Transpn Res.-B Vol 13B, pp 189–205
© Pergamon Press Ltd., 1979 Printed in Great Britain

0191-2615/79/0901-0189/$02.00/0

ESTIMATING CHOICE PROBABILITIES AMONG NESTED ALTERNATIVES

YOSEF SHEFFI

Civil Engineering Department, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

and

Transportation System Center, U.S. Department of Transportation, Cambridge, MA 02139, U.S.A.

(*Received* 17 *October* 1977; *in revised form* 6 *November* 1978)

Abstract—In several travel choice situations (e.g. automobile ownership level and trip frequency) the alternatives available to an individual randomly chosen from the population exhibit some internal choice-related ranking: the choice of a given alternative implies that all lower-ranked alternatives have been chosen. Such alternatives are referred to as "nested". This paper presents a model for estimating choice probabilities among nested alternatives. The model is devised from the well known logit model and uses existing logit maximum-likelihood estimation techniques (and computer packages). The approach is shown to be more attractive than the multinomial logit and linear regression models, from a theoretical point of view, yet cheaper than the multinomial probit model. The model is developed in a disaggregate, utility maximization framework. An example application, estimating probabilities of trip frequencies by elderly individuals is presented.

1. INTRODUCTION

This paper presents an econometric model for estimating choice probabilities among nested alternatives. The term nested alternatives (or nested choice set) is used to describe a choice set where the alternatives are associated with some ranking and the choice of any alternative implies that all lower-ranked alternatives have been chosen as well. The objective of this paper is to present a model that is based on a set of assumptions that differs from the ones leading to the multinomial logit (MNL), thereby overcoming some undesirable properties of the MNL model, yet retaining the choice theory base and the computational ease of the MNL model.

In order to establish the basis for the model presented here, a short review of the disaggregate demand modelling framework and the MNL model is presented below.†

Underlying disaggregate demand models is the hypothesis that in a choice situation, an individual associates a value with each available alternative. This value is commonly referred to in the travel demand literature as "utility". The utility of an alternative is a function of the decision-maker's characteristics and the alternative's attributes, and the decision-maker is assumed to choose the alternative which yields the greatest utility. Since utilities are not observable, they are modelled as random variables distributed across the population of decision-makers.

Most operational models assume a functional form of the utility which is linear in the parameters and with additive disturbance term. Specifically, the utility of alternative i to an individual randomly chosen from the population, U_i, is given by:

$$U_i = \beta Z_i + \xi_i, \tag{1}$$

where β is a vector of parameters, Z_i is a vector of functions of characteristics of the individual under consideration and the attributes of alternative i, and ξ_i is a random variable representing an unobserved disturbance or error term. The term βZ_i is denoted V_i and termed the observed utility (or mean utility, since without loss of generality, it can be assumed that $E[U_i] = V_i$).

Let I denote the index set of S, the set of alternatives available to a randomly chosen decision-maker. The probability that alternative i is chosen, P_i, is given by:

$$P_i = Pr(U_i \geq U_j; \forall j \in I); \forall i \in I. \tag{2}$$

†A more detailed discussion of disaggregate demand models and the related choice theory can be found in a variety of references including the books by Domencich and McFadden (1975) and Richards and Ben-Akiva (1975).

In order to solve for the choice probabilities (eqn (2)), one has to assume a probability law for the error terms ξ_i (see eqn (1)). When the ξ_i's are assumed to be independently and identically distributed (i.i.d.) Gumbel variates, i.e.

$$P_r(\xi_i \leq \omega) = \exp(e^{-\omega}), \tag{3}$$

eqn (2) reduces to the well known multinomial logit (MNL) formula:

$$P_i = \frac{e^{v_i}}{\sum_{\forall j \in I} e^{v_j}}, \tag{4}$$

as shown by Beilner and Jacobs (1972) and McFadden (1973).

There have been numerous applications of the MNL model to travel demand analysis, and computer packages for estimating the vector of parameters, β, using maximum likelihood are readily available (e.g. the National Bureau of Economic Research's TROLL System). These maximum likelihood estimates are consistent, asymptotically efficient and asymptotically normal.

The major drawback of the MNL model is that it exhibits the so-called "Independence from Irrelevant Alternatives" (IIA) property (Luce, 1959). The IIA characteristic of the MNL model gives rise in some instances, to predicted behavior which is unacceptably counter-intuitive. As shown by many researchers (e.g. Mayberry (1970), Schneider (1973), Florian and Fox (1976) and Daganzo and Sheffi (1977)) the logit model tends to overestimate the choice probability of correlated alternatives. This pecularity of the MNL model is rooted in the assumption that the error terms, ξ_i's, are independent random variables, thus correlations among the utility functions cannot be captured.

This paper deals with another choice situation where (as shown in Section 2) the independence (among the alternatives' utility functions) assumption leads to counter intuitive results, and therefore the multinomial logit cannot be utilized. This is the case of the choice set comprising nested alternatives, for which an alternative model is developed.

The case of nested alternatives is presented in the next section. Section 3 presents the postulates on the structure of the utility functions, upon which the model is based and Section 4 derives the model from these postulates. Sections 5–9 demonstrate the use of the proposed model to estimate and predict trip frequencies among elderly individuals. The model is also compared (costwise) with other available models, in Section 8. Section 10 concludes the paper, summarizing the model's features.

2. THE CASE OF NESTED ALTERNATIVES

The choice situation that this paper deals with is characterized by the alternatives being naturally rank-ordered. An example of this might be the number of automobiles owned by a given household. In this case the alternatives include: owning no cars, one car, two cars, etc. Another example involves trip generation (daily trip frequency) in which households (or individuals) are assumed to choose to undertake no trips, one trip per day, two trips per day, etc. A third example of ranked, nested alternatives (outside the context of travel demand models) is the family size decision, which has a very similar structure to the car ownership example, with regard to the number of children that a family chooses to have.

The basic characteristic of the choice situation under concentration here is that the ith alternative cannot be chosen without choosing all the alternatives $0, 1, \ldots (i-1)$, beforehand. In fact, alternative i is considered only if alternative $(i-1)$ (and all those preceding it) have been chosen. At this point, the decision-maker presumably cannot reverse former decisions and the choice is only between accepting i or rejecting it. In case of rejection, the final choice is $(i-1)$. The $(i+1)$th alternative (and all higher ranked alternatives) are not even considered if i is rejected.

