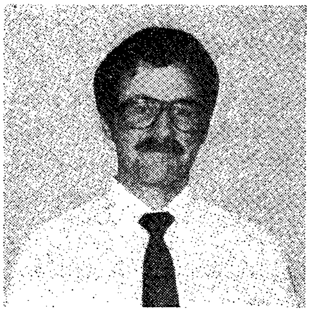
[PROC. OF JSCE No. 347/IV-1 (Infrastructure Planning and Management) July 1984]

Invited Paper

DISAGGREGATE TRIP DISTRIBUTION MODELS

By Moshe Ben-Akiba, Hugh F. Gunn** and Lionel A. Silman***



1. INTRODUCTION

The principal method used in transportation studies to forecast matrices of person trips between origins and destinations is usually an aggregate gravity model. This paper presents the theory and methods for the application of disaggregate estimation of probabilistic choice models to trip distribution. The most important practical differences between the aggregate and the disaggregate models are in the model estimation stage. Disaggregate models can be estimated with smaller traveller surveys and place an emphasise on the specification of utility functions with a larger set of explanatory variables.

Disaggregate models were initially employed during the 1960's to study the choice of travel mode. The first application of disaggregate estimation techniques and the multinomial logit model to the choice of trip destination was by CRA (1972). (See also McFadden, 1974, and Domencich and McFadden, 1975.) This model was based on an unsatisfactory specification of the level of service variables by different modes of travel. It was corrected by Ben-Akiva (1973) who estimated nested and joint logit models of mode and destination choice.

These prototypical models led to several developments and applications of complete systems of disaggregate urban travel demand models. In a research context, Ben-Akiva *et al.* (1977) developed a system of disaggregate models and used it for aggregate forecasting with network equilibration. Subsequently, operational disaggregate model systems were implemented by Ruiter and Ben-Akiva (1978) in the U. S., and by Ben-Akiva *et al.* (1978) in the Netherlands. An overview of these earlier studies is given in Ben-Akiva (1977).

* Associate Professor, Transportation Systems Division, Department of Csvil Engıneering, Massachwsetts Institute of Technology

** Cambridge Systematics Europe, The Hague, The Netherlands

The destination choice models presented in this paper incorporate several innovations in disaggregate modelling techniques since these earlier applications. The paper is divided into three main parts. Section 2 deals with the issue of aggregation of actual destinations into traffic zones, also addressed by aggregate trip distribution models. However, the section presents an aggregation of alternatives theory that was developed for disaggregate logit models based on the concept of random utility.

The major advances in the state-of-the-art, following those of the earlier models, were in the application of more efficient statistical estimation methods. These developments are presented in Section 3. Of direct relevance to destination choice is the development by McFadden (1978) of a consistent estimator for the logit model with sampling of alternatives. An important result which will be applied more in future studies is the maximum likelihood estimator for generalized choice-based survey designs developed by Cosslett (1981).

Finally, Section 4 presents destination choice models that were estimated in two recent studies in Paris, France, and in Maceio, Brazil, using the methods described in Sections 2 and 3.

2. The Alternatives in Destination Choice

Destination choice is characterized by a very large number of alternatives. In some situations, such as shopping trips, it may be possible to identify a small set of distinct shopping centers (see, for example, Richards and Ben-Akiva, 1975, Koppelman and Hauser, 1978, and Morichi *et al*, 1984). However, for other trip purposes and in dense urbanized areas there are no natural boundaries that define alternative destinations.

Furthermore, data on the attractivity of alternative destinations are only available at some level of geographical aggregation, which is usually a system of traffic zones. Thus, in most applications the alternatives in a destination choice model must be based on aggregate alternatives. The relationship between the underlying elemental destinations and aggregated destination alternatives is developed below. A theory of aggregation of alternatives is presented and its implications for empirical models are shown.

(1) The Concept of Elemental Alternatives

Define the actual destinations that travellers are choosing as elemental alternatives. An example of an elemental alternative would be an individual store as the destination of a shopping trip. Any definition of alternative destinations for modelling can therefore be viewed as a scheme of grouping together elemental alternatives.

In some situations it may not be possible to define precisely what constitutes an elemental alternative. In a choice of a residential location an elemental alternative is a dwelling unit. However, for a recreational trip it may not be possible to identify a unique establishment, or a well defined cluster of establishments, that consititutes an elemental destination alternative. In these situations it will be assumed that a destination alternative is an aggregation of some elemental alternatives whose precise definition is unknown to the analyst and, as a consequence, the number of elemental destinations included in an aggregate destination alternative is also undefined.

The elemental destination alternatives are, by definition, mutually exclusive and exhaustive; that is, the traveller chooses one and only one elemental alternative. Denote by L the set of all elemental destinations and by $p_n(l)$ the nth traveller's choice probability for elemental destination $l \in L$. Partition the set L into non-overlapping subsets L_i, $i=1,\cdots,J$, where J is the number of aggregate destination alternatives.

The choice probability of an aggregate destination is equal to the probability that the traveller chooses one of its elemental alternatives. Thus the choice probability for traveller n of destination i is defined by

$$p_n(i)=\sum_{l \in L_i} p_n(l), \quad i=1,\cdots,J \qquad (1)$$

(2) Random Utilities of Aggregate Alternatives

Denote the random utility of an elemental destination l to traveller n by U_{ln} and assume, without loss of generality, that

$$U_{ln}=V_{ln}+\varepsilon_{ln} \quad (2)$$

where V_{ln} and ε_{ln} are, respectively, the systematic and the random components of the utility for elemental alternative l and traveller n. Since the elemental alternatives are mutually exclusive, the utility of an aggregate destination alternative can be defined by

$$U_{in}=\max_{l\in L_i}(V_{ln}+\varepsilon_{ln}),\quad i=1,\cdots,J \quad (3)$$

The utility of an aggregate destination can also be expressed as a sum of its expectation, denoted by V_{in}, and a random component, denoted by ε_{in}, as follows

$$U_{in}=V_{in}+\varepsilon_{in},\quad i=1,\cdots,J \quad (4)$$

where

$$V_{in}=E\left[\max_{l\in L_i}(V_{ln}+\varepsilon_{ln})\right]$$

Define the average of the elemental alternative utilities by

$$\bar{V}_{in}=\frac{1}{M_i}\sum_{l\in L_i}V_{ln},\quad i=1,\cdots,J \quad (5)$$

where M_i is the number of elemental alternatives in the set L_i. The relationship between the mean utility of the aggregate destination (V_{in}) and the average utility of its elemental alternatives ($\bar{V}_{in}$) depends on the joint distribution function of the elemental alternatives utilities.

(3) Extreme Value Distribution of Random Utilities

If aggregate destination i contains a large number of elemental alternatives, and if the utilities of the elemental alternatives are independently and identically distributed (*i. i. d.*), then the distribution of the utility of destination i approaches the following extreme value (or Gumbel) distribution (see Johnson and Kotz, 1970):

$$F(U_{in})=\exp\{-\exp[-\mu'(U_{in}-\alpha)]\} \quad (6)$$

where

$F(\ )$ denotes a cumulative distribution function; μ' is a positive scale parameter; and α is the mode of the distribution. (The mean is equal to $\alpha+\frac{\gamma}{\mu'}$, where γ is Euler's constant, approximately equal to 0.577.)

