\mathbf{x}_{ijt},\boldsymbol{\alpha})$ indexed by a finite parameter vector $\boldsymbol{\alpha}$. Then, choice probability P_i of individual i can be represented based on a parametric approach as follows:

$$P_i = \int_{\delta_{i1}} \cdots \int_{\delta_{ij}} \cdots \left\{ h_i\,(y_{ijt}|\mathbf{x}_{ijt},\, \boldsymbol{\beta},\, \delta_{ij})\, dF_i\big(\delta_{i1}|x_{i1t},\, \alpha\big) \cdots dF_i\big(\delta_{ij}|x_{ijt},\, \alpha\big) \cdots \right\} \tag{3}$$

where, suffixes i, j and t are defined in eqn (1).

Eqn (3) is a model with random-effects assuming a continuous distribution of δ_{ij}. Concerning its estimation methods, one method directly calculates its integral and another approximates its integral (Davies and Crouchley, 1985; Davies, 1984). Beta, Gamma, normal, log-normal, exponential, Weibull and log-logistic distributions have been used in the parametric approach (Heckman and Willis, 1977; Pickles, 1983; Davies, 1984; Davies and Crouchley, 1985; Dunn and Wrigley, 1985; Hensher and Mannering, 1994). Many applications of this approach can be found in the fields of econometrics and geography (Heckman and Willis, 1977; Davies, Crouchley and Pickles, 1982; Davies, 1984; Dunn and Wrigley, 1985; Davies, Pickles and Crouchley, 1983). Uncle (1987) applied the Beta-logistic model of Heckman and Willis (1977) to analyze the mode choice behavior of shopping trips for the first time in the transportation field, and were able to confirm the nature of the unobserved heterogeneity. Though integral calculation or its approximation is needed in a parametric approach, no theoretical problems exist. However, the integral calculation or its approximation makes it difficult to calibrate models if the distribution form of δ_{ij} is complicated. Besides that, the estimated parameters of explanatory variables respond sensitively to different distributions of δ_{ij}(Heckman and Singer, 1984).

Since δ_{ij} cannot be precisely measured by the analyst, specifying the distribution of δ_{ij} is not an easy task. Unlike the parametric approach, choice probability P_i of individual i can be represented based on the MP approach as follows:

$$P_i = \sum_{k=1}^{m} h_i\,(\,y_{ijt}\,|\,\mathbf{x}_{ijt},\, \boldsymbol{\beta},\, \xi_{kj}\,)\,\rho_k \tag{4}$$

where, ξ_{kj} and ρ_k are position and weight parameters of kth MP, and m is the total number of MP.

Moreover, there exists a restriction condition imposed on ρ_k .

$$\rho_k \geq 0 \text{ and } \sum_{k=1}^{m} \rho_k = 1 \text{ for } k=1, 2, \ldots, m \tag{5}$$

Thus, MP must be a discrete point which reflects the distribution characteristics of δ_{ij}. Furthermore, if several conditions are satisfied, the model can be estimated using a sufficiently small number of MP (Laird, 1978; Davies and Crouchley, 1984; Heckman and Singer, 1984; Davies and Crouchley, 1985; Reader, 1993).

Initially, MP approach was applied using a fixed number of MP. After that, Simar (1976) extended the initial MP approach to compound Poisson distribution family and sets a bound on MP which has been improved by Lindsay. Simar (1976) also demonstrated consistency, i.e. the ML estimator converges weakly with probability one to the true measure.

Laird (1978) extended some of Simar's results to the general problem. In particular, the property of self-consistency was identified which lead to an extensive set of conditions under which the ML estimator was a discrete measure. It was conjectured that for "well-behaved (analytical) unimodal probability densities" there is no more MP needed than observations in the sample. It was shown that this is true for virtually all densities by Lindsay. Since then, Lindsay (1981, 1983a,b) established the general conditions under which the MP approach can be applied. Lindsay's research must be regarded as setting a benchmark in the history of studying the MP approach.

Much research has been done by using the MP approach in the fields of econometrics and geography (Davies and Crouchley, 1984; Heckman and Singer, 1984; Davies and Crouchley, 1985; Reader, 1993). However, little research has been carried out in the field of transportation. In 1987, two pieces of researches in the field of geography: Dunn, Reader and Wrigley (1987), Davies and Pickles (1987) were introduced in the special issue on longitudinal data methods in the journal of "Transportation Research". Kitamura *et al.* (1996) introduced the MP approach into the doubly-censored Tobit model to explain the time allocation of two types of discretionary activities.

2.2 Dynamic Discrete Choice Models Treating Unobserved Heterogeneity

The utility function of a dynamic model incorporating state dependence can be expressed as eqn (6).

$$U_{ijt} = f(\, y_{ij,t-1}\,, \dots, y_{ij,t-q} \text{ and } \mathbf{x}_{ij,t}\,, \mathbf{x}_{ij,t-1}\,, \dots, \mathbf{x}_{ij,t-q}\,) \tag{6}$$

where, $y_{ij,t-1}, \dots, y_{ij,t-q}$: choice results of individual i for alternative j at time t-1,...,t-q
$\mathbf{x}_{ij,t}, \dots, \mathbf{x}_{ij,t-q}$: explanatory variable vectors.

The dynamic model assumes that current behavior is influenced by the previous one. But, how to take this previous information into account in the model depends on the length of the interval in the panel data. The dynamic models of travel behavior are generally based on those developed in the field of econometrics from 1960s to 1980s. As far as we know, Burnett (1974) applied the first disaggregate dynamic discrete choice model in the field of transportation research to study spatial choice behavior. Her dynamic model considered only one spatial alternative at a time.

In the latter half of 1980s, a great deal of panel data including Dutch National Mobility Panel Data emerged in the field of transportation research. And a multinomial probit model which considers state dependence and serial correlation simultaneously by Daganzo and Sheffi (1982) and Markov model, etc. were developed. The dynamic analysis era of travel behavior had started.