A choice model such as the one just described implies a particular interdependency among the alternatives included in a decision-maker's choice set. This interdependency excludes the possibility of using the ordinary MNL model to estimate the model's parameters and predict the probability of choosing each alternative. To demonstrate this interdependency and to exemplify the IIA property of multinomial logit (which is the stumbling block in this case, as in many others—as mentioned in Section 1), consider an auto ownership model. With the estimation of a MNL model one can predict the probability that a given household will own i cars, $i = 0, 1, 2, \ldots$. Now consider (for the sake of argument only) predicting the implications of a new car use restraint policy. The policy includes an extremely high tax on each car owned by any household from the third up. The prediction of the MNL model will be that most of the owners of three or more cars will now be distributed among zero, one, and two cars according to the relative share of these alternatives before the new policy was implemented. If the sum of probabilities of owning three or more cars was, say, 0.25 before the new policy implementation, the MNL model's prediction would involve a 25% increase (approximately) in the share of carless families. The share of one and two car families would also be predicted to grow by 25%. This is obviously not realistic since one would expect the number of carless and one car families to remain almost unchanged as a result of the above mentioned policy. It is more realistic to assume that most of the previous multicar families will now choose to own two cars. This example demonstrates the effect of the IIA property of the MNL model.

Several models, other than MNL-based models, might have been considered for the case of nested alternatives. These include several variants of linear regression, and the multinomial probit (MNP) model. Using the notation introduced in Section 1, the estimated equation in the case of a linear regression is:

$$E[i] = \beta Z + \xi \tag{5}$$

where $E[i]$ is the expected number of alternatives chosen (e.g. trips per day, car ownership level, etc.). Also included in the class of linear regression model are estimation techniques such as the ones developed by McKelvey (1973) and Tobin (1958) for the analysis of models with limited dependent variables. All the above mentioned regression models cannot incorporate interdependencies between the alternatives such as the one described in this section. Another drawback of this class of models is that they do not offer an interpretation within the framework of utility maximization and choice theory. A third drawback is that when utilized in a prediction mode, only the means of the dependent variable is predicted (at the disaggregate level) rather than the whole distribution (as is the case with choice-based models).

One can also estimate probabilities, using linear forms for aggregated data to estimate the equation:

$$Pr(i) = \beta Z + \xi. \tag{6}$$

A model of this form was estimated by CRA (1972) and others. Aside from all the above mentioned drawbacks this model can be estimated only on aggregate data. Furthermore, in prediction, certain combinations of characteristics and attributes (values of Z) might produce "probabilities" that are out of the zero-one range.

The multinomial probit (MNP) is another possibility for the estimation of choice probabilities among nested alternatives. The model is based on the assumption that the error term vector $\xi = (\ldots, \xi_i, \ldots)$ in eqn (1) is multivariate normally distributed. Since the multivariate normal distribution admits full parametrization in terms of a covariance matrix, the MNP does not exhibit the IIA property and can be used to construct a choice model in the environment of any correlations among the alternatives' utilities. Unfortunately, albeit recent developments in MNP estimation procedures (see Daganzo, Bouthelier and Sheffi, 1977a, and 1977b) the model is still relatively expensive to use, for problems involving a large number of parameters and alternatives.

In the next two sections, a model of the choice among nested alternatives based upon the MNL model is suggested. The proposed model is developed within the choice theory and utility maximization framework, and does not exhibit the IIA property of MNL. Prediction with the new model and estimation of its coefficients are generally an order of magnitude cheaper than probit and, do not require the development of a special computer code.

3. BASIC POSTULATES

The choice process under consideration can be characterized by two basic postulates:

A. No alternative can be chosen without it implying that all the lower ranked alternatives had been chosen. If an alternative is not chosen, no higher-ranked alternative can be chosen.

B. The marginal utilities of the alternatives in the choice set are independent random variables.

Postulate A is straightforward and is really an observation that is implied from the description of the process (it is impossible to choose to have a second child before the first is born or to purchase the third car first). To formalize this postulate we will use the choice theory basis discussed in Section 1. Consider an arbitrary decision maker facing a choice set S, containing mutually exclusive alternatives each associated with a utility level U_i. The subscript i is part of an index set I, containing all integers ordered by the nesting relationships.† The first part of postulate A states that if i is the chosen alternative, the event:

$$(U_i \geq U_{i-1} | U_{i-1} \geq U_j; \forall j < i-1, j \in I) = \{ \bigcap_{i>j>2} (U_j \geq U_{j-1} | U_{j-1} \geq U_{j-2}) \cap (U_1 \geq U_0\} \forall i \in I. \quad (7)$$

This part of Postulate A implies that the utilities are a monotonically increasing function of their index set for all alternatives ranked lower than the chosen one.

The second part of Postulate A implies that the utility of any alternatives ranked higher than the one following the chosen alternative, is lower than the utility of the alternative following the chosen one. In other words, the event:

$$(U_j > U_i | U_i > U_{i+1}; \forall j > i+1, j \in I) = \phi. \quad (8)$$

In probability space, this means that the probability of choosing an alternative equals zero if its predecessor was not chosen.

Postulate B is the core of the model and this section will be devoted to explaining and analyzing it. To begin with, this postulate means that the choices are made "one at a time". This might not be so in reality as a family may decide *a priori* on a desired family size level and not re-evaluate the conditions continuously, or to purchase two cars at the same time. In some cases such as the auto ownership example, this type of behavior is rare enough to be ignored. However, in other applications (such as the household size) this behavior might be quite common and invalidate the use of the model presented in this paper. Thus the validity of this assumption should be checked before the model is considered for use.

The second implication of Postulate B is that the random variables formed from pairwise subtraction of the utilities of the given alternatives are independent random variables. In probability space, it follows that the probability of the intersection of any two events, each defined by an arbitrary alternative's utility being greater than or equal to the utility of its adjacent alternative, equals the product of the corresponding probabilities:

$$\begin{aligned} Pr\{(U_{i+1} \geq U_i) \cap (U_{J+1} \geq U_j)\} &= Pr\{\Delta U_{i+1} \geq 0) \cap (\Delta U_{j+1} \geq 0)\} \\ &= Pr\{\Delta U_{i+1} \geq 0\} \cdot Pr\{\Delta U_{j+1} \geq 0\} \quad \forall i \neq j; i, j \in I \end{aligned} \quad (9)$$

where

$$\Delta U_i = U_i - U_{i-1}$$
$$\Delta U_j = U_j - U_{j-1}.$$

An intuitive argument for Postulate B, leading to the above result (eqn (9)) is developed in the remainder of this section. The argument is based on a particular specification of a multinomial probit model for the problem of nested alternatives, and through a simple transformation leads to independently distributed marginal utilities.

