CRT-527

PROBABILITY: The P-2 program for the Standard and Generalized BOX-COX LOGIT models with disaggregate data

by

Tran Liem
Marc Gaudry

Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7

This work was supported at various stages by Transport Canada, by the F.C.A.R. program and the M.E.S.S.T. of Quebec, and by the S.S.H.R.C. and the N.S.E.R.C. of Canada.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Centre de recherche sur les transports — Publication #527
June 1987 — Revised April 1993

ABSTRACT

The P-2 program is designed for the maximum likelihood estimation of the Multinomial LOGIT model in which the dependent variables are of discrete type. In transportation, the basic unit of observation is often an individual (disaggregate data), and the behavior of the latter faced with the choice of travel mode between two or more alternatives is by nature a discrete event. Advantages of disaggregate demand analysis include consistency of the disaggregate approach with the utility maximization concept of consumer demand theory, and a rich specification which can capture the main characteristics of the individual as well as the degree of intermodal competition due to the availability of a broad range of available explanatory variables.

To introduce flexible functional forms other than the usual linear and logarithmic forms of the independent variables, different Box-Cox transformations can be applied on these variables: the Box-Cox parameters can be fixed or estimated jointly with the coefficients of the variables. Program outputs include two types of elasticities (elasticity at sample means of the variables and weighted aggregate elasticity) as well as simulated forecasts.

Keywords: Box-Cox Multinomial Logit, Discrete Choice Models, Disaggregate Data, Maximum Likelihood Estimation, Elasticities, Simulated Forecast.

RÉSUMÉ

Le programme P-2 est destiné à l'estimation selon la méthode du maximum de vraisemblance du modèle LOGIT Multinomial dans lequel les variables dépendantes sont de type discret. Dans les problèmes de transport, il est fréquent que l'unité d'observation de base soit un individu (données désagrégées), et que le comportement de ce dernier face au choix d'un mode de déplacement entre deux ou plusieurs alternatives soit par nature un événement discret. Les avantages de l'analyse de la demande désagrégée comprennent la consistence de l'approche désagrégée avec le concept de la maximisation de l'utilité dans la théorie de la demande du consommateur, et une spécification beaucoup plus riche qui peut capter les principales caractéristiques de l'individu ainsi que le degré de concurrence intermodale grâce à la disponibilité d'un ensemble beaucoup plus vaste de variables explicatives.

Pour introduire des formes fonctionnelles flexibles autres que les formes usuelles, linéaire et logarithmique, des variables indépendantes, différentes transformations de Box-Cox peuvent être appliquées à ces variables: les paramètres de type Box-Cox peuvent être fixés ou estimés conjointement avec les coefficients des variables. Le logiciel calcule notamment deux types d'élasticités (élasticité à la moyenne des variables et élasticité agrégée-pondérée) ainsi que des prévisions simulées.

Mots-clés: Logit Multinomial avec Box-Cox, Modèles de Choix Discret, Données Désagrégées, Estimation du Maximum de Vraisemblance, Élasticités, Prévision par Simulation.

OTHER TRIO DOCUMENTATION AND ORDER FORM . . . . See last two pages

Contents

List of Tables

INTRODUCTION

In transportation demand analysis, most data are collected from surveys of individual households. The statistical models using this kind of information are called disaggregate demand models since the basic unit of observation is an individual, in contrast to the aggregate demand models where data on travel behavior are aggregated into shares at the regional or national level. A number of advantages of disaggregate over aggregate demand analysis includes the consistency of the disaggregate approach with the utility maximization concept of the consumer demand theory, and a much richer specification which can capture the main characteristics of the individual as well as the degree of intermodal competition due to the availability of a much broader range of explanatory variables. Nevertheless, the disaggregate analysis has its own practical limitations such as the huge costs associated with the data collection and the difficult problem of model estimation in the case of a large number of alternative choices and a complex specification of demand.