However, in order to capture the changes in travel behavior properly, it is important to distinguish state dependence and unobserved heterogeneity. While the early approaches, referred to above, treated these two variables together, several researchers subsequently have attempted to separate state dependence and unobserved heterogeneity.

Heckman has developed dynamic models with fixed-effects and random-effects in the analysis of labor force participation, in which only two alternatives were considered in late 1970s. Tardiff (1979) extended the dynamic models to the case of more than two alternatives in the analysis of spatial choice. Tardiff was one of the first to make an attempt to extend discrete choice methodology by introducing state dependence effects and serial correlation into the utility functions (Fischer and Nijkamp, 1987). He proposed a dynamic model which included the previous choice behavior as an explanatory variable.

$$U_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \sum_{j'} \gamma_{jj'} C_{ij',t-1} + \delta_{ij} + \varepsilon_{ijt} \tag{7}$$

where, $\mathbf{x}_{ijt}$: explanatory variable vector of individual i for alternative j at time t,
$\boldsymbol{\beta}$: parameter vector of $\mathbf{x}_{ijt}$,
$C_{ij',t-1}$: dummy variable of alternative j with $C_{ij',t-1} = 1$, if individual i chose j' at time t-1 and $C_{ij',t-1} = 0$, otherwise,
$\gamma_{jj'}$: parameter of $C_{ij',t-1}$,
ε_{ijt} : error term following identical and independent distribution for i, j, t.

Tardiff (1979) considered that serial correlation is caused by unobserved heterogeneity. The second term on the right-hand side of eqn (7) accounts for the first-order Markov effects. Tardiff suggested treating δ_{ij} as fixed-effects rather than random-effects.

In contrast to Tardiff (1979), Heckman (1981) derived a general dynamic model which can analyze the structure of discrete choices made over time from a direct consideration of the complex error component structure (random-effects approach). Heckman's dynamic model can be expressed as follows:

$$y_{it} = v_{it} + \varepsilon_{it} \tag{8}$$

$$v_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \sum_{k=1}^{\infty} \gamma_{t-k,t}\, d_{i,t-k} + \sum_{k=1}^{\infty} \lambda_{k,t-k} \prod_{q=1}^{k} d_{i,t-q} + G(L)\, y_{it} \tag{9}$$

$$\begin{cases} y_{it} \geq 0 & \text{if and only if } d_{it} = 1 \\ y_{it} < 0 & \text{if and only if } d_{it} = 0 \end{cases} \tag{10}$$

where, covariance of ε_{it} and v_{it} is zero, d_{it} is a dummy variable denoting the occurrence of the event under consideration. And, $G(0) = 0$, $G(L) = g_1L + g_2L^2 + \ldots + g_KL^K$ is a general lag operator, $L^k y_{it} = y_{it-k}$.

The distribution of d_{it} is generated by the distributions of ε_{it} and v_{it}, while ε_{it} is assumed to be normally distributed with mean zero. This normality assumption generates a general model which is able to account for a wide variety of error structures for serially correlated unobserved variables. The initial conditions for y_{it} and d_{it} are assumed to be predetermined or exogenous.

The second term in eqn (9) which is assumed to be finite, represents state dependence effects. The third term denotes the cumulative effect on current choices. The last term captures the action of habit persistence. By imposing various restrictions on the parameters of the general model, a variety of models such as Markov model, renewal processes, the multinomial probit model of Daganzo and Sheffi (1982), etc. emerge as special cases.

Subsequently, extending the research of Heckman (1981), Kitamura and Bunch (1990) distinguish explicitly between unobserved heterogeneity and state dependence in a dynamic ordered-response probit model, in order to analyze car ownership using panel data based on a parametric approach.

Much progress has been made in the field of dynamic discrete choice modeling. However, it is not doubtful that several problems are not yet satisfactorily solved, those are, attrition bias, initial conditions, uncertainty of choice set, model identifiability and data requirements when explanatory variables are "slow moving" or behavioral changes of interest are infrequent (Fischer and Nijkamp, 1987; Kitamura, 1990).

3. DESCRIPTION OF MASS POINT APPROACH

3.1 Laird-type Mass Point Approach

To apply Laird-type approach, several conditions on $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$ in eqn (3) must be ensured (Laird, 1978; Davies and Crouchley, 1984). The basic conditions are:

1) The functions $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$ are differentiable with respect to δ_{ij} for every possible value of δ_{ij}.
2) There does not exist a set of coefficients $\{a_k\}$ such that $\sum_{i\in L} a_i h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})=1$ for every value of δ_{ij}, where L indexes the $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$ which are distinct. This is an assumption of linear independence of the $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$.
3) For some values of δ_{ij} and for some k, the kth derivatives with respect to δ_{ij} of the $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$ are either all non-negative or non-positive and at least one is non-zero.
4) δ_{ij} varies over a finite interval.
5) There is a range of values of δ_{ij} such that, for any δ_{ij} within the range and for any δ'_{ij} outside the range, $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij}) > h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta'_{ij})$ for all i.