Assume a multivariate normal (MVN) distribution of the error term vector ξ associated with

†The index set I can include only a subset of all integers, according to the problem at hand and without loss of generality; S (and I) can also vary among decision makers. In fact, the size of the choice set does not have to be defined prior to estimation, as will become apparent in the following sections.

the utilities of the alternatives open to an arbitrary decision-maker. It follows from eqn (1) that:

$$U \sim MVN(V, \Sigma) \tag{10}$$

where U is a vector of utilities of the alternatives, V is the vector of observed (measured) utilities, and Σ is the covariance matrix of the error term vector ξ, associated with U.

The structure of the covariance matrix Σ, can be specified following the description of the choice process among nested alternatives. Since the choice of any alternative implies that all lower-ranked alternatives have been chosen (Postulate A) a reasonable specification might be the following:

$$\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_0^2 & \sigma_0^2 & \sigma_0^2 & - & - & - & - \\ \sigma_0^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & - & - & - & - \\ \sigma_0^2 & \sigma_1^2 & \sigma_2^2 & \sigma_2^2 & - & - & - & - \\ \sigma_0^2 & \sigma_1^2 & \sigma_2^2 & \sigma_3^2 & - & - & - & - \\ - & - & - & - & & & & \\ - & - & - & - & & & & \end{bmatrix}$$

or: $(\Sigma_{ij}) = \sigma_k^2;\ k = \min\{i, j\};\ \forall i, j \in I$

where σ_i^2 is the variance associated with the error term of the utility of the ith alternative.†

Before transforming the MNP model to a model with independently distributed marginal utilities, the above mentioned specification should be motivated. However, note that it cannot be "justified" since the "true" econometric model is generally unknown, thus only an intuitive motivation is offered here.

As shown by several researchers, many choice problems can be visualized as a choice of a path in an imaginary network.‡ The paths of such a network are associated with utilities (the alternatives' utilities), and each decision-maker is assumed to choose a route from his "origin" to his "destination", i.e. choose an alternative. Since the utilities are random variables, the concepts used in stochastic network assignment problems (see Daganzo and Sheffi (1977)) can be used to find the chosen routes.

A network structure corresponding to the case of nested alternatives is shown in Fig. 1. This network illustrates the nesting relationships among the alternatives (i.e. if alternative i is chosen, all lower-ranked ones are chosen as well), and motivates eqn (11) if one accepts the view that the covariance among "routes" i and j is only in the "links" they share in common. In other words, the covariance structure is related to the overlap between routes.

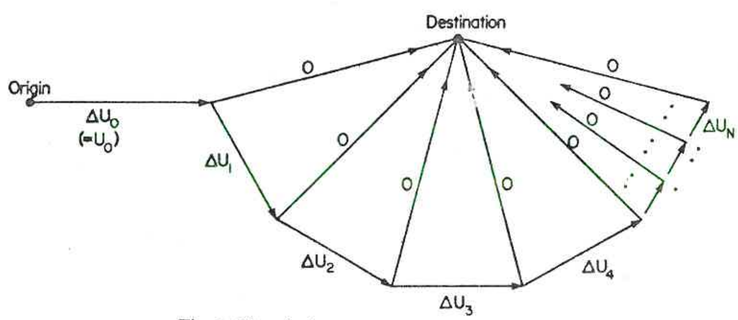

Fig. 1. The choice structure as a network.

†Similar specification of a covariance matrix for an auto-ownership choice model utilizing MNP, has been used by Daganzo, Bouthelier and Sheffi (1977b).

‡The visualization of travel choices in the transportation market as a network problem was suggested by Dafermos (1974) whos developed a deterministic model of choice. The idea was further developed by Danzig *et al.*, who demonstrated that the elastic demand traffic assignment problem can be solved as a fixed demand problem on an expanded network. Recently, Sheffi and Daganzo (1978) and Sheffi (1978) developed a framework for visualizing most transportation problems as network problems within the framework of choice-theory based travel demand models.

At this point, once the utility functions themselves have been specified, and eqn (11) is accepted as the covariance specification, we have a completely specified choice model.

However, as is the case with any other choice model, only the differences in utilities are estimable. Thus, consider now the random vector $\Delta U = M \cdot U$ where M is the linear transformation matrix given by:

$$(M_{ij}) = \begin{cases} -1 & \text{for } i - j = 1 \\ \delta_{ij} & \text{otherwise} \end{cases} \tag{12}$$

where δ_{ij} is the Kroneker delta. M is a square matrix whose order equals the number of alternatives ("routes" in Fig. 1), open to the arbitrary decision maker under consideration. The random vector ΔU corresponds to the circumferencial links of the network shown in Fig. 1 or, more generally to the difference in the utilities of the alternatives:

$$\Delta U = MU = \begin{Bmatrix} U_0 \\ U_1 - U_0 \\ U_2 - U_2 \\ U_3 - U_2 \\ --- \\ --- \\ --- \end{Bmatrix}. \tag{13}$$

The probability density function of the new random vector is given by:

$$\Delta U \sim MVN(MV, M\Sigma M^T) \tag{14}$$

where:

$$MV = \begin{Bmatrix} V_0 \\ V_1 - V_0 \\ V_2 - V_1 \\ V_3 - V_2 \\ --- \\ --- \\ --- \end{Bmatrix} \tag{15a}$$

and:

$$M\Sigma M^T = \begin{Bmatrix} \sigma_0^2 & 0 & 0 & 0 & --- & --- \\ 0 & \sigma_1^2 - \sigma_0^2 & 0 & 0 & --- & --- \\ 0 & 0 & \sigma_2^2 - \sigma_1^2 & --- & --- & --- \\ 0 & 0 & 0 & \sigma_3^2 - \sigma_2^2 & --- & --- \\ 0 & 0 & 0 & 0 & --- & --- \\ --- & --- & --- & --- & --- & --- \\ --- & --- & --- & --- & --- & --- \end{Bmatrix} \tag{15b}$$

This result gives an intuitive background for Postulate B. Note that the non-diagonal entries of the covariance matrix shown in eqn (15b) are all zero, i.e. the differences in utilities are independently distributed variates.

In our case we do not use the normal distribution assumption (eqn (10)) but rather the logit model. The above mentioned MVN-based model was only intended to provide some intuitive basis for Postulate B and motivate the use of the logit model for the differences in utilities, since those can be assumed to be independent.

An alternative treatment of Postulate B is a comparison with the MNL model assumptions. In the MNL case, the utilities themselves are assumed independent while in our model only the marginal utilities are assumed to be independent. As shown in the sequel, this assumption obviates the IIA property of the MNL model.

The two postulates of this section are used in the next one to set up the model.