The *i. i. d.* assumption above means that

$$V_{ln}=\bar{V}_{in},\quad V\ l\in L_i$$

i. e., the utilities of the elemental alternatives in a destination have equal means and that the ε_{ln}, for all $l\in L_i$, are *i. i. d.*. It can be shown that

$$\alpha=\bar{V}_{in}+\frac{l}{\mu'}lnM_i \quad (7)$$

where the additional term $\frac{l}{\mu'}lnM_i$ is a measure of the "size" of aggregate destination i.

The extreme value distribution is maintained under maximization. Hence, if the elemental utilities have unequal means but the random components of the elemental utilities are *i. i. d.* extreme value, as follows,

$$F(\varepsilon_{ln})=\exp[-\exp(-\mu'\varepsilon_{ln})],\quad l\in L_i \quad (8)$$

then the distribution of the aggregate utility of destination i is also extreme value with the same scale parameter μ' and

$$\alpha=\frac{l}{\mu'}l_n\sum_{l\in L_i}\exp[\mu' V_{ln}]$$

$$=\bar{V}_{in}+\frac{l}{\mu'}l_n\left[\frac{l}{M_i}\sum_{l\in L_i}\exp[\mu'(V_{ln}-\bar{V}_{in})]+\frac{l}{\mu'}l_nM_i\right. \quad (9)$$

Thus, the utility for a "heterogenous" destination alternative i can be modelled by

$$U_{in}=\bar{V}_{in}+\frac{1}{\mu'}lnM_i+\frac{1}{\mu'}lnB_{in}+\varepsilon_{in} \qquad (10)$$

where

$$B_{in}=\frac{1}{M_i}\sum_{l\in L_i}\exp[\mu'(V_{ln}-\bar{V}_{in})]$$

is a measure of the heterogeneity of elemental alternatives in aggregate destination i.

The convexity of the exponential function, the monotonicity of the logarithmic transformation and Jensen's inequality (Rao, 1973) imply that

$$\frac{1}{\mu'}ln\frac{1}{M_i}\sum_{l\in L_i}\exp[\mu' V_{ln}]\geq \bar{V}_{in} \qquad (11)$$

or alternatively, that

$B_{in}\geq 1$

The correction for heterogeneity is therefore non-negative. The equality holds in the case of homogeneous elemental alternatives in an aggregate destination.

The properties of the measure for the variability of the attributes among elemental alternatives were analyzed by Lerman (1975), McFadden (1978), and Kitamura *et al.* (1979). A useful insight can be gained by assuming that the number of elemental alternatives in an aggregate destination is large and the distribution of the V_{ln}, for $l\in L_i$, around $\bar{V}_{in}$, is approaching a normal distribution with variance σ^2_{in}. Under this assumption it can be shown that the term $\frac{1}{\mu'}lnB_{in}$ approaches $\frac{\mu'}{2}\sigma^2_{in}$. The same result can also be obtained by a second-order Taylor series expansion about $V_{ln}=\bar{V}_{in}, l\in L_i$.

Thus, the utility term for the variability of attributes among elemental alternatives may be omitted if the aggregate destination alternatives are defined with equal within zone variances such that

$\sigma^2_{in}=\sigma^2_n,\quad i=1,\cdots,J$

Lerman (1975) and Kitamura *et al.* (1979) have shown that in situations of unequal destination zones the size and the variance effects play an important role in correcting for the aggregation of elemental destinations.

(4) A Logit Model of Aggregate Destination Alternatives

It was shown above that the average utility of aggregate destination i, denoted as V_{in}, can be expressed as the sum of

$\bar{V}_{in}$: the average utility of the elemental alternatives in destination i;

$\frac{1}{\mu'}lnM_i$: the measure of the size of destination i; and

$\frac{1}{\mu'}lnB_{in}$: the measure of the variability of the attractiveness of the elemental alternatives in destination i.

This aggregate destination utility can be used in a logit destination choice model as follows:

$$p_n(i)=\frac{\exp[\mu^* V_{in}]}{\sum_{j=1}^{J}\exp[\mu^* V_{jn}]},\quad i=1,\cdots,J \qquad (12)$$

where μ^* is a positive scale parameter. This model assumes that $\varepsilon_{in}, i=1,\cdots,J$, are *i. i. d.* Gumbel. Because the V_{in} 's were derived from a logit model of the choice among elemental alternatives, (12) has the nested logit form that can be derived directly as a special case of the multivariate extreme value distribution of McFadden(1978). The ratio of scale parameters of the ε_{in} and ε_{in} Gumbel distributions is

$$\mu=\frac{\mu^*}{\mu'}$$

The square of the scale parameter of a logit model is proportional to the inverse of the variance of the random utilities, and since the variance cannot decrease with aggregation, the following must hold:

$0\leq\mu<1$

The correlation between elemental utilities of different aggregate destinations is assumed to be zero. For elemental utilities within the same aggregate destination the correlation coefficient is equal to $(1-\mu^2)$ (see Ben-Akiva and Lerman, 1984). Ideally, if the average utilities are well specified it may be possible to impose the restriction $\mu=1$.

In this case the parameters of the destination choice model are not dependent on the definitions of the aggregate destination alternatives. This means that a model estimated for one set of aggregate destinations can also be applied for other aggregation schemes (given, of course, that the size and variance measures are adjusted appropriately).

(5) Modelling Destination Choice with Unknown Size and Variance

The average utility $\bar{V}_{in}$ is a function of the average attributes of the the destinations in zone i for traveller n. The variance terms lnB_{in} can also be treated as an attribute of a destination zone. Thus, with known destination sizes the logit model utility for a linear-in-the-parameters specification with K' coefficients of the attribute variables becomes

$$\mu^* V_{in}=\sum_{k=1}^{K'} \theta_K X_{ink}+\mu lnM_i, \quad i=1,\cdots,J \qquad (13)$$

where

θ_K is the coefficient of the kth attribute ; and

X_{ink} is the value of the kth attribute for destination i and traveller n.

This utility is linear in the unknown parameters ($\theta_k, k=1,\cdots,K'$, and μ) ; it can be estimated with a standard linear logit estimation procedure, which can also be used to estimate the model under the equality restriction $\mu=1$.

As discussed earlier, in empirical applications the "size" of a destination is often unobservable. However, it is possible to hypothesize a relationship between the size of an alternative and a set of observable "size variables", such as population, employment, and area. For consistency among different levels of aggregation, the relationship between multiple size variables and the destination size must be linear, as follows :

$$M_i=\sum_{k=K'+1}^{K} \theta_k X_{ink}, \quad i=1,\cdots,J \qquad (14)$$

where

$(K-K')$ is the number of size variables ; and

X_{ink}, $k=K'+1,\cdots,K$, are the values of the size variables for destination i and traveller n.