Conditions 1) and either 2) or 3) are sufficient to show that the nonparametric characterization of $F_i(\delta_{ij}|\mathbf{x}_{ijt},\boldsymbol{\alpha})$ is a discrete distribution; the additional condition 4) or 5) is sufficient to show that the number of MP is finite. To satisfy Laird's conditions, Davies and Crouchley (1984) rewrite δ_{ij} as eqn (11) and specify a binary Laird-type MP model as in eqn (12).

$$\mu_{ij} = \frac{1}{1+\exp(\delta_{ij})} \tag{11}$$

$$\begin{aligned}&\mathrm{Prob}\left(S_i \,|\, \boldsymbol{\xi}, \boldsymbol{\rho}, \boldsymbol{\beta}; \{x_{it}\}\right)\\ &= \sum_{k=1}^{m} \prod_{t=1}^{T_i} \left\{ \frac{(1-\xi_k)^{(1-y_{it})}\left[\xi_k \exp(-\boldsymbol{\beta}'\mathbf{x}_{it})\right]^{y_{it}}}{1+\xi_k\left[\exp(-\boldsymbol{\beta}'\mathbf{x}_{it})-1\right]} \right\} \rho_k\end{aligned} \tag{12}$$

where, $S_i = \{y_{it}\,|\,t=1, \ldots, T_i, y_{it}=(0, 1)\}$, T_i is total number of time points that individual i participated in the survey. Note that T_i can differ depending on individuals. ξ_k, ρ_k are defined as eqn (4).

In Laird-type model, eqn (13) is also imposed in addition to eqn (5).

$$0 < \xi_k < 1 \tag{13}$$

In the case of multinomial logit model, eqn (12) will turn to be more complicated.

3.2 Lindsay-type Mass Point Approach

As described above, Laird's conditions are strict and complicated. By contrast, Lindsay's conditions are more flexible (Lindsay, 1981, 1983a, b).

Theorem 1: For a fixed $\boldsymbol{\beta}$ and a finite sample, the ML estimator by MP approach in an identified model is a finite mixture with at most N points, where N is the number of distinct values of $\{y_{ijt}, \mathbf{x}_{ijt}\}$ in the sample. For this property to hold, it is required that $h_i(y_{ijt}|\mathbf{x}_{ijt},\boldsymbol{\beta},\delta_{ij})$ be a bounded function of δ_{ij} for fixed $\boldsymbol{\beta}\in\mathbf{B}$ and $\mathbf{x}_{ijt}\in\mathbf{X}$, and $F_i(\delta_{ij}|\mathbf{x}_{ijt},\boldsymbol{\alpha}) \geq 0$ is non decreasing and right continuous. The form of the log-likelihood function is given:

$$L = \sum_{i=1}^{N} \ln \sum_{k=1}^{N} h_i (y_{ijt} \mid \mathbf{x}_{ijt}, \boldsymbol{\beta}, \delta_{ij}) \rho_k \tag{14}$$

where,

$$\rho_k \in R_N , R_N = \left\{ (\rho_1 , \rho_2 , \ldots , \rho_N) ; \rho_k \geq 0 ; 1 \leq k \leq N ; \sum_{k=1}^{N} \rho_k = 1 \right\} \tag{15}$$

In the Lindsay-type approach, eqn (13) is relaxed as eqn (16).

$$\xi_k \in E \, , \; E \subset (-\infty , \infty) \tag{16}$$

Theorem 2: For each $\boldsymbol{\beta} \in \mathbf{B}$ and $\mathbf{x}_{ijt} \in \mathbf{X}$, if $h_i (y_{ijt} \mid \mathbf{x}_{ijt}, \boldsymbol{\beta}, \delta_{ij})$ is in the exponential family and provided one condition is met, then

$$\sup_{\xi_k \in E,\, \rho_k \in R_N} \left\{ \sum_{i=1}^{N} \ln \sum_{k=1}^{N} \mathbf{h}_i (\mathbf{y} \mid \mathbf{x}, \boldsymbol{\beta}, \xi_k) \rho_k \right\} \tag{17}$$

is attained for a unique mixing distribution $\{ \xi_k , \rho_k \}$. The required condition is that no MP comes from the boundary of E (Lindsay, 1983b). Based on Theorems 1 and 2, a binary Lindsay-type MP model can be built as follows:

$$\text{Prob}\left(S_i \mid \boldsymbol{\xi} , \boldsymbol{\rho} , \boldsymbol{\beta} ; \{ x_{it} \} \right) = \sum_{k=1}^{m} \prod_{t=1}^{T_i} \left\{ \frac{\left[\exp (\boldsymbol{\beta}' \mathbf{x}_{it} + \xi_k) \right]^{1 - y_{it}}}{1 + \exp (\boldsymbol{\beta}' \mathbf{x}_{it} + \xi_k)} \right\} \rho_k \tag{18}$$

Unlike the Laird-type model, the multinomial MP model can be easily developed similarly to eqn (18), which will be explained in section 5.

4. ESTIMATION OF BINARY MASS POINT MODELS

4.1 Ajina Three-period Panel Data

In order to examine the effectiveness of the MP approach, binary Laird- and Lindsay-type models are estimated by using three-period individual panel data collected in June 1989, November 1989 and October 1991 before and after the opening (August 1989) of a new Ajina railway station, located at the west of Hiroshima. The Ajina panel data used in this study was collected for four modes: car, bus, tram and rail, and obtained from a sample of 169 people, all of whom consistently participated in the panel three times each. However, in order to simplify the estimation and comparison of two MP models, binary panel data for car and tram is examined in this section. The modal shares of car and tram in the binary panel data and the shares of each mode in original four modes panel data are shown in Table 1.

Table 1. Shares of each mode in Ajina binary and four modes panel data (%)

Mode	Binary data				Four modes data			
	Jun. '89	Nov. '89	Oct. '91	TM	Jun. '89	Nov. '89	Oct. '91	TM
Car	64.6	66.7	69.7	67.0	41.4	42.6	45.6	43.2
Tram	35.4	33.3	30.3	33.0	29.0	23.7	19.5	23.7
Bus					20.1	25.4	25.4	24.0
Rail					9.5	8.3	9.5	9.1

(TM: temporal mean)

4.2 Estimations of Binary Mass Point Models

Since it cannot be ensured that the log-likelihood functions in Laird- and Lindsay-type MP models are convex over the search area, there is always a risk that numerical optimization will produce misleading results by 'homing in' on a local rather than a global maximum. Davies and Crouchley (1984) used several different initial values for the optimization algorithm to reduce this risk. The same method is adopted in this study.