4. THE MODEL

The following derivation of the model is based on the random utility theory reviewed in Section 1, and uses the nomenclature developed in the preceding section. Repeating eqn (2), the probability that alternative i will be chosen is:

$$P_i = Pr[U_i \geq U_j;\ \forall j \in I] \quad \forall i \in I,$$

which is the utility maximization principle. Focusing on the chosen alternative, i, the index set I corresponding to the choice set S, can be divided into two mutually exclusive subsets as follows:

$$I' = \{\forall j < i; j \in I\}$$
$$I'' = \{\forall j > i; j \in I\}$$

Utilizing the independence property of Postulate B, expressed in eqn (9), eqn (2) becomes:

$$P_i = Pr(U_i \geq U_j; \forall j \in I') \cdot Pr(U_i \geq U_j; \forall j \in I''). \tag{16}$$

Focusing our attention on the subset I', the first term in the product above can be expanded as a product of marginal probabilities:

$$\begin{aligned} Pr(U_i \geq U_j; \forall j \in I') = {} & Pr(U_i \geq U_{i-1} | U_{i-1} \geq U_j;\ \forall j < i-1). \\ & \cdots Pr(U_k \geq U_{k-1} | U_{k-1} \geq U_l;\ \forall l < k-1) \cdots \\ & \cdots Pr(U_2 \geq U_1 | U_1 \geq U_0) \cdot Pr(U_1 \geq U_0). \end{aligned} \tag{17}$$

But, from eqn (7) (first part of Postulate A) we conclude that:

$$Pr(U_i \geq U_j; \forall j \in I') = \left\{ \prod_{i \geq k \geq 2} Pr(U_k \geq U_{k-1} | U_{k-1} \geq U_{k-2}) \right\} \cdot Pr(U_1 \geq U_0).$$

Invoking again eqn (6) for the pairwise independence of all terms in the above product, we get:

$$Pr(U_i \geq U_j; \forall j \in I') = \prod_{k=1}^{i} Pr(U_k \geq U_{k-1}). \tag{18}$$

The second term in the product on the right hand side of eqn (16) corresponds to the set I'' and to the alternatives that are ranked higher than the chosen one. It follows directly from eqn (8) that:

$$Pr(U_i \geq U_j; \forall j \in I'') = Pr(U_i \geq U_{i+1}). \tag{19}$$

Combining the two index subsets I' and I'', and substituting eqn (18) and eqn (19) in eqn (16), the final result becomes:

$$P_i = Pr(U_i \geq U_{i+1}) \cdot \prod_{k=1}^{i} Pr(U_k \geq U_{k-1}). \tag{20}$$

This is the random utility model that corresponds to the choice among ordered integer alternatives. To simplify further treatment of the model, denote the above mentioned binary probabilities by $P_{i+1|i}$ [i.e. $P_{i+1|i} = Pr(U_{i+1} \geq U_i)$]. Using this notation, the model can be expressed as:

$$P_i = (1 - P_{i+1|i}) \prod_{k=1}^{i} P_{k|k-1}. \tag{21}$$

The model is a product of independent binary choices. Estimating each of the binary probabilities can be carried out through the use of a logit model. The use of a logit model is justified in the case of a binary choice problem since the difficulties arising from the IIA property of the multinomial logit do not exist in a binary model.

When this model is estimated, every decision maker, choosing alternative i is a source for $i+1$ data points. The observation is that he/she preferred 1 over 0; 2 over 1; ... ; i over $i-1$ and i over $i+1$.

This suggests one straightforward way to estimate the model's coefficients. Let $V_{j+1|j} = V_{j+1} - V_j$, where V_j is the observed utility of alternative j. To estimate the parameters of $V_{1|0}$, the whole sample is used with the choice model: $P_{1|0} = 1/[1+\exp(V_0 - V_1)]$ for all sampled individuals who chose Alternative 1 or a higher ranked one, and $(1-P_{1|0})$ for all individuals who chose alternative 0. Next all sampled individuals who have chosen zero are excluded from the sample and the coefficients of $V_{2|1}$ are estimated, with $P_{2|1} = 1/[1+\exp(V_1 - V_2)]$ for all individuals who chose alternative 2 or a higher ranked one, and $(1-P_{2|1})$ for all individuals who chose alternative 1. Next, the coefficients of $V_{3|2}$ are estimated, using only the part of the sample which includes individuals choosing alternative 2 or a higher ranked one, and so forth. In a sample including observed choices up to alternative k, one would have to estimate $k+1$ such binary models in order to estimate the coefficients of the utility functions associated with all the alternatives. The prediction, then, would be performed using eqn (21), utilizing any aggregation technique such as classification, simulation, or complete enumeration (see Koppelman (1976) for a discussion of these and other aggregation methods).

However, the above estimation procedure is very inefficient since there is a fixed cost involved in setting up many separate computer programs. A more important disadvantage of this procedure is that it does not enable the user to apply restrictions across alternatives or to investigate trends that are related to the utilities as a function of their index set. Such restrictions might be imposed due to prior theoretical hypotheses and in order to obtain predictions beyond the limit of the highest chosen alternative observed in the sample. (These points are further explained below.) Fortunately, the model can be estimated with all binary submodels considered simultaneously. The reason for this is that our model (eqn (21)) is in the form of a product of probabilities, and so is the likelihood function.

The likelihood function for any choice model is:

$$L = \prod_{t=1}^{N} P_{i,t} \tag{22}$$

where $P_{i,t}$ is the probability that individual t (from the sample) will choose alternative i, and there are N decision makers in the sample. Substituting the model of eqn (21), one gets:

$$L = \prod_{t=1}^{N} (1 - P_{i+1|i,t}) \prod_{k=1}^{i} P_{k|k-1,t}, \tag{23}$$

and since multiplication is an associative operator, the maximum likelihood computer package will maximize over the binary conditional probabilities, performing the equivalent of the cumbersome method described in the preceding paragraph, in only one run. (The methods are equivalent only for the unconstrained case, of course.)