The size variables must be non-negative

$X_{ink}\geq 0, \quad k=K'+1,\cdots,K, \quad i=1,\cdots,J$

and their coefficients must also be non-negative

$\theta_k \geq 0, \quad k=K'+1,\cdots,K$

with at least one non-zero coefficient. Moreover, not all the size varible coefficients can be identified, and it is necessary to impose a restriction such as $\theta_k=1$. Obviously, this restriction can only be applied to a non-zero coefficient.

The case of a single size variable is equivalent to known size. The size variable simply replaces the size measure and there are no additional parameters to estimate. If size variables are not available, the model must include a full set of destination specific constants to capture the size effect.

The general case with two or more size variables yields a logit model which is non-linear in the parameters $\theta_k, k=K'+1,\cdots,K-1$. This model requires a special logit estimation procedure, such as the one developed by Daly (1982) and applied by CSE (1983).

3. Estimation of Disaggregate Destination Choice Models

Using heterogeneous aggregate destinations may introduce measurement errors of the explanatory variables, e.g., travel time, and thereby reduce the accuracy of a destination choice model. On the other hand, using a large number of small traffic zones as the destination alternatives may cause the data preparation and the computational burden to be prohibitively expensive. The solution for this latter problem is found by utilizing the independence from irrelevant alternatives property of the logit model which permits consistent estimation with only a subset of the alternatives (including the chosen and a sample of non-chosen alternatives).

Thus, estimation of a disaggregate destination choice model may include two types of sampling :
-sampling of observations; and
-sampling of destinations for every observation.

This section of the paper is concerned with alternative sampling methods and their corresponding model estimation procedures. The first subsection describes the methods of sampling destinations and the second subsection develops the use of choice-based and enriched samples of observations in destination choice models.

(1) Sampling of Destinations

Denote by N the number of observations in the estimation sample and by $\tilde{J}_n(\tilde{J}_n \leq J)$ the number of destination alternatives assigned to observation $n, n=1, \cdots, N$. Clearly, with a subset of the alternatives it would only be possible to maximize a conditional likelihood function rather than the true likelihood. An observation of a chosen alternative for traveller n is assumed to be a random draw from a multinomial distribution whose probabilities are given by a parametric choice model, $p_n(i), i=1,\cdots,J$. A sampling of alternatives procedure assigns to observation n a subset of the destination alternatives, denoted by D_n, with $\tilde{J}_n$ elements. It is obvious that for estimation purposes D_n must include the chosen alternative. Denote by $\pi_n(D \mid i)$ the conditional probability of constructing for observation n the set D given that the chosen alternative is i. (These are such that $\pi_n(D \mid j)=0$ for $j \notin D$) The joint probability of drawing a chosen alternative and a subset of alternatives D is

$$\pi_n(i,D)=\pi_n(D \mid i)\cdot p_n(i), \quad i=1,\cdots,J \cdots\cdots (15)$$

Thus, the conditional probability of alternative i being chosen given a sample of alternatives D is

$$\pi_n(i \mid D)=\frac{\pi_n(D \mid i)\, p_n(i)}{\sum_{j \in D} \pi_n(D \mid j) p_n(j)}, \quad i=1,\cdots,J \cdots\cdots (16)$$

The conditional probability $\pi_n(i \mid D)$ exists if

$\pi_n(D \mid j)>0, \quad \forall j \in D$

This positive conditioning property was established by McFadden (1978) as a condition for a consistent estimator for the logit model with samples of alternatives. Substitute the logit choice probabilities given in equation (12) in the expression for the conditional probability (16) to obtain

$$\pi_n(i \mid D)=\frac{\exp[\mu^* V_{in}+ln\, \pi_n(D \mid i)]}{\sum_{j \in D} \exp[\mu^* V_{jn}+ln \pi_n(D \mid j)]}, \quad i \in D \cdots\cdots (17)$$

McFadden (1978) uses (17) to prove that maximization of the conditional likelihood function

$$\sum_{n=1}^{N} ln\, \pi_n(i \mid D) \cdots\cdots (18)$$

yields, under normal regularity conditions, consistent estimates of the unknown parameters θ. Note that the estimated logit model (17) includes an additive alternative specific correction for the bias introduced by the sampling of alternatives. The coefficient of this variable is constrained to 1.

Sampling of alternatives is an easily applicable technique for reducing the computational burden involved in estimating a destination choice model with a large number of traffic zones. The main issue which is unresolved is how to obtain the most effective sample of destinations.

The computational load is approximately linear in both the number of observations and in the number of alternatives. If sufficient observations are available, it would seem that more information is obtained by using many observations with rather few alternatives per observation, rather than having few observations and a large sample of alternatives for each observation.

Two principal types of sampling strategies have been applied to destination sampling : random sampling with uniform selection probabilities, and sampling with unequal selection probabilities (designed to reflect differential levels of importance of alternative destinations).

1) Simple Random Sampling of Destinations

The simplest approach to sample design is to draw a simple random sample of alternatives and to add the chosen alternative if it is not otherwise included. Thus, if J' alternatives are randomly drawn without replacement from the population of J alternatives $(J' \leq J)$ then the probability that the outcome is the subset D is

$$\pi_n(D \mid i) = C(J, J')^{-1}, \quad i \in D \cdots\cdots (19)$$

where $C(J, J')$ denotes the number of combinations of J items taken J' at a time. The size of the destination sample for observation n, denoted by $\tilde{J}_n$, is equal to J' if the observed choice was sampled, otherwise it is equal to $J'+1$. This sample design was used in earlier studies, as in, for example, Netherlands Ministry of Transport (1977).

To ensure that the various destination samples have the same size, it is possible to randomly draw (without replacement) J' alternatives from all the available destinations, except for the chosen alternative. In this case the set D always has $J'+1$ elements and

$$\pi_n(D \mid i) = C(J-1, J')^{-1}, \; i \in D \cdots\cdots (20)$$

This method was used in the study in Tel-Aviv by Silman (1980).

The above simple random sampling strategies are characterized by the uniform conditioning property,

$$\pi_n(D \mid i) = \pi_n(D \mid j), \quad \forall i, \; j \in D$$

This property implies that the correction terms for destination sampling bias in the logit model,

$$ln\pi_n(D \mid i), \quad i \in D$$

cancel out, and a standard logit model with a choice set given by D yields consistent estimates.

2) Importance Sampling of Destinations

It would seem that a simple random sample is not necessarily an efficient scheme since for any given traveller the vast majority of the traffic zones may have very small choice probabilities. It would seem to be more efficient to design a sample of alternatives in which the high probability zones have a higher probability of being selected. The basic idea of importance sampling is borrowed from Monte Carlo integration, where it is optimal to use sample selection probabilities which are proportional to the quantity being observed (see Hammersley and Handscomb, 1965). For example, consider the problem of estimating a sum of choice probabilities over a subset with J_0 alternatives,

$$\sum_{i=1}^{J_0} p_n(i)$$

It is efficient to select a sample from the J_0 destination alternatives with selection probabilities q_{in} such that the ratios

$$\frac{p_n(i)}{q_{in}}, \quad i=1,\cdots,J_0$$

vary as little as possible.