The estimation results of binary logit models over three time points and a pooled binary logit model are shown in Table 2. Here, a pilot study, based on a segmentation approach, showed that the observed socio-economic attributes, such as sex and age, do not affect the choice behavior significantly. Therefore, only service-of-level variables like time and cost are used.

From Table 2, it is clear that the explanatory variables whose parameters are statistically significant, are only in-vehicle time and egress time for the 1991 model and the pooled model.

Table 2. Estimation results of logit models

Explanatory variable	Jun. 1989	Nov. 1989	Oct. 1991	Pooled model
Access time (min.)	0.008 (0.10)	-0.028 (0.57)	0.070 (0.96)	-0.0004 (0.01)
In-vehicle time (min.)	-0.020 (0.67)	-0.036 (1.30)	-0.101 (2.41)*	-0.041 (2.37)*
Cost (100 yen)	-0.179 (1.23)	-0.022 (0.15)	0.008 (0.13)	-0.048 (1.09)
Egress time (min.)	-0.126 (1.95)	-0.081 (1.61)	-0.135 (2.04)*	-0.096 (3.37)**
No. of transfers	-0.286 (0.44)	0.444 (0.49)	-2.559 (1.46)	-0.214 (0.58)
Initial likelihood	-68.6	-68.6	-68.6	-205.9
Maximum likelihood	-20.4	-26.0	-20.5	-71.9
Adjusted likelihood ratio	0.686	0.601	0.685	0.645
Sample size	99	99	99	297

(t scores in parentheses; *:significant at 5%; **: at 1%)

The estimation results of the binary Laird- and Lindsay-type MP models are shown in Tables 3 and 4, respectively. Supplementary explanations for the Tables are given as follows:

1) The initial likelihood of the MP model is the value of the likelihood when all parameters are 0 except that one of ρ_k is given as 1 because of $\sum_{k=1}^{m}\rho_k = 1$.
2) Since $\rho_k \geq 0$ and $\sum_{k=1}^{m}\rho_k = 1$ hold, $\rho_k = \rho_k^* \times \rho_k^*$ and $\rho_m = 1 - \sum_{k=1}^{m-1}\rho_k^* \times \rho_k^*$ are used instead of ρ_k in the estimation. Since $\xi_k \geq 0$ hold, $\xi_k = \xi_k^* \times \xi_k^*$ is used instead of ξ_k in the Laird-type MP model.

As a result, the following facts are given.

1) The maximum likelihood gradually increases with the increase in the number of MP and converges to a certain value. The converging number of MP is 4 for Larid-type and 2 for Lindsay-type. It must be noted that the maximum likelihood of the models with more than the converging number of MP is approximate because of the limitation of the software used here. These results mean that the population can be classified into 4 or 2 homogeneous groups of mode choice. ρ_k can be interpreted as the probability that an individual belongs to group k.

2) The estimated parameters of explanatory variables in the MP models are larger than those in the pooled logit model (see Table 2) except for the number of transfers. Although the cost parameter is not significant in the pooled model, it turns out to be significant in the MP models. This result reveals that mode choice models, in which unobserved heterogeneity is not considered, will produce erroneous conclusions.

3) The MP models are superior to the pooled logit model in terms of goodness-of-fit (i.e. adjusted likelihood ratio).

4) The estimation results from the two types of models are not significantly different from each other.

Table 3. Estimation results of binary Laird-type Mass Point model

Explanatory variable	MP = 1	MP = 2	MP = 3	MP = 4	MP = 5	MP = 6
Access time (min.)	0.027 (0.67)	-0.038 (0.73)	-0.067 (0.31)	0.130 (1.88)	0.173 (3.90)**	0.150 (1.71)
In-vehicle time (min.)	-0.029 (1.91)	-0.059 (2.82)**	-0.086 (1.12)	-0.054 (1.39)	-0.059 (3.19)**	-0.062 (2.26)*
Cost (100 yen)	-0.089 (1.78)	-0.183 (3.31)**	-0.170 (1.97)*	-0.135 (2.12)*	-0.145 (2.47)*	-0.164 (2.36)*
Egress time (min.)	-0.075 (3.26)**	-0.148 (3.62)**	-0.122 (2.33)*	-0.139 (2.82)**	-0.136 (3.24)**	-0.148 (2.59)**
No. of transfers	-0.274 (0.73)	0.122 (0.12)	-0.198 (0.20)	-0.072 (0.07)	-0.075 (0.09)	0.341 (0.37)
ρ^*_1		-0.579 (4.76)**	0.724 (4.30)**	0.090 (0.09)	0.071 (0.19)	-0.011 (0.01)
ξ^*_1	-0.582** (6.78)	-0.263 (1.23)	0.881 (14.0)**	0.224 (0.22)	0.448 (0.45)	-0.325 (0.26)
ρ^*_2			0.626 (0.64)	0.204 (0.38)	0.321 (1.32)	0.441 (0.51)
ξ^*_2		0.784 (8.56)**	-0.134 (0.96)	-0.162 (0.16)	0.062 (0.32)	0.092 (0.69)
ρ^*_3				0.542 (9.11)**	0.733 (11.8)**	-0.002 (0.01)
ξ^*_3			-0.509 (1.46)	0.988 (35.0)**	0.065 (1.35)	0.667 (1.88)
ρ^*_4					0.291 (1.35)	0.684 (0.98)
ξ^*_4				0.111 (1.04)	0.449 (1.04)	0.069 (0.92)
ρ^*_5						0.553 (12.2)**
ξ^*_5					0.985 (48.8)**	0.961 (54.6)**
ρ^*_6						
ξ^*_6						-0.078 (1.36)
Initial likelihood	-205.90	-205.90	-205.90	-205.90	-205.90	-205.90
Maximum likelihood	-71.87	-68.78	-56.14	-53.55	-53.95	-53.58
Adjusted likelihood ratio	0.645	0.660	0.723	0.735	0.733	0.735
Sample size	297	297	297	297	297	297