Since the behavior of an individual faced with the choice of travel mode between two or more alternatives is by nature a discrete event, i.e. the dependent variable is a dummy variable (0 if there is no choice for a given mode, and 1 if there is a choice), discrete choice analysis must be used to explain this particular decision-making process in the disaggregate demand models. Discrete choice problems were first studied in mathematical psychology (Thurstone 1927, Luce 1959, Marschak 1960, Luce and Suppes 1965, Bock and Jones 1968, Tversky 1972) and have also been analyzed in the biometric field (Berkson 1944, Cox 1970, Finney 1971). Early applications of discrete choice models to disaggregate transportation demand include only two alternatives of travel mode (Warner 1962, Lisco 1967, Quarmby 1967, Lave 1969, Stopher 1969, Gronau 1970, de Donnea 1971, McGillivray 1972, Talvitie 1972, Wigner 1973, Watson 1974). The extension of mode choice models to more than two alternatives was made during the early 1970's. The major contributions to economic and statistical ideas from which a new area of research, called "behavioral demand modeling", emerged are due to McFadden (1974, Domencich and McFadden 1975) who established the theoretical grounding of discrete choice models in random utility maximization and the use of disaggregate small-sample data in the estimation of choice models via the maximum likelihood method. A review of econometric applications of discrete choice models is given in Amemiya (1981), Manski (1981) and McFadden (1982).

Among the available discrete choice models, the most computationally tractable and widely used is the Multinomial Logit (**MNL**) model which explains the behavior of an individual in terms of a probability associated with each choice of travel mode among the different alternatives, since the dependent variable is a dummy variable. Extensive treatment of this model and its applications in travel demand analysis can be found in Hensher and Johnson (1981) and Ben-Akiva and Lerman (1985).

The probability class is designed for users who are interested in: (1) estimating the coefficients of the independent variables in MNL model, or (2) estimating them jointly with the parameters of the Box-Cox transformations on the variables, by the maximum likelihood procedure. It is an extension to other existing programs which do not include (2) as an option [(MDA (Lerman 1982), BLOGIT (Crittle and Johnson 1980), MLOGIT in QUAIL 3.0 (McFadden et al. 1979), MLOGPROT (Hensher 1979) and PROLO (Cragg and Lisco 1968)].

Due to the use of the Box-Cox transformation which permits to choose flexible functional forms of the independent variables, three model types can be estimated within the probability class:

1. The **Standard Linear Model Type** if all the independent variables appear in linear form, without any Box-Cox transformation involved. One main specification constraint of this model type is that

the characteristics of an alternative (e.g. the network variables in transportation demand) can only be included in its own representative utility component.

2. The **Standard Box-Cox Model Type** which is similar to the Standard Linear Model Type in terms of the specification of the independent variables, but with the Box-Cox transformations on some variables of interest to let the data choose the optimal functional forms which can be different from the usual linear form.
3. The **Generalized Box-Cox Model Type** (Gaudry, 1978) which removes the specification constraint on the characteristics of an alternative in the Standard Box-Cox Model Type by allowing them to be included also in some or all of the other representative utility components. The specification of this model type is very similar to a demand equation system where each equation is a function of all the commodity prices in the context of the consumer demand theory.

1 STATISTICAL MODEL

1.1 Log-Likelihood Function

Let C be the universal (or master) choice set which includes all potential choices for a given population and define M to be the number of elements (or alternatives) in C. Given that each individual n has a feasible choice set denoted by $C_n \subseteq C$, the total utility of each alternative $i \in C_n$ is expressed as a sum of two components:

$$U_{in} = V_{in} + \epsilon_{in} \tag{1}$$

where V_{in} is the observed (or representative utility) component which is a known function of the attributes of the mode and the socioeconomic characteristics of the individual, and ϵ_{in} is an unobserved, random component which contains the unobserved attributes or taste variations and other influences due to measurement errors and the use of instrumental (or proxy) variables.