In specifying the nested-alternatives-logit model, one has to consider two different specification issues. The first one is common to all econometric models: the functional form of the utility function. Since (due to the unimodality characteristics of its likelihood functions) the logit model is typically specified as linear in the parameters, one has only to specify the functional form of the explanatory variables. Also, as in any other choice model, one has to specify which variables are alternative-specific and which are generic.†

The second specification issue has to do with the ranking of the alternatives; one can specify a functional form of the generic variables with respect to the utilities' index set. For example, in an auto ownership model, Burns, *et al.* (1975) dealt directly with a specification with respect to

†Generic variables are those whose parameters are constrained to be equal across alternatives. Some variables can, of course, be specified as generic only across some of the alternatives.

the index set (in a MNL model framework) by introducing two composite variables. The first one had to do with the increased accessibility provided which each additional automobile and the second one reflected the decreasing available income for consumption of other goods, as a larger component of the income is tied up in owning automobiles. An "available income" variable for alternative i (owning i cars) can be specified as:

$$INC - f(PR_i)$$

where INC is the total annual household income, PR_i is the annual cost of owning and operating i automobiles and $f(\cdot)$ is a functional form to be specified. For example, a simple specification of $f(\cdot)$ might be $f(PR_i) = \beta \cdot i \cdot PR_i$, where β is a parameter to be estimated, i.e. linear specification with respect to the index set. (The above mentioned researchers used a different specification.)†

Note that in the model proposed in this paper, the specification with respect to the index set in a testable hypothesis in the same sense that a generic entry is. In the above mentioned example, one can specify the "available income" variable for alternative i as:

$$INC - \beta_i \cdot PR_i.$$

Once the estimation has been carried out, any hypothesis of the functional form of β_i with respect to i can be tested.

Note also that it is the simultaneous estimation of the nested alternatives logit that enables one to specify generic variables in this model and to specify the measured utilities with respect to the index set.

In the second part of this paper, from Section 5 on, this model is used to estimate a trip generation example and the results are compared with ordinary least squares estimation. The emphasis in this example is on demonstrating the estimation technique and the prediction procedure, as well as comparison to other models.

5. THE PROBLEM AND THE DATA‡

The structure described in the preceding section was used to estimate a trip generation model. The choice of daily travel frequency falls in the category of nested alternatives; possible choices might be 0 (not to take any trip) or 1, 2, 3 ... daily trips. The model is applied to non-work vehicular trips of elderly individuals.

The decision-making unit used in this example is the elderly members of a household taken as a group. Elderly individuals are defined here as persons over 65 years of age. The household elderly group was chosen as the behavioral unit rather than individuals because *household* characteristics and interactions between travel of household members were expected to be the significant determinants of frequency-of-travel choice.

In general, a trip generation problem might not conform with our model of ordered nested alternatives in two aspects. First, there is a problem with using the entire household as the behavioral unit. Trips might be decided upon simultaneously and carried out by more than one person. The model cannot account for this phenomenon since the "one choice at a time" assumption is basic to its structure. The second difficulty is that multi-destination trip chains (in which a number of trips are combined in a single tour from the residence) cannot be accounted for in our model, and tours have to be counted as trips.

Fortunately both those deficiencies of modelling trip generation are not significant in the sample. Less than 5% of the trips were carried out by more than one individual and only 3% of the total trips in the sample were found to be portions of tours. Thus in the case of the elderly, our model is quite suitable for modelling trip generation behavior.

The sample used to estimate the model is taken from a 25% random subsample tape of the 1968 home interview survey in the Washington, D.C. Metropolitan area. All complete inter-

†As an aside, note that Burns *et al.* (1975) tried to capture some of the interdependence among the alternatives through a careful specification of the measured utilities (i.e. specification with respect to the index set). Such specification should always be attempted since it can add to the explanatory power of the model.

‡A more detailed description of this example can be found in Hendrickson and Sheffi (1976).

views of households with elderly members were used for the estimation. In all, 684 households, 774 elderly persons and 509 home-origin, one way trips were represented.

The variables of interest which are available from the home interview dataset include:

(1) Household data: location of residence, household size, auto ownership level, income (by category) and availability of transit service.

(2) Individual data: age, possession of driver's license and employment status.

(3) Trip data: trip purpose, destination and mode of travel.

In addition to the data available from the home interview survey itself, travel impedance and some land use data were available from a companion dataset prepared for the Washington Council of Governments by R. H. Pratt and Associates. Variables of interest from this dataset include:

(1) Trip Data: transit fare, transit time, auto travel time and automobile travel cost. (For each origin-destination pair.)

(2) Land Use Data: commercial, residential and total area per district.

Two pieces of information that might have been important in determining trip generation behavior of elderly individuals, were unfortunately unavailable. There were no data concerning the physical condition of the interviewed persons and there was no information about walk trips. However other variables were used to approximate and capture the effect of those. The possession of driver licenses and the number of elderly workers serve as indicators of physical conditions, and the commercial area in residence district was expected to capture the opportunity for walking trips.

6. SPECIFICATION OF THE UTILITY FUNCTIONS

The utility functions specified for the model consist of three types of variables: household characteristics, district land use and travel impedance. The variables used are defined in Table 1. Means and standard deviations of the variables are summarized in Table 2.

Table 1. Definition of variables

Household Characteristics	
ELDERLY:	number of elderly members of a household (65 years of age or older)
NON-ELDERLY:	number of non-elderly members of a household
CARS:	number of automobiles available to a household
LICENSES:	number of elderly household members possessing driver's licenses
WORKERS:	number of employed <u>elderly</u> household members
INCOME:	annual household income (in hundreds of dollars)
Land Use Characteristics	
COMM AREA:	percentage of area in a district devoted to commercial purposes
Travel Impedance	
TRANSIT:	transit availability; binary variable (1 if transit route is within 1/2 mile; 0 otherwise)
AUTO TIME:	average automobile travel time for a district (minutes)
TRANSIT FARE:	average transit fare for a district (cents)
Other	
TRIPS:	number of one-way, non-work vehicular trips by elderly persons from a household

Table 2. Variable means and standard deviations

Variable	Mean	Standard Deviation
TRIPS	.87	1.35
ELDERLY	1.31	.49
NON-ELDERLY	1.13	1.83
CARS	.92	.87
LICENSES	.53	.63
WORKERS	.10	.30
INCOME	84.50	63.40
COMM. AREA	48.24	33.55
AUTO. TIME	15.80	4.30
TRANSIT FARE	21.90	19.00

Travel times and costs represent dimensions of travel impedance. Some form of composite, average impedance variables must be constructed for every household to represent the impedance for trips which might have been taken but were not. The level of service variables used in our model for this purpose are averages across all trips actually made from each district. As shown by Ben-Akiva and Lerman (1977), this measure of accessibility is not consistent with random utility theory. They (and other researchers) suggest the expectation of the maximum utility alternative among all combinations of mode, destination and route choice for this purpose. Although theoretically superior, this measure was not used since it requires the estimation of mode choice models, destination choice models and an assignment model, tasks which were beyond the scope of this work.