(t scores in parentheses; *: significant at 5%; **: at 1%)

Table 4. Estimation results of Lindsay-type Mass Point model

Explanatory variable	MP = 1	MP = 2	MP = 3	MP = 4	MP = 5
Access time (min.)	0.028 (0.68)	-0.038 (0.57)	-0.018 (0.24)	-0.012 (0.12)	0.084 (1.28)
In-vehicle time (min.)	-0.039 (2.27)*	-0.055 (2.08)*	-0.058 (1.76)	-0.054 (1.27)	-0.041 (2.16)*
Cost (100 yen)	-0.064 (1.28)	-0.100 (1.94)	-0.103 (2.15)*	-0.111 (2.14)*	-0.112 (2.35)*
Egress time (min.)	-0.086 (3.29)**	-0.116 (2.94)**	-0.113 (2.62)**	-0.120 (2.86)**	-0.124 (3.31)**
No. of transfers	-0.178 (1.10)	-0.065 (0.07)	-0.101 (0.10)	-0.184 (0.19)	-0.686 (0.96)
ρ^*_1		0.818 (17.5)**	0.003 (0.02)	-0.003 (0.003)	0.192 (1.14)
ξ^*_1	0.481 (2.81)**	2.179 (2.95)**	0.021 (0.02)	0.024 (0.02)	-1.705 (1.66)
ρ^*_2			0.830 (20.2)**	0.823 (16.3)**	0.525 (7.43)**
ξ^*_2		-2.807 (4.50)**	2.514 (3.39)**	2.353 (2.96)**	-5.633 (4.01)**
ρ^*_3				-0.055 (0.10)	-0.072 (0.33)
ξ^*_3			-2.796 (4.32)**	0.061 (0.06)	-0.183 (0.34)
ρ^*_4					0.104 (0.55)
ξ^*_4				-3.288 (5.45)**	-0.952 (0.95)
ρ^*_5					
ξ^*_5					2.929 (3.91)**
Initial likelihood	-205.90	-205.90	-205.90	-205.90	-205.90
Maximum likelihood	-71.43	-54.31	-53.99	-53.53	-53.69
Adjusted likelihood ratio	0.647	0.732	0.733	0.736	0.735
Sample size	297	297	297	297	297

(t scores in parentheses; *: significant at 5%; **: at 1%)

5. DYNAMIC MASS POINT MODEL TREATING UNOBSERVED HETEROGENEITY

5.1 Specification of Multinomial Mass Point Model

As mentioned in section 4, the difference between estimation results from the two types of models is not significant. Furthermore, since the Lindsay-type approach has a simple structure, this approach is easier to extend to multinomial logit model. Consider the following multinomial logit model incorporating unobserved heterogeneity, in which choice set differs across individuals and J_{it} is the choice set of individual i at time t.

$$\text{Prob}\left(y_{ijt} = 1 \,|\, \boldsymbol{\delta}_{ij}, \boldsymbol{\beta}, \{ \mathbf{x}_{ijt} \} \right) = \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{ijt} + \delta_{ij})}{\sum_{j'=1}^{J_{it}} \exp(\boldsymbol{\beta}' \mathbf{x}_{ij't} + \delta_{ij'})} \tag{19}$$

Probability that individual i chooses $S_i = \{ y_{ijt} \,|\, y_{ijt} = (0, 1); j = 1,\ldots,J_{it}; t = 1,\ldots,T_i \}$ can be represented as eqn (20), which is referred to as multinomial MP model.

$$\mathrm{Prob}\left(S_i \mid \boldsymbol{\xi}, \boldsymbol{\rho}, \boldsymbol{\beta}, \{\mathbf{x}_{ijt}\}\right) = \sum_{k=1}^{m} \left\{ \prod_{t=1}^{T_i} \frac{\prod_{j=1}^{J_{it}} \left[\exp(\boldsymbol{\beta}'\mathbf{x}_{ijt} + \xi_{kj})\right]^{y_{ijt}}}{\sum_{j'=1}^{J_{it}} \exp(\boldsymbol{\beta}'\mathbf{x}_{ij't} + \xi_{kj'})} \right\} \rho_k \tag{20}$$

where, y_{ijt} is a variable of choice result for mode j at time t.

The original multinomial Ajina panel data is used to estimate the pooled multinomial logit (MNL) model and multinomial MP model. The estimation results are shown in the 1st and 2nd columns of Table 5. ξ_{k2}, ξ_{k3} and ξ_{k4} are position parameters of MP k for bus, tram and rail, respectively. For the MP models, only the estimation results at the time of convergence is shown because of space limitations. The main conclusions from Table 5 are similar to those from Tables 3 and 4.

1) The converging number of MP in the MNL MP model is estimated to be 5 and most of the estimated parameters of ξ_{kj} and ρ_k are statistically significant. The adjusted likelihood ratio of the multinomial MP model is 0.440, higher than that of the pooled model (i.e. 0.072), which indicates the significance of the MP model.

2) The estimated parameter of in-vehicle time which is insignificant in the pooled model turns out to be significant in the multinomial MP model and that of cost turns out to have the expected sign, even though that of access time is insignificant. This means that the in-vehicle time which is believed to be more important than the access time in mode choice, is properly evaluated in the presence of heterogeneity, so that considering heterogeneity can make the biases in the estimated parameters smaller.