Thus, an importance destination sampling strategy is based on preliminary estimates of the choice probabilities. Usually two factors are considered, distance and size, which may be combined by a "gravity" type function

$$\tilde{M}_i \exp[-\beta \, d_{in}], \quad i=1,\cdots,J \cdots\cdots (21)$$

where $\tilde{M}_i$ is an approximate measure of size of destination zone i ; d_{in} is a measure of distance between the origin of traveller n and destinations in zone i ; and β is a parameter that represents the sensitivity to distance. This approach was developed for predictions with a sample of destinations by Ben-Akiva and Watanatada (1981), who have shown that a reasonable value of β for a uniform spatial distribution of elemental destination alternatives is $2/\bar{d}$, where $\bar{d}$ is the average trip length. For $\beta=0$ the selection probability of a zone is proportional to its size. In other words, the elemental destination alternatives have equal selection probabilities.

This method has optimal properties for model application but is only an intuitively reasonable sampling strategy for model estimation. The following paragraphs describe importance sampling strategies that have been applied in recent studies.

Independent Importance Sampling :

Perform $J-1$ independent draws, one for each element in the set of all alternative zones, excluding the chosen alternative, selecting destination j with probability q_{jn}, and add the chosen zone to the estimation choice set. The resulting sample of alternative zones is characterized by the following probability :

$$\pi_n(D\mid i)=\prod_{\substack{j\in D\\ j\neq i}} q_{jn}\prod_{j\notin D}(1-q_{jn}),\quad i\in D$$

which can also be expressed as

$$\pi_n(D\mid i)=\frac{1}{q_{in}}Q_n(D),\quad i\in D \qquad (22)$$

where

$$Q_n(D)=\prod_{j\in D} q_{jn}\prod_{j\in D}(1-q_{jn})$$

is the unconditional selection probability, and is independent of the chosen alternative. In this method of sampling the size of the choice set is unknown and may range from 1 to J. Therefore, this procedure is mostly useful for applications (e.g., CSI, 1979) and appears to be useful for model estimation when the number of aggregate destinations is relatively small. This deficiency of a highly variable destination sample size is reduced by the following sampling procedure.

Importance Sampling with Replacement :

Draw a sample of size J' from the set of all J destinations, selecting alternative j with probability q_{jn} at each draw. Delete duplicate zones and add the chosen zone, if it was not sampled, to obtain the set D with selection probability

$$\pi_n(D\mid i)=\prod_{\substack{j\in D\\ j\neq i}} q_{jn}\left(\sum_{j\in D} q_{jn}\right)^{J'+1-\tilde{J}},\quad i\in D$$

where $\tilde{J}$ is the size of the set D. Note that $(J'+1-\tilde{J})$ is the number of duplicate zones. The conditional selection probability of the set D can also be expressed by

$$\pi_n(D\mid i)=\frac{l}{q_{in}}Q_n(D),\quad i\in D \qquad (23)$$

where

$$Q_n(D)=\prod_{j\in D} q_{jn}\left(\sum_{j\in D} q_{jn}\right)^{J'+1-\tilde{J}}$$

is again independent of the chosen alternative. In this strategy the size of the set is bounded by

$$l\leq\tilde{J}\leq J'+1$$

where the lower limit is realized when all J' draws yield the chosen alternative. Thus, this importance sampling technigue is advantageous over the previous method because it produces substantially reduced variability of $\tilde{J}$

Stratified Importance Sampling :

The technique of stratified importance sampling avoids the need to specify a selection probability q_{in} for every zone $i=1,\cdots,J$.

The set of J destinations is stratified into A disjoint subsets such that

$$\sum_{a=1}^{A} J_{an}=J$$

where J_{an} is the number of destinations in stratum a for traveller n. The composition of the strata may differ, for example, by the origin location, hence the subscript n. The importance sampling criteria is realized by assigning different selection probabilities in different strata, while maintaining uniform selection probabilities within strata. Let $\tilde{J}_{an}$ be the assigned sample size for stratum $a, a=1,\cdots,A$, and denote by $a(i)$ the stratum of alternative i. Draw a simple random sample (without replacement) of size $\tilde{J}_{an}$ from every stratum except that from the stratum of the chosen alternative i draw only a sample of $(\tilde{J}_{a(i)n}-1)$ destinations and add the chosen alternative. Note that the size of the resulting set D can be fixed at a predetermined value, such that

$$\tilde{J}=\sum_{a=1}^{A}\tilde{J}_{an}$$

and the size of D is uniform across all observations. The probability of selecting a set of destinations D is

$$\pi_n(D\mid i)=C(J_{a(i)n}-1,\ \tilde{J}_{a(i)n}-1)^{-1}\prod_{\substack{a=1\\ a\neq a(i)}}^{A} c(J_{an},\tilde{J}_{an})^{-1},\quad i\in D$$

It can also be expressed as

$$\pi_n(D\mid i)=\frac{J_{a(i)n}}{\tilde{J}_{a(i)n}}Q_n(D),\quad i\in D \qquad (24)$$

where

$$Q_n(D)=\prod_{a=1}^{A}C(J_{an},\tilde{J}_{an})^{-1}$$

The main advantages of this method are its fixed sample size

$$\tilde{J}=J'+1$$

where J' is the total number of random draws. The selection probabilities, given by

$$q_{in}=\frac{\tilde{J}_{a(i)n}}{J_{a(i)n}},\quad i=1,\cdots,J$$

are easier to quantify than in the previous methods. This method was employed in the application in Paris, described in Section 4 of this paper.

Applications of Importance Sampling :

In estimation the importance sampling techniques described above require the estimation of a logit model of the form

$$\pi_n(i\mid D)=\frac{\exp[\mu^* V_{in}-lnq_{in}]}{\sum_{j\in D}\exp[\mu^* V_{jn}-lnq_{jn}]},\quad i\in D \qquad (25)$$

The correction term for sampling of alternatives reduces to $-lnq_{jn}$, $j\in D$, by cancelling out the unconditional probabilities of selecting a set D.

In model applications, if a sample of destinations is constructed with selection probabilities q_{jn}, $j=1,\cdots,J$, the term

$$\sum_{j\in D}\frac{1}{q_{jn}}\exp[\mu^* V_{jn}]$$

is a natural estimator of the denominator of the logit model because $1/q_{jn}$ is the expansion factor for alternative j. The properties of this estimator for the logit model were analyzed by watanatada and Ben-Akiva (1977 and 1979) , who noted that it produces an unbiased numerator and an unbiased denominator though the ratio is consistent but not unbiased.

If the sample used for estimation is also used for prediction, it is necessary to take into account the role of the chosen alternative in the sample selection and the possibility that duplicates were omitted. Denote by $w(i)$ the number of times alternative i was drawn. The expansion factor is $\frac{w(i)}{q_{in}}$ for a non-chosen destination, and for a chosen destination that is always added to the sample the expansion factor is

$$\frac{w(i)+1}{q_{in}+\left(\sum_{j\in D} w(j)\right)^{-1}}$$

(2) Sampling of Observations

The previous subsection defines conditions under which a model describing choice among a large number of alternatives can be estimated on the basis of data describing the selected option and a subset of the rejected options. It confers an ability to reduce the number of alternatives for which data must be gathered and processed, and is the key to a practical resolution of the problem of estimating the unknown parameters.