According to the utility-maximizing principle, each individual is assumed to select alternative i with the highest utility $U_{in} > U_{jn}$ for all $i \neq j$. Since the utilities are random across the individuals, this event is associated with a probability:

$$P_n(i) = \text{ Prob}\left(V_{in} + \epsilon_{in} > V_{jn} + \epsilon_{jn}\;,\ \text{all } j \in C_n\right). \tag{2}$$

Depending on the assumption made about the joint probability distribution of the set of disturbances $\{\epsilon_{in}\;,\ i \in C_n\}$, a specific random utility model can be obtained. For the **MNL** model, each error ϵ_{in} is assumed to be independently and identically distributed over the population and for each individual according to the Type 1 extreme value (or Gumbel) distribution which has the following cumulative distribution function:

$$\text{Prob}\left(\epsilon_{in} \leq \epsilon\right) = \exp\left[-\exp\left(-\sqrt{\pi^2/6\sigma^2}\;\;\epsilon\right)\right] \tag{3}$$

with mode zero and variance σ^2 for each alternative $i \in C_n$.

The probability that individual n chooses alternative i can now be expressed as:

$$P_n(i) = \exp\left(V_{in}\right) / \sum_{j \in C_n} \exp\left(V_{jn}\right) \tag{4}$$

where the representative component of utility V_{jn} is a linear function in the K transformed variables $X_{jnk}^{(\lambda_k)}$'s:

$$V_{jn} = \sum_{k=1}^{K} \beta_k X_{jnk}^{(\lambda k)}\ . \tag{5}$$

The $X_{jnk}^{(\lambda_k)}$ is the basic Box-Cox transformation with parameter λ_k on the positive variable X_{jnk}:

$$X_{jnk}^{(\lambda_k)} = \begin{cases} \left(X_{jnk}^{\lambda_k} - 1\right)/\lambda_k & \text{if } \lambda_k \neq 0, \\ \ln X_{jnk} & \text{if } \lambda_k = 0. \end{cases} \tag{6}$$

This transformation is a class of power transformations already considered by Anscombe and Tukey (1954), Tukey (1957), Box and Tidwell (1962), Box and Cox (1964), Zarembka (1974) and Spitzer (1978). If λ_k is equal to one (or zero), then the variable is entered in its linear (or log-linear) form.

Since the transformation is continuous for all possible values of the λ-parameter, but defined only for a positive variable, it is clearly understood that in (5) some of the X_{jnk}'s cannot be subject to the Box-Cox transformation, namely the constant, the dummies and the ordinary variables which are not strictly positive.

The likelihood function for a random sample of size N can be viewed as the product of the choice probabilities associated with M subsets of observations, in which the first subset includes N_1 individuals observed to have chosen alternative 1, the next one N_2 individuals to have chosen alternative 2, etc., all observations being independent:

$$\mathcal{L} = \prod_{n=1}^{N_1} P_n(1) \prod_{n=N_1+1}^{N_1+N_2} P_n(2) \ldots \prod_{n=N_{M-1}+1}^{N} P_n(M). \tag{7}$$

This expression can be simplified by defining a dummy variable d_{in} such that $d_{in} = 1$ if individual n has chosen alternative $i \in C_n$ and $d_{in} = 0$ otherwise:

$$\mathcal{L} = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{d_{in}}. \tag{8}$$

The corresponding log-likelihood function can now be written as:

$$L = \ln \mathcal{L} = \sum_{n=1}^{N} \sum_{i \in C_n} d_{in} \ln P_n(i). \tag{9}$$

1.2 Computational Aspects

To maximize the log-likelihood function (9) with respect to the β-coefficients and the λ-parameters in (5), the Davidon-Fletcher-Powell optimization method (Fletcher and Powell 1963) which requires the first derivatives of the function is used.

From now on, the summations Σ_n^N , $\Sigma_{i \in C_n}$, $\Sigma_{j \in C_n}$, $\Sigma_{h \in C_n}$ and $\Sigma_{k=1}^K$ will be simply denoted by Σ_n , Σ_i, Σ_j, Σ_h and Σ_k respectively.