5.2 Specification of Dynamic Mass Point Model

Since the time-serial independence of travel behavior was assumed in previous sections, joint choice probability was calculated as the products of choice probability at each time-point. However, this assumption seems unrealistic because individual behavior interacts temporally. Here, dynamic mode choice models dealing with previous behavior are specified:

$$P(S_i) = \prod_{t=2}^{T_i-1} \prod_{j=1}^{J_{it}} \{ \mathrm{Prob}[\, y_{ijt} \mid (Q_{ij1}, \dots, Q_{ijt-1})] \}^{y_{ijt}} \tag{21}$$

where, $Q_{ij1}, \dots, Q_{ijt-1}$ are information related to the previous behavior.

Eqn (21) assumes that an individual's travel behavior at time t is influenced by the entire history of his/her behavior. Because there are only three time points in the Ajina panel data, the following dynamic model is adopted here.

$$P(S_i) = \prod_{t=2}^{T_i-1} \prod_{j=1}^{J_{it}} \{ \mathrm{Prob}[\, y_{ijt} \mid Q_{ijt-1}] \}^{y_{ijt}} \tag{22}$$

where,

$$\mathrm{Prob}[\, y_{ij1} \mid Q_{ij0}] = \mathrm{Prob}(y_{ij1}) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ij1} + \delta_{ij})}{\sum_{j'=1}^{J_{it}} \exp(\boldsymbol{\beta}'\mathbf{x}_{ij'1} + \delta_{ij'})} \tag{23}$$

Table 5. Estimation results of Pooled MNL, Multinomial MP and Dynamic MP models

Explanatory variable	Pooled MNL Model	MNL MP Model (MP=5)	Dynamic MP Model 1 [1] (MP = 3)	Dynamic MP Model 2 [2] (MP = 4)	Dynamic MP Model 3 [3] (MP = 3)
Access time (min.)	-0.030*	-0.055	-0.014	-0.073	-0.057
In-vehicle time (min.)	-0.002	-0.036**	-0.029**	-0.079**	-0.041**
Cost (100 yen)	0.001	-0.039	-0.158**	-0.079	-0.037
Egress time (min.)	-0.040**	-0.098**	-0.082**	-0.123**	-0.095**
Effect of state dependence			3.199**	0.169**	2.304**
ρ^*_1		0.510**	0.377**	0.708**	0.435**
ξ^*_{12}		1.675*	6.157	-1.072	12.07**
ξ^*_{13}		4.995**	-5.595	-3.340**	-2.522
ξ^*_{14}		-7.214**	2.904	-9.216**	9.069**
ρ^*_2		0.115*	0.171*	-0.132*	0.756**
ξ^*_{22}		-1.650	0.327	-1.738	-0.679
ξ^*_{23}		5.941	1.796	3.106**	-1.806**
ξ^*_{24}		12.75	7.944**	9.001**	-9.565**
ρ^*_3		0.430**		0.420**	
ξ^*_{32}		17.82**	-0.192	11.12**	1.818
ξ^*_{33}		-22.41**	-0.502	-4.481**	5.731
ξ^*_{34}		10.90**	-9.167*	5.916**	-1.106
ρ^*_4		0.101			
ξ^*_{42}		11.82**		-6.523**	
ξ^*_{43}		-5.928**		0.707	
ξ^*_{44}		9.137**		3.239**	
ρ^*_5					
ξ^*_{52}		-1.362*			
ξ^*_{53}		-3.160**			
ξ^*_{54}		-15.89**			
Initial likelihood	-359.29	-359.29	-359.29	-359.29	-359.29
Maximum likelihood	-331.08	-194.00	-163.00	-169.00	-188.61
Adjusted likelihood ratio	0.072	0.440	0.535	0.515	0.461
Sample size	507	507	507	507	507

([1] with previous choice result; [2] choice utility; [3] choice dummy variable; *: significant at 5%; **: at 1%)

$$\mathrm{Prob}\left[\, y_{ijt} \mid \Omega_{ijt-1} \,\right] = \frac{\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ijt} + \gamma\,\Omega_{ijt-1} + \delta_{ij}\right)}{\sum_{j'=1}^{J_{it}} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij't} + \gamma\,\Omega_{ij't-1} + \delta_{ij'}\right)} \tag{24}$$

Concerning Ω_{ijt-1}, three kinds of previous information are considered here:

1) $\Omega_{ijt-1} = y_{ijt-1}$: observed previous choice behavior,

It is plausible to consider the endogeneity of y_{ijt-1} when y_{ijt-1} is treated as an explanatory variable in eqn (24). Here, we assume that the endogeneity of y_{ijt-1} does not affect the estimation for the sake of simplicity.

2) $\Omega_{ijt-1} = V_{ijt-1}$ and $V_{ijt-1} = \boldsymbol{\beta}'\mathbf{x}_{ijt-1} + \xi_{kj}$: previous choice utility, and

3) $\Omega_{ijt-1} = d_{ijt-1} = \begin{cases} 1\text{, if } \widetilde{P}_{ijt-1} \geq 1/J_{it-1} \\ 0\text{, otherwise} \end{cases}$: previous choice dummy variable

where, $\widetilde{P}_{ijt-1}$ is the estimated probability that individual i chooses mode j at time t-1 and it is obtained from a logit model like eqn (23) without δ_{ij}.