The nature of destination choice, involving large numbers of alternatives, many of which occur with very low probability; raises another issue with implications for both survey design and analysis. In many studies an accurate estimate of the demand for particular alternative is of central importance whereas a less accurate estimate of the demand for other alternatives will suffice. The sample of observations should be designed with this in mind. For example, long distance movements are the major determinant of the demand for major highways, yet occur relatively infrequently in comparison with shorter distance travel.

Various authors have contributed to the development of a body of theory which addresses this problem (Cosslett, 1981; Manski and McFadden, 1981; Manski and Lerman, 1977). This involves the use of samples which have been enriched with observations of those selecting the important alternatives ("choice-based" designs).

The analysis of data from choice-based samples is straightforward when the model is of the logit form and only a small number of alternatives is involved. For example, many mode choice studies have used choice-based sampling to focus on some modes which are rarely chosen overall but are of major interest to the study.

The key condition under which the logit analysis of choice-based or enriched data becomes straightforward is the specification of a full set of alternative specific constants. In this case the logit model can be estimated as if the data were a random sample. Simple correction factors can be calculated to adjust the alternative-specific constants (see Cosslett, 1981).

The attraction balancing factors in conventional doubly constrained gravity models are equivalent to a full set of destination specific constants. This means that choice-based enrichment can be achieved for aggregate gravity models, and in fact, mixed survey designs involving household surveys enriched with interviews with travellers have been used routinely for many years. Note, however, that these models are specified with very few unknown parameters (often only one).

The difficulty with the estimation of disaggregate destination choice models with a richer specification including a larger number of explanatory variables is that the large number of alternatives implies a requirement for the collection and processing of a correspondingly large amount of data to estimate the full set of alternative-specific constants, not to mention the ensuing computational difficulties in estimation.

There are two responses to the difficulty with choice-based enrichment in estimating disaggregate destination choice models. One is to avoid choice-based enrichment by destinations altogether. This approach characterized previous applications of disaggregate models of destination choice. Choice-based enrichment involved only differential sampling rates for different modes. For example, the sample used in the Paris study described in Section 4 was enriched with trips by public transport.

Recent advances in the analysis of choice-based data opens the way to a second approach allowing sample enrichment by destination choice without estimating a model with a full set of destination specific constants.

The following describes the survey design and analysis devised for a study being conducted by the

Netherlands Ministry of Transport to construct regional travel demand models based entirely on disaggregate modelling technigues.

Firstly, following Cosslett (1981), the theory is developed for the analysis of a data set consisting of a sample of household interviews enriched with traveller interviews. The theory is then extended to apply to the nested logit model and a survey design which enriches alternatives at different levels in its hierarchy. Finally, the implications for the use of a sample of alternatives for model estimation are set out. This theory treats the sample of trips as being composed of a number of subsamples. Moreover, a giver trip can appear in more than one subsample. Each subsample is a simple random sample from some subset of all trips and this subset may be defined with reference to the dimension of travel behavior being modelled. Thus, only some subset of the choice alternatives may be found among the trips in a given subsample. For the Dutch application, there were five subsamples, three strata from a household interview survey, plus a subsample consisting of car-drivers crossing given screenlines, and another subsample of rail passengers crossing the same screenlines. The road and rail surveys involved sampling at a number of different sites, but sampling fractions were equalized as much as possible to allow the whole system of screenlines to be treated as defining just two subsamples : rail passengers and car-drivers.

To develop the theory, it is necessary to introduce some notation :

X_n : the vector of explanatory variables characterising the full set of alternatives, including that selected by observation n, $n=1,\cdots,N$;

i_n : the chosen alternative in observation $n, n=1,\cdots,N$;

θ : a vector of unknown parameters ;

$f(X)$: a probability density function for X ;

$D(s)$: is the set of available alternatives for subsamples, such that

$\bigcup_{s=1}^{S} D(s)=\{1,\cdots,J\}$;

N_i : the number of observations choosing alternative $i, i=1,\cdots J$;

$\tilde{N}_s$: the number of observations in subsample s, for $s=1,\cdots,S$

: $=\sum_{i\in D(s)} N_i$

$\tilde{H}_s$: $\tilde{N}_s/N$;

w_i : the proportion of the population choosing alternative i ; and

$\tilde{w}_s$: $\sum_{i\in D(s)} w_i$

In terms of a choice model with specified probabilities $p(i|X,\theta)$, define

$$p(D(s)|X,\theta)=\sum_{j=1}^{J} n_{js}\, p(j|X,\theta)$$

where

$$n_{is}=\begin{cases}1 \text{ if } i\in D(s)\\ 0 \text{ otherwise ;}\end{cases}$$

The likelihood of observation n for a generalized choice-based sample is given by

$$l_n(\theta,f)=\frac{p(i_n|X_n,\theta)f(X_n)\sum_{s=1}^{S} n_{i_n s}\tilde{H}_s}{\sum_{s=1}^{S} n_{i_n s}\tilde{w}_s} \quad (26)$$

and the kernel of the sample log likelihood is

$$L_N(\theta,f)=\sum_{n=1}^{N} p(i_n|X_n,\theta)+\sum_{n=1}^{N} l_n f(X_n)-\sum_{s=1}^{S}\tilde{N}_s\, ln\tilde{w}_s \quad (27)$$

where

$$\tilde{w}_s=\int dX\, f(X)\, p(D(s)|X,\theta),\quad s=1,\cdots,S \quad (28)$$

This likelihood function is to be maximized over the unknown parameters θ and all possible values of the unknown $f(X)$. The computational difficulty arises from the last term of the likelihood function (27), which contains the integral in (28). Cosslett (1981) developed an efficient estimator that overcomes this difficulty.

Cosslett demonstrates the further simplifications that occur when the aggregate shares of the alternatives $w_i, i=1,\cdots,J$, are known. His treatment easily extends to the case of known stratum shares. When the $\tilde{w}_s, s=1,\cdots,S$, are fixed quantities, the problem is to maximize $L_N(\theta,f)$, given by (27), over θ and f, subject to the stratum shares constraints in (28).

Referring back to the application, the "stratum shares" are the proportions of all trips in the study area falling into the five strata, i. e. having origins in one of the three household survey area types, or being car-driver trips crossing the screenline system, or being train passenger trips crossing the screenline system.

In practice, of course, these shares will not be known with complete accuracy. However, by supplementing the survey data with counts at the cordons and with information on trip rates from large conventional travel surveys, it is possible to form good estimates.