Specification of the utility functions is shown in Fig. 2, with the alternatives defined up to six daily trips (the maximum reported in the sample). All the variables, excluding the constants, entered in generic form. The specification of all variables with respect to the index set was chosen simply as linear (i.e. $V_i = i \cdot \beta_j X_j$, where X_j is the generic variable under consideration and β_j its coefficient), excluding two variables which were hypothesized to influence only the

Utility Functions	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}
V_0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V_1	1	0	0	0	0	0	1/ELD	NE	LCS	WKR	CRS	CA	TR	AT	LIF
V_2	0	1	0	0	0	0	2/ELD	2*NE	LCS	2*WKR	2*CRS	2*CA	TR	2*AT	2*LIF
V_3	0	C	1	0	0	0	3/ELD	3*NE	LCS	3*WKR	3*CRS	3*CA	TR	3*AT	3*LIF
V_4	0	0	0	1	0	0	4/ELD	4*NE	LCS	4*WKR	4*CRS	4*CA	TR	4*AT	4*LIF
V_5	0	0	0	0	1	0	5/ELD	5*NE	LCS	5*WKR	5*CRS	5*CA	TR	5*AT	5*LIF
V_6	0	0	0	0	0	1	6/ELD	6*NE	LCS	6*WKR	6*CRS	6*CA	TR	6*AT	6*LIF

LEGEND (see Table 1 for variables definitions)
ELD - ELDERLY
NE - NON-ELDERLY
LCS - LICENSES
WKR - WORKERS
CRS - CARS
CA - COMM AREA
TR - TRANSIT
AT - AUTO TIME
LIF - $\ln\left(\frac{\text{INCOME}}{\text{TRANSIT FARE}}\right)$

Fig. 2. Specification of the utility functions.

"no trip" versus "take a trip" (i.e. one or more) decision. These two variables are the transit availability (TRANSIT) and the driver licenses (LICENSES). The reason for this specification is that no more than one daily transit trip was expected to be taken and thus, it was expected that TRANSIT would influence only the 0/1 decision. As to LICENSES, the above mentioned specification was chosen in order to reflect the expected role of this variable as a proxy for physical condition, distinguishing between "home staying" and "outgoing" elderly. The rest of this section deals with the specification of the variables themselves.

The number of trips made by the elderly was expected to be positively correlated with the number of elderly individuals in the household (ELDERLY). However, this effect was hypothesized to exhibit diminishing returns due to a substitution effect among the elderly for some trip purposes. Thus, this variable entered as an hyperbolic function (1/ELDERLY, with the expectation of a negative coefficient, β_7).

The specification of the cost impedance variable as natural log of the ratio of income to transit fare [Ln(INCOME/TRANSIT FARE)], requires some explanation. The ratio specification was chosen in order to capture an income effect. That is, individuals with higher income tended to live in the suburbs, encountering higher travel cost. Since income was expected to have a non-linear effect on trip generation, the natural log of the ratio was specified.

Two level-of-service variables, automobile cost and transit time were omitted from the final specification in earlier trial runs. Both of them were very low in magnitude and not statistically significant. It was hypothesized that since the trips taken were relatively short, the fixed cost of owning a car (captured by the car ownership variable) was a much more important determinant of travel behavior than the variable cost. The insignificance of the transit travel time variable is probably due to the relatively low income and the low number of workers among the elderly which cause, in turn, a low value of their time. As mentioned above, including all the dimensions of travel impedance in a consistent way, would have required estimation of other travel demand models.

7. ESTIMATION RESULTS

The results of the estimation are shown in Table 3 for the variables in the estimated utility functions—$V_{i+1|i}$. In all cases except for auto time (AUTO.TIME), commercial area (COMM. AREA), transit availability (TRANSIT), and for the constant for $V_{1|0}$, the coefficients are statistically different from zero at the 1% confidence level. Estimation of the complete model required 0.16 min of CPU time on the MIT IBM 370/168 computer. Standard errors are reported in parentheses below each coefficient estimate.

As expected, the number of driver's licenses held by elderly members of the household is very significant in explaining the decision to travel (that is, to take one or more trips). Complementary estimation runs, using other specifications for this variable, verified the initial hypothesis as to the role of this variable.† It serves to divide the elderly population into relatively mobile and immobile groups. In this context, it is believed that the holding of a driver's license is an indicator of physical fitness.

The negative sign for the number of non-elderly household members, (NON-ELDERLY), indicates that elderly individuals living in households with non-elderly members travel relatively less frequently than individuals in exclusively elderly households. It is likely that non-elderly household members substitute for the elderly for some necessary trips (such as shopping). To some extent, this effect may also be caused by differences in health; healthier elderly individuals might tend to live independently.

The relatively large standard error for commercial area density, COMM.AREA and transit availability, TRANSIT, suggests that these variables are misspecified in the data. The commercial area density variable was intended to represent the opportunity for walking trips. Of interest are the availability of stores and other potential trip destinations within walking distance of each residence. Since the COMM.AREA variable used is specific to each district, (no other data was available), microscale opportunities for each household are not accurately presented.

†In those runs this variable was specified as alternative-specific and all coefficients for utilities higher than the $V_{1|0}$ were found to be statistically indistinguishable from zero at the 10% significance level.

Table 3. Estimated coefficients of the utility functions

VARIABLE	$V_{1\|0}$	$V_{2\|1}$	$V_{3\|2}$	$V_{4\|3}$	$V_{5\|4}$	$V_{6\|5}$
CONSTANTS	.626 (.447)	1.29 (.378)	1.33 (.390)	1.28 (.431)	1.57 (.492)	1.28 (.549)
1/ELDERLY	-.988 (.286	-.988 (.286)	-.988 (.286)	-.988 (.286)	-.988 (.286)	-.988 (.286)
NON-ELDERLY	-.247 (.064)	-.247 (.064)	-.247 (.064)	-.247 (.064)	-.247 (.064)	-.247 (.064)
CARS	.385 (.110)	.385 (.110)	.385 (.110)	.385 (.110)	.385 (.110)	.385 (.110)
LICENSES	1.00 (1.76)	0	0	0	0	0
WORKERS	.561 (.198)	.561 (.198)	.561 (.198)	.561 (.198)	.561 (.198)	.561 (.198)
COMM.AREA	-.00322 (.00198)	-.00322 (.00198)	-.00322 (.00198)	-.00322 (.00198)	-.00322 (.00198)	-.00322 (.00198)
TRANSIT	.237 (.241)	0	0	0	0	0
AUTO.TIME	-.0287 (.0152)	-.0287 (.0152)	-.0287 (.0152)	-.0287 (0.152)	-.0287 (.0152)	-.0287 (.0152)
ln(INCOME / TRANSIT.FARE)	.233 (.081)	.233 (.081)	.233 (.081)	.233 (.081)	.233 (.081)	.233 (.081)

Log likelihood at zero	-668.9	Number of observations	1093
Log likelihood at convergence	-345.8	(NOTE: STANDARD ERROR IN PARENTHESES)	
Percent correctly predicted	66.6		

Transit availability is a binary variable which has the value 1 if a transit route is within 1/2 mile of residence and zero otherwise. However, later analysis of the data indicated that for all transit trips actually made by elderly individuals, the maximum distance walked to a transit stop was three blocks, which is considerably less than the 1/2 mile criterion. Transit might be perceived by elderly as available only within 1/4 mile of a transit route. As a result, this variable is misspecified in the data and this might have caused the large dispersion in the estimator of its coefficient.