Let the column vector of parameters be $\Pi = [(\beta_1, \beta_2, \ldots, \beta_K), (\lambda_1, \lambda_2, \ldots, \lambda_K)]'$. The first derivatives of L with respect to the elements of Π, say β_m and λ_m, are computed as follows:

$$\begin{aligned}
\frac{\partial L}{\partial \beta_m} &= \sum_n \sum_i \frac{d_{in}}{P_n(i) \Sigma_j \exp(V_{jn})} \left[\exp(V_{in}) X_{inm}^{(\lambda_m)} - P_n(i) \sum_h \exp(V_{hn}) X_{hnm}^{(\lambda_m)} \right] \\
&= \sum_n \left[\sum_i d_{in} X_{inm}^{(\lambda_m)} - \sum_i d_{in} \sum_h P_n(h) X_{hnm}^{(\lambda_m)} \right] \\
&= \sum_n \left[\sum_i d_{in} X_{inm}^{(\lambda_m)} - \sum_h P_n(h) X_{hnm}^{(\lambda_m)} \right] \text{ since } \Sigma_i d_{in} = 1 \text{ for every } n \\
&= \sum_n \sum_i [d_{in} - P_n(i)] X_{inm}^{(\lambda_m)}
\end{aligned} \tag{10}$$

and similarly

$$\frac{\partial L}{\partial \lambda_m} = \sum_n \sum_k \beta_k \sum_i [d_{in} - P_n(i)] \frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_m} \tag{11}$$

where

$$\frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_m} = \begin{cases} 0 & \text{if } \lambda_k \neq \lambda_m, \\ \dfrac{1}{\lambda_k}\Big[X_{ink}^{\lambda_k} \ln X_{ink} - X_{ink}^{(\lambda_k)}\Big] & \text{if } \lambda_k = \lambda_m \neq 0, \\ \dfrac{1}{2}\ln^2 X_{ink} & \text{if } \lambda_k = \lambda_m = 0. \end{cases} \tag{12}$$

At the maximum point of the log-likelihood function L, the covariance matrix for the estimated parameters $\widehat{\Pi}$ is evaluated as:

$$\text{Var}\Big(\widehat{\Pi}\Big) = \left[-\frac{\partial^2 L}{\partial \Pi \partial \Pi'}\right]^{-1}_{\Pi=\widehat{\Pi}} \tag{13}$$

where

$$\frac{\partial^2 L}{\partial \beta_r \partial \beta_m} = -\sum_n \sum_i P_n(i)\left[X_{inr}^{(\lambda_r)} - \sum_j X_{jnr}^{(\lambda_r)} P_n(j)\right]\left[X_{inm}^{(\lambda_m)} - \sum_j X_{jnm}^{(\lambda_m)} P_n(j)\right], \tag{14}$$

$$\frac{\partial^2 L}{\partial \lambda_r \partial \beta_m} = -\sum_n \sum_i \left\{X_{inm}^{(\lambda_m)} \frac{\partial P_n(i)}{\partial \lambda_r} - [d_{in} - P_n(i)]\frac{\partial X_{inm}^{(\lambda_m)}}{\partial \lambda_r}\right\} \tag{15}$$

and

$$\frac{\partial^2 L}{\partial \lambda_r \partial \lambda_m} = -\sum_n \sum_k \beta_k \sum_i \left\{\frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_m}\frac{\partial P_n(i)}{\partial \lambda_r} - [d_{in} - P_n(i)]\frac{\partial^2 X_{ink}^{(\lambda_k)}}{\partial \lambda_r \partial \lambda_m}\right\}. \tag{16}$$

The derivative $\partial P_n(i)/\partial \lambda_r$ in (15) and (16) is computed as:

$$\frac{\partial P_n(i)}{\partial \lambda_r} = P_n(i) \sum_k \beta_k \left[\frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_r} - \sum_j P_n(j)\frac{\partial X_{jnk}^{(\lambda_k)}}{\partial \lambda_r}\right]. \tag{17}$$

The derivatives $\partial X_{inm}^{(\lambda_m)}/\partial \lambda_r$ in (15), $\partial X_{ink}^{(\lambda_k)}/\partial \lambda_m$ in (16), $\partial X_{ink}^{(\lambda_k)}/\partial \lambda_r$ and $\partial X_{jnk}^{(\lambda_k)}/\partial \lambda_r$ in (17) are computed as in (12).