Following Cosslett's approach, it is possible to show that the values of θ which maximize this constrained log likelihood function are also those which can be derived by maximizing the following expression jointly for θ and δ :

$$L'_N=\sum_{n=1}^{N} ln \frac{p(i_n|X_n,\theta)\sum_{s=1}^{S} n_{i_n s}\delta_s}{\sum_{j=1}^{J}(p(j|X_n,\theta)\sum_{s=1}^{S} n_{js}\delta_s]} - \left[\sum_{n=1}^{N} ln \sum_{s=1}^{S} n_{i_n s}\delta_s\right] + N\sum_{s=1}^{S}\delta_s\tilde{w}_s \cdots\cdots (29)$$

where $\delta=\{\delta_1,\cdots,\delta_s,\cdots,\delta_S\}$ is a vector of unknown stratum specific constants.

This in turn proves to be a relatively simple process providing that the choice model is logit and an appropriate system of strata dummy variables is introduced into the model specification. This can be done in such a way that the first expression on the right-hand side of (29) contains no information about δ; crucially, the system of dummy variables is not required to be as large as the number of distinct alternatives, but only as large as the number of distinct strata overlaps minus one.

The term added to the utility of alternative j is given by the expression

$$ln \sum_{s=1}^{S} n_{js}\delta_s, \quad j=1,\cdots,J$$

An overlap of strata occurs when an observed choice can potentially be included in two or more subsamples. For example, a screen line crossing rail trip could also have been observed in the household survey. For every potential overlap there is a particular combination of δ_s's.

In the Dutch application, there are nine such distinct strata overlaps according to the

(i) trip origin from three different areas, and

(ii) the screen-line-crossing by car-driver, rail passenger, or neither of these.

This three-by-three cross-classification defines nine different overlaps, which can be distinguished by eight dummy variables.

With this choice of specification, δ can be estimated by maximizing over just the last two terms on the right-hand side of (20). The resulting values can be used to correct the alternative-spcific constants in the separately fitted choice model.

Alternatively, the adjustment factors can be calculated prior to model estimation, and used in the same way as the destination-sampling correction factors. In this case, no further adjustment is necessary.

This estimator can also be used for a nested logit model, for example, destination and mode, by the usual approach of estimating the sub-models first and passing on expected maximum utilities via the "logsum" variable. Given the estimated "lower level" model, it is unnecessary to make any adjustment for the fact that the enrichment of the compound alternatives is divided unequally between the component sub-alternatives.

The presentation above has outlined the theoretical and the practical considerations involved in the use of samples enriched with choice-based observations for the purpose of destination choice modelling. Given a carefully chosen survey design, and conditioning on estimates of stratum shares, the estimation of nested logit models can proceed in completely conventional way provided only that appropriate correction factors are included in the model specification.

Finally, there is the question of the construction of random sub-samples of alternatives. Here it is necessary to note only that the inclusion of the choice-based correction factors has allowed the problem to be reduced to one of maximizing a (pseudo) likelihood function containing terms which are equivalent to a simple logit model in which all alternatives are available.

As a result, the sample of alternatives for each observation should also be drawn from the full set of alternatives, despite the fact that for the choice-based observations, only a subset of alternatives could have been observed. The asymmetry introduced by the choice-based design is fully compensated by the correction procedure.

4. Estimation Results

The model theory and estimation methods of disaggregate destination choice models presented in the previous sections are illustrated by the following estimation results from two recent studies.

(1) Models of Non-Work Travel for the Paris Region

As part of a larger study, the Regie Autonome des Transports Parisien estimated the models of joint choice of travel mode and destination for personal business and shopping trips presented in Table 1 and 2 (CSE, 1984).

The Ile-de-France study area was divided into 595 zones which constitute the alternative destinations. Of these zones, 269 are in Paris and 326 are in the suburbs. The travel modes are : walk (W), moped or two wheels (2 W), Car (C), and public transport (T). The sampling of destinations was conducted using the stratified importance sampling procedure described in Section 3. (1) 2) For every origin zone the 595 destinations were divided into four non-overlapping strata :

(i) the origin zone;

(ii) the 10 zones closest to the origin zone but not including it ("adjacent zones");

(iii) the remaining zones in the city of Paris;

(iv) the remaining zones outside the city of Paris.

The sample consisted of the single zone in stratum (i) and two zones selected from each of the other three strata, giving a total destination sample size of 7 for each trip. The chosen zone was included in the sample with certainty and the other zones in each stratum were taken as a simple random sample. This sampling strategy results in high probabilities of selection for zones near the zone of residence and low probabilities for all other zones.

In a 1976 survey some 10 000 households in the Ile-de-France region reported all the trips made during a single day and provided information on household and trip-maker characteristics. The estimation samples were constructed by sub-sampling from this survey. Because of the large proportions of walk and car trips for non-work purposes it was decided to use different sub-sampling rates by mode to increase the proportion of public transport and two-wheel trips in the estimation data set. This is a simple application of choice-based sampling, which with a logit model requires alternative-specific correction factors that were first derived by McFadden in Manski and Lerman (1977) and extended by Cosslett (1981).

The level of service attributes were calculated from road and transit networks. The available measures of size of alternative destinations included the usual area, population and employment

Table 1 Estimation Results for a Joint Mode and Destination Choice Model for Personal Business Trips in Paris.

Variable Number	Variable Description	Coefficient Estimate	"t" Statistic
	Mode Constants (Before correction)		
1	Walk (W) constant	2.63	9.9
2	Two Wheels ($2W$) constant	−0.43	1.0
3	Car (C) constant	0.19	0.5
	Level of Service Attributes		
4	Travel time (W)	−0.16	15.3
5	Travel time ($2W$)	−0.16	8.6
6	In-vehicle travel time ($\leq$10 min.) (C)	−0.145	3.8
7	In-vehicle travel time ($>$10 min.) (C)	−0.053	5.2
8	Park-seek time (C)	−0.35	3.4
9	In-vehicle travel time (Transit-T)	−0.046	6.0
10	Out-of-vehicle travel time (T)	−0.073	6.8
11	Travel cost/income ($\leq$.06) (C)	−31.1	4.0
12	Travel cost/income ($>$.06) (C)	−8.42	1.8
13	Parking cost/income (C)	−21.9	2.6
14	Travel cost/income (T)	32.7	3.2
15	(Travel cost)^2/income (T)	−34.7	3.6
	Mode Availability Characteristics		
16	Number of cars (C)	0.85	4.5
17	Head of household dummy (C)	1.04	4.7
18	Male dummy ($2W$)	1.00	2.6
	Destination Area Constants (Before correction)		
19	City of Paris constant	3.61	17.6
20	Suburb constant	2.97	12.4
21	"Adjacent zones" constant	0.41	3.2
	Size Variables ($\ln\theta$) (see eq. 14)		
22	Number of banks, post offices and medical doctors	−1.72	5.1
23	Number of offices of local government and personal services	−4.83	2.3
24	Number of hospitals and clinics	−0−	—
25	Scale parameter (μ) (see eq. 13)	0.85	10.9

Number of observations 906
Loglikelihood at zero −2 633
Loglikelihood at $\hat{\theta}$ −1 510

Table 2 Estimation Results for a Joint Mode and Destination Choice Model for Shopping Trips in Paris.