The WORKERS variable, the number of employed elderly, has a positive sign. It is probably capturing some health status and "outgoing" tendency among the elderly. It might also indicate some income level, but probably more than this, independency that leads to more trips for all purposes.

Before demonstrating the use of the model, the above results are compared mainly costwise with the linear regression and the multinomial probit models.

8. COMPARISON WITH LINEAR REGRESSION AND MULTINOMIAL PROBIT

For comparison, a least squares regression was applied to the same disaggregate data. The same variables as in the nested-alternatives-logit were used with the same functional form. The estimation results were:

$$\text{TRIPS} = \underset{}{1.08} - \underset{(0.240)}{0.415} * 1/\text{ELDERLY} - \underset{(0.034)}{0.080} * \text{NON-ELDERLY} + \underset{(0.079)}{0.150} * \text{CARS} + \underset{(0.098)}{0.741} *$$

$$\text{LICENSES} + \underset{(0.168)}{0.610} * \text{WORKERS} - \underset{(0.0014)}{0.002} * \text{COMM.AREA} + \underset{(0.132)}{0.086} * \text{TRANSIT} - \underset{(0.012)}{0.010} *$$

$$\text{AUTO.TIME} + \underset{(0.062)}{0.113} * \text{Ln}\left(\frac{\text{INCOME}}{\text{TRANSIT FARE}}\right)$$

$$R^2 = 0.25 \qquad\qquad S = 1.18$$

with standard errors reported in parentheses.

The results of the least square regression are good, in the sense that all the signs agree with our prior expectations. The magnitude of the estimated coefficients, cannot be meaningfully compared between the models since the scales of the integer-alternatives-logit and the OLS above are different. The same holds true for the standard "goodness of fit" measures: R^2 and "per cent right" for the regression and logit, respectively, which are not comparable measures.

The primary explanatory variables in the model are driver's licenses (LICENSES), number of non-elderly members of the household (NON-ELDERLY) and the number of elderly workers (WORKERS) (these variables are significantly different from zero, based on t-statistic at the 1% confidence level). The results tend to indicate that the two variables that are very likely to be highly correlated with physical fitness are the most significant ones.

Estimation of the least squares equation required only 0.021 min of CPU time on the MIT IBM 370/168 computer. This is an order of magnitude cheaper than the nested-alternative-logit estimation.

A multinomial probit model, the other alternative to the nested-alternatives logit was not calibrated but the cost of applying it can be estimated using the formula developed by Daganzo, Bouthelier and Sheffi (1977b). According to this formula, the CPU time required to estimate a multinomial probit model for the same problem is approximately 8 min. This applies to the same computer and the same number of iterations to convergence of the maximum likelihood procedure, as the logit estimation. Thus, estimation of the nested alternatives logit is more than an order of magnitude cheaper than a MNP model.

We now turn to demonstrate the use of the model in prediction carrying on with the above example.

9. AGGREGATION AND PREDICTION

For most practical purposes and policy implications one would want to predict the probability mass function of trip generation in the population. The aggregation technique used here is that of complete enumeration of all data observations. For each household t, the expected choice probabilities:

$$P_{k+1|k,t} = \frac{1}{1+e^{V_{k+1|k,t}}} \tag{24a}$$

were computed. Then the probability of choosing each number of trips was calculated by eqn (21):

$$P_{i,t} = (1 - P_{i+1|i,t}) \prod_{k=1}^{i} P_{k|k-1,t}, \tag{24b}$$

and these probabilities were summed over all observations in the data and normalized:

$$\bar{P}_i = \frac{1}{N} \sum_{t=1}^{N} P_{i,t} \tag{24c}$$

where $\bar{P}_i$ is the share of the population choosing alternative i, and N is the number of observations (households) in the sample.

The number of trips in the predicted distribution is, of course, not limited to six trips since variables that were specified in generic form (and all of them but the constants were), could be used to predict the probability of taking any number of trips. The observed shares in the sample and the predicted shares ($\bar{P}_i$) are shown in the second and third columns, respectively, of Table 4 below.

The fourth column of this table exhibits the aggregate probabilities obtained from using average household and alternatives' characteristics for prediction.† Comparing it to the complete enumeration method one can conclude that the aggregation bias (created by using the naive approach) seems relatively small in this example.

†This aggregation method was termed "The Naive Approach" by Koppelman (1976).

To illustrate the use of the model for prediction, the impacts of three transit improvement strategies were simulated and the expected aggregate trip generation probabilities calculated, using complete enumeration. The three simulated strategies were:

(1) a 50% reduction in transit fare;

(2) assuring transit availability to all elderly individuals;

(3) strategies 1 and 2 in combination.

To perform these predictions, the data observations were altered to reflect the impact of the various strategies. For example, strategy 3 was modelled by imposing TRANSIT = 1 on all households, and reducing all observed transit fares by 50%. The results are summarized in the last three columns of Table 4. Changes in travel cost seem to have a larger effect on trip generation than changes in transit availability. This is not surprising since only 22.6% of the elderly households do not have transit available (within 1/2 mile).

The predicted impact from reducing travel cost implies a value for the (arc) fare elasticity of demand for travel. A 50% reduction in cost resulted in a 14.6% increase in travel, so the implied arc elasticity of trip generation by the elderly with respect to the transit fare is -0.30. It should be emphasized, however, that this does not imply that transit will capture the entire increase in travel; transit fare acts as a composite variable reflecting travel costs in the model. Differentiating the modal impact of impedance variable changes would require re-specification of the model to differentiate among modal choices. Alternatively, a modal split model may be applied using the predictions of travel obtained from the trip-generation model.