Finally, the second derivative $\partial^2 X_{ink}^{(\lambda_k)}/\partial \lambda_r \partial \lambda_m$ in (16) is computed as:

$$\frac{\partial^2 X_{ink}^{(\lambda_k)}}{\partial \lambda_r \partial \lambda_m} = \begin{cases} 0 & \text{if } \lambda_k \neq \lambda_r \text{ or } \lambda_k \neq \lambda_m, \\ \dfrac{1}{\lambda_k}\left[X_{ink}^{\lambda_k} \ln^2 X_{ink} - 2\dfrac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_k}\right] & \text{if } \lambda_k = \lambda_r = \lambda_m \neq 0, \\ \dfrac{1}{3}\ln^3 X_{ink} & \text{if } \lambda_k = \lambda_r = \lambda_m = 0 \end{cases} \tag{18}$$

where

$$\frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_k} = \frac{1}{\lambda_k}\Big[X_{ink}^{\lambda_k} \ln X_{ink} - X_{ink}^{(\lambda_k)}\Big]. \tag{19}$$

1.3 Model Specification

The specification of a particular logit model presents some difficulties which the uninitiated user will frequently encounter. Since the estimation procedure fails to converge when the model is not correctly specified, some guidelines will be provided here to avoid the most common errors.

In transportation demand analysis three categories of independent variables are currently specified in the representative utility components V_{jn}'s:

i) the constant terms which have a value of one in each alternative where they appear and zero elsewhere;

ii) the socioeconomic variables which indicate the social and economic attributes of the individuals independently of the mode choice and therefore vary only across the individuals but not across the alternatives, e.g. the individual's income, the education level, the number of children in the family;

iii) the network variables which represent the characteristics of the transportation modes and which vary both across the individuals and the alternatives, e.g. the access time, the waiting time, the in-vehicle time, the travel cost associated with each mode.

An independent variable is called alternative-specific if it appears in one alternative only, or in some or all alternatives with distinct β-coefficients, otherwise it is termed generic if it is present in some or all alternatives with one common β-coefficient.

Three model types can be estimated in the program depending on the specification of the independent variables across the alternatives and the use of the Box-Cox transformations on the strictly positive variables:

1. The **Standard Linear MNL Model Type** where the term "Linear" indicates that all the independent variables enter linearly without any Box-Cox transformation involved and the term "Standard" refers to the case where a socioeconomic variable may appear in $M-1$ alternatives at most and each representative utility component V_{in} includes only the characteristics of the i-th transportation mode, e.g. the travel time or cost of an i-th mode are specified only in the associated V_{in}, but not in the other V_{jn}'s ($j \neq i$). Furthermore, if all the socioeconomic and network variables are specified as generic, the model is also called the Classical Linear MNL Model.
2. The **Standard Box-Cox MNL Model Type** which is analogous to the Standard Linear MNL Model in terms of the specification of the independent variables, but with the fixed or estimated Box-Cox transformations on the strictly positive variables. Note that this model includes the Standard Linear MNL Model as a special case if every λ-parameter is fixed at the value of one.
3. The **Generalized Box-Cox MNL Model Type** which allows a more general specification of the independent variables than the Standard Box-Cox MNL Model in the sense that the characteristics of a particular mode as well as the socioeconomic variables can appear in all the alternatives provided that distinct Box-Cox parameters are specified across the alternatives so that there is no underidentification problem for the β-coefficients. Note that for the network variables, if a time or cost variable of a given mode is specified only in $M-1$ alternatives at most, then distinct Box-Cox parameters are not required across the alternatives since the associated β-coefficients are always identified in this case. The specification of this model as far as the network variables are concerned is very similar to a demand equation system where each equation is a function of all commodity prices in the context of the consumer demand theory. The model also provides a particular estimable form for McFadden's unspecified Universal Logit (1975). The only difficulty in the estimation of this model arises from the much greater number of distinct Box-Cox parameters required if the number of alternatives is large and that several characteristics of a particular mode are jointly specified in all the alternatives.