Variable Number	Variable Description	Coefficient Estimate	"t" Statistic
	Mode Constants (Before correction)		
1	Walk (W) constant	2.16	7.0
2	Two Wheel ($2W$) constant	0.29	0.8
3	Car (C) constant	−0.44	0.9
	Level of Service Attributes		
4	Travel time (W)	−0.20	16.5
5	Travel time ($2W$)	−0.36	9.8
6	In-vehicle travel time ($\leq$10 min.) (C)	−0.28	9.8
7	In-vehicle travel time ($>$10 min.) (C)	−0.15	10.9
8	Park-seek time (C)	−0.58	4.8
9	In-vehicle travel time (T)	−0.084	8.7
10	Out-of-vehicle travel time (T)	−0.13	9.7
11	Travel cost/income/distance (C)	−24.4	1.2
12	Parking cost/income (C)	−15.2	1.5
13	Travel cost/income/distance (T)	−11.7	2.8
	Mode Availability Characteristics		
14	Number of cars (C)	1.07	5.4
15	Employed dummy (C)	1.12	5.6
16	Age $\leq$18 dummy (W)	0.74	2.2
17	Male dummy ($2W$)	1.46	4.6
	Destination Area Constants (Before correction)		
18	City of Paris constant	3.86	16.6
19	Suburb constant	3.14	11.6
20	"Adijacent zones" constant	0.86	6.8
	Mode/Destination Constants (Before correction)		
21	Central business district and transit (T) constant	0.60	3.6
	Size Variables ($\ln\theta$)		
22	Number of department stores and hyper markets	6.48	15.1
23	Number of supermarkets	4.37	7.1
24	Number of other stores	−0−	—
25	Scale Parameter (μ)	0.59	8.6

Number of observations 1 020
Loglikelihood at zero −2 956
Loglikelihood at $\hat{\theta}$ −1 396

variables, and unusually detailed data on the numoer of establishments and their employment for very detailed breakdown of activity types.

The estimated models include a large number of travel time and travel cost variables with significant coefficients. The scale parameter value is significantly different from one in the shopping model, but the coefficients obtained from a constrained estimation of this model with $\mu=1$ were only sightly different from those of the constrained model. It was found that the size variables representing the number of establishments perform better than the employment variables. It is interesting to examine the size measure for the shopping model in Table 2, which is

exp (6.48) · [Number of department stores and hyper-markets]
+exp (4.37) · [Number of supermarkets]
+1 · [Number of other stores]

This means, for example, that a supermarket attracts 79 (=exp (4.37)) more trips than a small shop at the same location.

(2) Model of Work Trips for Maceio

As a part of a joint MIT/GEIPOT study of urban transportation policies and planning methods in Brazil, the model shown in Table 3 was estimated for the city of Maceio, located in the economically depressed Northeast of Brazil (see Kozel and Swait, 1982). The model predicts the choice of a daily home-based work tour, characterized by a travel mode, a work place (or destination), and a frequency of one or two round trips (when returning home for the mid-day meals). Many workers return home for lunch, thus creating two peak periods from 1100 to 1300 hours.

The travel modes include bus (B), taxi (T), auto passenger (AP) and auto driver (AD). The auto driver mode was only available to workers from car-owning households who are 19 years of age or older. If the round-trip travel time by a travel mode between home and a destination is greater than 2 hours the option to return home for lunch was eliminated. The Maceio urbanized area was divided into 35 traffic zones. The origin zone was omitted and a simple random sample of 1/3 of the remaining zones, minus the chosen, was drawn.

The estimation data set of 1016 motorized daily work trips was randomly sub-sampled from a 1977 travel survey of a sample of about 3200 households in Maceio. Neither household nor individual income was available in this survey; instead, the housenold electricity consumption was collected as a proxy. This was found to be highly correlated with household income in studies in other Brazilian cities similar to Maceio.

Table 3 Estimation Results for a Joint Mode, Frequency and Destination Choice Model for Work Trips in Maceio.

Variable Number	Variable Description	Coefficient Estimate	"t" Statistic
	Mode and Frequency Constants		
1	One round trip, taxi (T)	−3 15	4.9
2	One round trip, auto passenger (AP)	−3.83	7.9
3	One round trip, auto driver (AD)	−2.09	4.2
4	Two round trips, bus (B)	0.71	5.9
5	Two round trips, taxi (T)	−0 85	1 4
6	Two round trips, auto passenger (AP)	−2.01	4.8
7	Two round trips, auto driver (AD)	0 15	0.3
	Level of Service Attributes		
8	Total travel time	−0.008	5.8
9	Total travel cost/ln (electricity consumption)	−0 170	4 8
	Mode Availability Characteristics		
10	Household electricity consumption (T)	0 008	3 7
11	Household electricity consumption (AP)	0.010	5 6
12	Household electricity consumption (AD)	0.009	5.2
13	Number of household members (T)	−0.111	1.6
14	Number of household members (AP)	−0.163	2 8
15	Number of household members (AD)	−0 186	3 1
16	Number of vehicles/number of workers (T)	1 78	3.4
17	Number of vehicles/number of workers (AP)	3 16	8.8
18	Number of vehicles/number of workers (AD)	3 01	7.3
	Destination Area and Frequency Constant		
19	Two round trips and CBD	1 01	6.5
	Scale Parameter (μ)		
20	ln (total number of jobs), CBD	0.99	14.6
21	ln (total number of jobs), non CBD	1.14	12 4

Number of observations 1 016
Loglikelihood at zero −3 641
Loglikelihood at $\hat{\theta}$ −2 071

Separate bus and highway networks and information on bus and taxi fares were used by GEIPOT to generate estimates of travel times and travel costs. Employment data was available to represent the sizes of alternative zones as work places.

The possibility of committing important specification errors are many with a complex joint choice model of employment location, travel mode to work and frequency of a mid-day round trip for lunch at home. Accordingly, several tests of parameter stability across choice dimensions were performed by estimating conditional choice models of one or two choice dimensions holding the remaining dimension (s) constant, e. g. choice of mode and frequency given destination. These tests evidenced a high degree of parameter stability, which supports the joint choice structure which has been adopted. In addition, a market segmentation test did not reject the hypothesis of equal model parameters for professional and non-professional occupation types. These results are due in large part to the inclusion in the model specification of a large set of socio-economic variables.

It is encouraging to note that the single size variables have coefficients near one, for CBD and non-CBD destinations, indicating the appropriateness of the zonal aggregation scheme and the model specification.

REFERENCES

1) Ben-Akiva, M., : Passenger Travel Demand Forecasting : Applications of Disaggregate Models and Directions for Research, *Proceedings of the 3rd world conference on Transport Research*, Martinus Niihoff, The Hague, 1977

2) Ben-Akiva, M. : Structure of Passenger Travel Demand Models, Ph. D. Dissertation, Department of Civil Engineering,

M. I. T., 1973.

3) Ben-Akiva, M., Adler, T., Jacobson, J. and Manheim, M. L. : Experiments to Clarify Priorities in Urban Travel Forecasting Research and Development, Final Report to the U. S. Department of Transportation, Program of University Research, 1977.