10. CONCLUSIONS

The elderly trip generation behavior discussed in the second part of this paper was brought only as a demonstration of the model presented in the first part. It exemplifies a whole class of problems to which the nested-alternatives-logit can be applied. This class of problems is characterized by ranked and nested alternatives in the choice set, i.e. an observed choice implies that all lower-ranked alternatives have been chosen as well. The essence of the model is in capturing the special correlation implied by the definition of nested alternatives and overcoming the difficulty arising from applying the multinomial logit to this problem: the independence from irrelevant alternatives. From the mathematical point of view the principle upon which the model is based is to transform the logit model and the data itself using the same

Table 4. Observed and predicted aggregate probabilities

# of Trips	Observed	Base Case	Average Household	Transit Improvement Strategies: Fare Reduction	Improved Availability	Both
0	.567	.567	.559	.536	.557	.525
1	.221	.224	.229	.225	.230	.230
2	.101	.098	.111	.102	.100	.108
3	.053	.050	.048	.057	.051	.058
4	.022	.022	.025	.026	.022	.026
5	.015	.016	.014	.020	.016	.020
6	.000	.016	.009	.021	.016	.021
7	.000	.005	.003	.007	.005	.007
8	.000	.002	.002	.002	.002	.003
9	.000	.001	.001	.001	.001	.001
10	.000	.000	.000	.000	.000	.000
Mean	.872	.886	.865	1.017	.906	1.038
% Increase				14.8	2.3	17.2

transformation shown in eqn (13); thus no posterior back-transformation of the result is needed and one gets the desired coefficient directly from the computer output.

The advantages of the nested-alternatives-logit presented in this paper are the following:

(1) It is a disaggregate choice model and as such has all the advantages of logit, probit and other choice models (i.e. relative statistical efficiency in the use of the data, a choice-theory based functional form, and the ability to predict the whole distribution of choice, rather than the mean only).

(2) The problem of independence from irrelevant alternatives, crippling logit analysis is overcome with the use of this model, albeit specific correlations among the alternatives.

(3) The use of existing logit computer packages for the estimation of the model is straightforward and no special maximum likelihood routine is needed.

(4) No prior definition of the number of alternatives in each individual's choice set is required (e.g. for a MNL or a MNP model the alternatives have to be specified as (for example) 0, 1, 3 or more). This eliminates possible biases that might have been introduced by lumping the higher ranked alternatives as a single alternative.

(5) By having to specify the model with respect to the utilities' index set as well, one is provided with a convenient tool for introducing patterns of behavior that relate simply to the ranking of the alternatives.

(6) The computation costs are modest: an order of magnitude less than a probit model and generally comparable with MNL analysis.

As an aside, note that the example application of the model might suggest several conclusions regarding trip generation behavior by elderly, including:

(1) The elderly may be divided into mobile and relatively immobile groups on the basis of driver's license ownership.

(2) Elderly individuals living independently of non-elderly tend to travel more frequently.

(3) Travel impedance and level of service seem to significantly affect trip generation.

(4) In future surveys concerning the elderly population, the following issues have to be taken care of: (a) Specification of the transit availability variable. (b) Information regarding walking trips. (c) Information regarding the physical condition of the interviewed individuals.

More working experience with the model would be needed in order to establish it as a common tool for estimating choice probabilities among nested, ordered alternatives. However, its theoretical basis and the results of the example above seem to warrant its usefulness.

Acknowledgements—The author would like to thank Moshe Ben-Akiva of M.I.T. who initiated the idea of this model (based on a private communication with Daniel McFadden of M.I.T.) and provided many insightful suggestions. Thanks also go to Carlos Daganzo of the University of California, Berkeley and S. Lerman of M.I.T. for their helpful comments.

REFERENCES

Ben-Akiva M. and Lerman S. (1977) Disaggregate travel and mobility choice models and measures of accessibility. Paper presented at the *3rd Int. Conf. on Behavioral Travel Modelling*, Adelaide, Australia.

Burns L., Golob T. and Nicolaidis G. (1976) Theory of urban household automobile ownership decisions. *Trans. Res. Record* **569**, 56–73.

Burrell J. (1976) Multipath route assignment: a comparison of two methods. In *Traffic Equilibrium Methods*; *Lecture Notes in Economics and Mathematical Systems*, Vol. 118. Springer-Verlag, New York.

Dafermos S. (1976) Integrated equilibrium flow models for transportation planning. In *Traffic Equilibrium Methods*; *Lecture Notes in Economics and Mathematical Systems*, Vol. 118. Springer-Verlag, New York.

Charles River Associates (C.R.A.) (1972) *A Disaggregate Behavioral Model of Travel Demand*, Report to FHWA, U.S. DOT, Washington, D.C.

Daganzo C., Bouthelier F. and Sheffi Y. (1977a) An efficient approach to estimate and predict with multinomial probit models. A paper presented at the *56th Annual TRB Meeting*, Washington, D.C.

Daganzo C., Bouthelier F. and Sheffi Y. (1977b) Multinomial probit and qualitative choice: a computationally efficient algorithm. *Transpn Sci.* 11([illegible]), 338–358.

Daganzo C. and Sheffi Y. (1977) On stochastic models of traffic assignment. *Transpn Sci.* 11(3), 253–274.

Danzig G., Maier F. and Landsdowne Z. (1976) *The Application of Decomposition to Transportation Network Analysis.* Control Analysis Coop., Report DOTOTSC-76-26.

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

Florian M. and Fox B. (1976) On the probabilistic origin of Dial's multipath traffic assignment model. *Transpn Res.* 10(6), 339–341.

Hendrickson C. and Sheffi Y. (1976) *A Disaggregate, Serially Structured Model of Trip Generation by Elderly Individuals.* M.I.T. Center of Trans. Studies Report No. 76-16.

Luce R. (19[illegible]) *Individual Choice Behavior.* Wiley, New York.

Mayberry J. (1970) Structural requirements for abstract-mode models of passenger transportation. In *The Demand for Travel-Theory and Measurement* (Edited by R. Quandt). Heath, Lexington, Mass.

McFadden D. (1973) Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (Edited by P. Zarembecka). Academic Press, New York.

Richards M. and Ben-Akiva M. (1975) *A Disaggregate Travel Demand Model*. Saxon House, Westmead, England.

Schneider M. (1973) Probability maximization in networks. *Proc. of Int. Conf. on Trans. Res.*, Bruges, Belgium, pp. 748–755.

Sheffi Y. (1978) *Transportation Networks Equilibration with Discrete Choice Models*, Ph.D. Thesis, Department of Civil Engineering, M.I.T., Cambridge, Mass.

Sheffi Y. and Daganzo C. (1977) A unified approach to transportation forecasting: hypernetwork and supply-demand equilibrium with disaggregate demand models. Paper presented at the *57th TRB Annual Meeting*.