Then, dynamic MP model of mode choice can be specified:

$$P(S_i) = \sum_{k=1}^{m} \left(\frac{\prod_{j=1}^{J_{it}} \left[\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij1}+\xi_{kj}\right)\right]^{y_{ij1}}}{\sum_{j'=1}^{J_{it}} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij'1}+\xi_{kj'}\right)} \cdot \prod_{t=2}^{T_i} \frac{\prod_{j=1}^{J_{it}} \left[\exp\left(\boldsymbol{\beta}'\mathbf{x}_{ijt}+\gamma\,\Omega_{ijt-1}+\xi_{kj}\right)\right]^{y_{ijt}}}{\sum_{j'=1}^{J_{it}} \exp\left(\boldsymbol{\beta}'\mathbf{x}_{ij't}+\gamma\,\Omega_{ij't-1}+\xi_{kj'}\right)} \right) \rho_k \quad (25)$$

5.3 Empirical Analysis

The estimation results of dynamic MP models including previous choice behavior (i.e. choice result), choice utility and choice dummy variable are shown in the last three columns of Table 5. Moreover, the previous dummy variable is defined based on $\widetilde{P}_{ijt-1}$ in section 5.2 which is obtained by estimating the multinomial logit model without state dependence and unobserved heterogeneity. Here, it is assumed that the observed heterogeneity due to individual socio-economic attributes, such as sex and age, does not exist. The last three columns of Table 5 indicate that:

1) The effects of state dependence obtained from three types of dynamic MP models are all statistically significant.

2) The converging numbers of MP from the dynamic MP models with previous choice result, choice utility and choice dummy variable are 3, 4 and 3, respectively, which are smaller than that from the MNL MP model. The ranges of the estimated ξ_k for four models from the 2nd to the last column are {-22.41, 17.82} (40.230 in width), {-9.17, 7.94} (17.111 in width), {-9.22, 11.12} (20.34 in width) and {-9.57, 12.07} (21.64 in width), respectively. The ranges of estimated ξ_k for the dynamic MP models are smaller than that for the MNL MP model, and the dynamic MP model with previous choice result has the smallest range. This means that one aspect of the effects of omitted variables, which is the unobserved heterogeneity in the MNL MP model, may be substituted by the effects of state dependence incorporated in the dynamic MP models.

3) The goodness-of-fit indices (i.e. adjusted likelihood ratios) of three dynamic MP models are superior to that of the MNL MP model. The model with previous choice result is the best of all three dynamic models. Since the dynamic MP model with previous choice result has a simple structure and can be easily calibrated, it is practical to employ the previous choice result as a state dependence variable when short-term panel data is used.

On the other hand, it is also important to discuss the temporal transferability of the proposed dynamic MP models. Consequently, the following four models are newly estimated using the first two-period Ajina panel data (Jun. 1989 and Nov. 1989); Model-A: static pooled logit model with constant terms, Model-B: dynamic logit model with constant terms, Model-C: static MP model and Model-D: dynamic MP model with previous choice result.

Unlike the first two models, model-C and D deal with unobserved heterogeneity. The estimation results of models-A~D are shown in Table 6. Then the estimated parameters of explanatory variables are applied to the Ajina panel data at the third time-point (Oct. 1991) to calculate the share of each mode.

Table 6. Estimation results of model-A~D

Explanatory variable	Model-A	Model-B	Model-C (MP = 4)	Model-D (MP = 3)
Access time (min.)	-0.035	-0.064*	-0.119	-0.061
In-vehicle time (min.)	-0.043**	-0.047**	-0.168**	-0.064**
Cost (100 yen)	-0.002**	-0.002**	-0.011**	-0.006**
Egress time (min.)	-0.026	-0.030*	-0.170**	-0.065*
Effect of state dependence		3.159**		3.631**
ρ^*_1			-0.439**	-0.431**
ξ^*_{12}	-0.902**	-0.662	23.04**	-2.384
ξ^*_{13}	-1.036**	-0.831*	-18.24**	-12.55**
ξ^*_{14}	-2.489**	-1.960**	19.55**	-1.383
ρ^*_2			-0.432**	0.355**
ξ^*_{22}			-2.473	10.37**
ξ^*_{23}			4.545	5.420**
ξ^*_{24}			-16.40**	6.429**
ρ^*_3			-0.220**	
ξ^*_{32}			16.35**	-2.147**
ξ^*_{33}			10.24**	-1.656
ξ^*_{34}			4.315	-21.02**
ρ^*_4				
ξ^*_{42}			-7.429**	
ξ^*_{43}			-5.571**	
ξ^*_{44}			-42.41**	
Initial likelihood	-231.93	-231.93	-231.93	-231.93
Maximum likelihood	-170.06	-111.13	-118.81	-103.49
Adjusted likelihood ratio	0.253	0.511	0.461	0.535
Sample size	338	338	338	338

(*: significant at 5%; **: at 1%)

The temporal transferability of these models is analyzed by comparing the calculated share of each mode with the observed share based on the following absolute error, which indicates that a smaller value of the AE means higher temporal transferability of the model.

$$\text{AE (absolute error) of the share} = \sum_{k=1}^{K} \left| S_k^F - S_k^{P\text{-}F} \right| \tag{26}$$

where, S_k^F : observed share of mode k at time-point F (i.e. Oct. 1991) and

$S_k^{P\text{-}F}$: estimated share of mode k at time-point F using parameters at time-point P (i.e. June and Nov. 1989).

The rank of the models in terms of transferability in time is model-D, C, B and A according to the AE values in Table 7. Concerning the improved degree of accuracy, the MP models (i.e. model-C and D) are better than the models without unobserved heterogeneity (i.e. model-A and B), while dynamic models (i.e. model-B and D) is better than the models without state dependence (i.e. model-A and C). Besides, the MP model-C has a higher level of temporal transferability than the dynamic model-B. This result means that unobserved heterogeneity has a stronger effect on the estimation result of model than state dependence.

Table 7. Comparison of estimated shares for 1991 using 1989 models

Mode	Observed share	Estimated share			
		Model-A	Model-B	Model-C	Model-D
Car	45.6	37.6	42.2	43.5	40.9
Bus	25.4	33.7	31.1	22.9	27.1
Tram	19.5	21.5	19.4	24.5	20.5
Rail	9.5	7.3	7.1	9.3	11.6
AE	0.0	20.4	11.4	9.4	9.3

6. CONCLUSIONS

It is very significant to consider unobserved heterogeneity in developing discrete choice models. Therefore, binary and multinomial MP model and dynamic MP models have been developed in this paper to consider unobserved heterogeneity using short-term panel data in Hiroshima. As a result, some important conclusions can be stated:

1) The estimated models using Laird- and Lindsay-type MP approaches are not significantly different from each other. Because of simple model structure, the Lindsay-type approach is useful to build dynamic travel models.