4) Ben-Akiva, M., Kullman, B. C., Sherman, L. and Daly, A. : Aggregate Forecasting with a System of Disaggregate Travel Demand Models, *Proceedings of the PTRC Summer Annual Meeting*, 1978.

5) Ben-Akiva, M. and Lerman, S. R. : *Travel Behavior : Theories, Models and Prediction Methods*, in process for forthcoming publication, MIT Press Cambridge, MA, 1984.

6) Ben-Akiva, M. and Watanatada, T. : Application of a Continuous Spatial Choice Logit Model, in Manski, C. F. and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge MA, 1981.

7) Cambridge Systematics Europe (CSE) : Estimation and Application of Disaggregate Models of Mode and Destination Choice, draft report prepared for Direction des Etudes Generales, Regie Autononome des Transport Parisien, 1984.

8) Cambridge Systematics Europe (CSE) : Travel Demand Models, Zuidvleugel Study Report 7, prepared for the Dienst Verkeerskunde Rijkswaterstaat, The Netherlands Ministry of Transport, The Hague, 1983.

9) Cambridge Systematics, Inc. (CSI) : The Relationship of Changes in Urban Highway Supply to Vehicle Miles of Travel, report prepared for the National Co-operative Highway Research Program, Washington DC, 1979.

10) Charles River Associates (CRA) : A Disaggregate Behavioral Model of Urban Travel Demand, report prepared for the Federal Highway Administration, U. S. Department of Transportation, Washington, D. C., 1972.

11) Cosslett, S. R. : Efficient Estimation of Discrete Choice Models, in Manski, C. F., and McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge MA, 1981.

12) Daly, A. : Estimating Choice Models Containing Attraction Variables, *Transportation Research B*, Vol. 16 B, No. 1, pp. 5~15, 1982.

13) Domencich, T. and McFadden, D. : *Urban Travel Demand ; A Behavioral Analysis*, North-Holland, Amsterdam, 1975.

14) Hammersley, J. M. and Handscomb, D. C. : *Monte Carlo Methods*, Methuen, London, 1965.

15) Johnson, N. L. and Kotz, S. : *Continuous Univariate Distributions*, Wiley, New York, 1970.

16) Kitamura, R., Kostyniuk, L. P. and Ting, K-L : Aggregation in Spatial Choice Modelling, *Transportation Science*, Vol. 13, No. 4, pp. 325~342, 1979.

17) Koppelman, F. S. and Hauser, J. R. : Destination Choice Behavior for Non-Grocery-Shopping Trips, *Transportation Research Record* 673, pp. 157~165, 1978.

18) Kozel, V. and Swait, J. : Maceio Travel Demand Model System Calibration Results, Studies of Urban Travel Behavior and Policy Analysis in Maceio, Volume II, Center for Transportation Studies, M. I. T., Cambridge MA, 1982.

19) Lerman, S. R. : A Disaggregate Behavioral Model of Urban Mobility Decisions, Ph. D. Dissertation, Department of Civil Engineering, M. I. T., Cambridge MA, 1975.

20) Manski, C. F. and Lerman, S. R. : The Estimation of Choice Probabilities from Choice-Based Samples, *Econometrica*, Vol. 45, pp. 1977~1988, 1977.

21) Manski, C. F., and McFadden, D. : Alternative Estimators and Sample Designs for Discrete Choice Analysis, in Manski, C. F. and McFadden, D. (eds.), *Strcuctural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA, 1981.

22) McFadden, D. : Modelling the Choice of Residential Location, in Karlqvist, A., Lundqvist, L., Snickars, F., and Weibull, J. W. (eds.), *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 1978.

23) McFadden, D. : Conditional Logit Analysis of Qualitative Choice Behavior, in Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1974.

24) Morichi, S., Ishida, H. and Yai, T. : Comparison of Various Utility Functions for Behavioral Travel Demand Models, *Procedings of the world Conference on Transport Research*, SNV, Hamburg, 1984.

25) Netherlands Ministry of Transport : SIGMO Final Reports, Volumes I-IV, Projectbureau IVVS, The Hague, 1977.

26) Rao, C. R. : *Linear Statistitical Inference and its Applications*, Second Edition, Wiley, New York, 1973.

27) Richards, M. G. and Ben-Akiva, M. : *A Disaggregate Travel Demand Model*, D. C. Heath, Lexington MA, 1975.

28) Ruiter, E. R. and Ben-Akiva, M. : Disaggregate Travel Demand Models for the San Francisco Area : System Structure, Component Models and Application Procedures, *Transportation Research Record* 673, 1978.

29) Silman, L. A. : Disaggregate Travel Demand Models for Short-Term Forecasting, Information Paper 81, Israel Institute of Transportation Planning and Research, Tel-Aviv, 1980.

30) Watanatada T. and Ben-Akiva, M. : Forecasting Urban Travel Demand for Quick Policy Analysis with Disaggregate Choice Models : A Monte Carlo Simulation Approach, Transportation Research, Vol. 13 A, pp. 241~248, 1979.

31) Watanatada, T. and Ben-Akiva, M. : Development of an Aggregate Model of Urbanized Area Travel Behavior, Final Report prepared for the U. S. Department of Transportation, Washington DC, 1977.

(RECIEVED MAY 2 1984)

論 文 要 旨

非集計分布交通モデル

土木学会論文集 No. 347／Ⅳ-1 1984. 7

本論文は，非集計行動モデルによるOD交通量の推定方法を検討したものである．

非集計交通行動モデルは，基礎となる論理の明確さ，高い説明力，従来の集計モデルと比して少ないサンプル数でモデル構築できるという効率性等の利点から注目されているが，適用例のほとんどは交通機関選択の予測であった．トリップの目的地選択問題（OD交通量の推定問題）では，目的地をゾーンとして把える必要があり，個人の意識する目的地とゾーンの差異，ゾーニングの確定方法や選択代替肢数（目的地数）の多さ等，非集計モデル構築に対する多くの難しさを含んでいる．

本論文は大別して3つの部分から構成されている．

第1は実際の目的地をゾーンとして集計化することに対する選択モデルの理論展開を行った2．であり，その基礎となっているのはNested Logit Modelの考え方である．

第2は，非集計目的地選択モデル構築のためのサンプリングと推定問題に関する3．である．3.1では，莫大な要素数の代替肢全集合からサンプリングにより小数の目的地代替肢を抽出してパラメーター推定を行う方法を，McFadden (1978) が住宅地選択問題に際して展開したものを基礎に検討している．3.2ではモデル構築に用いるトリップデータのサンプリング問題を，Choice-based Sampling, Enriched Samplingの理論に基づいて検討している．

第3は，フランスとブラジルで適用した事例を示した4．である．パリに関しては，595ゾーンに対して，906サンプルを用いた私用トリップの目的地・交通手段同時選択モデルと1 020サンプルを用いた買物トリップの同種モデル構築結果を示している．ブラジルのマセイオ市の例は，ゾーン数35に対し1 016サンプルを用いた通勤トリップの目的地，交通手段さらに昼食時に一時帰宅するか否かの同時選択モデルである．

（文責：森地　茂）