2) The estimation results show that the maximum likelihood of models is improved greatly and finally converges to a certain value as the number of MP increases.

3) The estimated parameters of explanatory variables by MP models differ from those of models without unobserved heterogeneity. This result means that conventional models excluding unobserved heterogeneity, may derive erroneous conclusions.

4) MP models are superior to that of models without unobserved heterogeneity, while dynamic MP models are superior to static MP models without state dependence with respect to the goodness-of-fit of models.

5) In developing dynamic MP models with short-term panel data, the most useful and practical way is to incorporate the previous choice results into the model.

REFERENCES

Andersen E.B. (1970) Asymptotic properties of conditional maximum likelihood estimators, **Journal of the Royal Statistical Society, Series B32**, 283-301.

Burnett P. (1974) Disaggregate behavioral models of travel decisions other than mode choice: a review and contribution to spatial choice theory, **Transportation Research Board Special Report 149**, 207-222.

Chamberlain G. (1980) Analysis of covariance with qualitative data, **Review of Economic Studies XLVII**, 225-238.

Daganzo C.F. and Sheffi Y. (1982) Multinomial probit with time-series data: unifying state dependence and serial correlation models, **Environment and Planning A 14**, 1377-1388.

Davies R.B. (1984) A generalised beta-logistic model for longitudinal data with an application to residential mobility, **Environment and Planning A 16**, 1375-1386.

Davies R.B. and Crouchley R. (1984) Calibrating longitudinal models of residential mobility and migration: an assessment of a nonparametric marginal likelihood approach, **Regional Science and Urban Economics 14**, 231-247.

Davies R.B. and Crouchley R. (1985) Control for omitted variables in the analysis of panel and other longitudinal data, **Geographical Analysis 17**, 1-15.

Davies R.B., Crouchley R. and Pickles A.R. (1982) A family of hypothesis tests for a collection of short event series with an application to female employment participation, **Environment and Planning A 14**, 603-614.

Davies R.B. and Pickles A.R. (1987) A joint trip timing store-type choice model for grocery shopping, including inventory effects and nonparametric control for omitted variables, **Transportation Research 21A**, 345-361.

Davies R.B., Pickles A.R. and Crouchley R. (1983) Some methods for the testing and estimation of dynamic models which use panel data, **Environment and Planning A 15**, 1475-1488.

Dunn R., Reader S. and Wrigley N. (1987) A nonparametric approach to the incorporation of heterogeneity into repeated polytomous choice models of urban shopping behaviour, **Transportation Research 21A**, 327-343.

Dunn R. and Wrigley N. (1985) Beta-logistic models of urban shopping center choice, **Geographical Analysis 17**, 95-113.

Fischer M.M. and Nijkamp P. (1987) From static towards dynamic discrete choice modelling: a state of the art review, **Regional Science and Urban Economics 17**, 3-27.

Heckman J.J. (1981) Statistical models for discrete panel data, in C.F. Manski and D. McFadden (eds.), **Structural Analysis of Discrete Data with Econometric Applications**, MIT Press, Cambridge.

Heckman J. and Singer B. (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data, **Econometrica 52**, 271-320.

Heckman J.J. and Willis R.J. (1977) A beta-logistic model for the analysis of sequential labor force participation by married women, **Journal of Political Economy 85**, 27-58.

Hensher D.A. and Mannering F.L. (1994) Hazard-based duration models and their application to transport analysis, **Transport Review 14**, 63-82.

Kitamura R. (1990) Panel analysis in transportation planning: an overview, **Transportation Research 24A**, 401-415.

Kitamura R. and Bunch D. S. (1990) Heterogeneity and state dependence in household car ownership: a panel analysis using ordered-response probit models with error components, In M. Koshi (ed.), **Transportation and Traffic Theory**, Elsevier.

Kitamura R., Yamamoto T., Fujii S. and Sampath S. (1996) A discrete-continuous analysis of time allocation to two types of discretionary activities which accounts for unobserved heterogeneity, In J. B. Lesort (ed.), **Proceedings of the 13th International Symposium on Transportation and Traffic Theory**, Elsevier Science.

Laird N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution, **Journal of American Statistical Association 73**, 805-811.

Lindsay B.G. (1981) Properties of the maximum likelihood estimator of a mixing distribution, In C. Tallie (ed.), **Statistical Distributions in Scientific Work**, Kluwer Academics Publishers.

Lindsay B.G. (1983a) The geometry of mixture likelihoods, Part I, **Annals of Statistics 11**, 86-94.

Lindsay B.G. (1983b) The geometry of mixture likelihoods, Part II, **Annals of Statistics 11**, 783-792.

Neyman J. and Scott E.L. (1948) Consistent estimates based on partially consistent observations, **Econometrika 16**, 1-32.

Pickles A.R. (1983) The analysis of residence histories and other longitudinal panel data, **Regional Science and Urban Economics 13**, 271-285.

Reader S. (1993) Unobserved heterogeneity in dynamic discrete choice models, **Environment and Planning A 25**, 495-519.

Simar L. (1976) Maximum likelihood estimation of a compound Poisson process, **Annals of Statistics 4**, 1200-1209.

Tardiff T.J. (1979) Definition of alternatives and representation of dynamic behaviour in spatial choice models, **Transportation Research Record 723**, 25-30.

Uncle M.D. (1987) A beta-logistic model of mode choice: goodness of fit and inter-temporal dependence, **Transportation Research 21B**, 195-205.