We will now examine in full detail the specification of the different categories of independent variables in the three model types.

a) Specification of the constant terms

Since the constant terms have a special meaning and that they can never be transformed in the Standard and Generalized Box-Cox MNL Model Types, the same specification is used for the three models: the constant terms should always be specified as alternative-specific and included in $M-1$ alternatives exactly. The choice of the alternative where the constant term is omitted is entirely arbitrary and leaves all the three models invariant in terms of the maximum value of the log-likelihood function, the predicted choice probabilities, the β-coefficients (except for the coefficients of the constant terms) and the λ-parameters.

A small example of the Standard Linear MNL Model Type with a three-mode choice will illustrate the problem. The system of the representative utility components V_{jn}'s ($j = 1, \ldots, 3$) for an individual n who is assumed to have a full choice set $C_n \equiv C$ can be written in matrix form as:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} X_{1n1} & 0 & 0 & X_{1n4} & 0 & 0 \\ 0 & X_{2n2} & 0 & 0 & X_{2n5} & 0 \\ 0 & 0 & X_{3n3} & 0 & 0 & X_{3n6} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{bmatrix} \tag{20}$$

where X_{1n1}, X_{2n2} and X_{3n3} are the alternative-specific constant terms ($\beta_1 \neq \beta_2 \neq \beta_3$), and X_{1n4}, X_{2n5} and X_{3n6} are the alternative-specific variables ($\beta_4 \neq \beta_5 \neq \beta_6$), say the travel times associated with the three modes respectively.

Since the constant terms X_{1n1}, X_{2n2} and X_{3n3} are all equal to one for each individual n, the choice probability $P_n(1)$ can be expressed without explicitly writing them as follows:

$$P_n(1) = \frac{\exp(V_{1n})}{\sum_{j=1}^{3} \exp(V_{jn})} = \frac{\exp(\beta_1 + \beta_4 X_{1n4})}{\exp(\beta_1 + \beta_4 X_{1n4}) + \exp(\beta_2 + \beta_5 X_{2n5}) + \exp(\beta_3 + \beta_6 X_{3n6})} \,. \tag{21}$$

This expression can be represented in three equivalent forms:

$$P_n(1) = \begin{cases} \dfrac{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n4})}{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n4}) + \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2n5}) + \exp(\beta_6 X_{3n6})} \\ \dfrac{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1n4})}{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1n4}) + \exp(\beta_5 X_{2n5}) + \exp(\beta_3 - \beta_2)\exp(\beta_6 X_{3n6})} \\ \dfrac{\exp(\beta_4 X_{1n4})}{\exp(\beta_4 X_{1n4}) + \exp(\beta_2 - \beta_1)\exp(\beta_5 X_{2n5}) + \exp(\beta_3 - \beta_1)\exp(\beta_6 X_{3n6})} \end{cases} . \tag{22}$$

Clearly in any of these three forms, only two differences of the constants, e.g. $\beta_1 - \beta_3$ and $\beta_2 - \beta_3$ in the first form, can be estimated but not separately β_1, β_2 and β_3. The three forms can also be viewed as if the constant term is omitted from the third, second and first alternatives respectively. Moreover, due to the special meaning of the differences of the constants, e.g. $\beta_1 - \beta_3$ and $\beta_2 - \beta_3$ in the first form, which reflect the sample means of the differences of the random utility components $\epsilon_{1n} - \epsilon_{3n}$ and $\epsilon_{2n} - \epsilon_{3n}$ respectively, it is unmeaningful to specify the constant terms in less than $M - 1$ alternatives. Hence the alternative-specific constant terms must be included in $M - 1$ alternatives exactly.

Naturally when the differences of the constants are estimated, they will change according to the form chosen, but the coefficients β_4, β_5 and β_6 as well as the predicted choice probabilities and hence the maximum value of the log-likelihood function will remain invariant. Note that the generic case, e.g.

$\beta_1 - \beta_3 = \ \beta_2 - \beta_3$ in the first form of (22), is irrelevant due to the special meaning of the differences of the constants.

b) Specification of the socioeconomic variables

1) **Standard Linear MNL Model Type**

The specification of the socioeconomic variables is similar to that of the constant terms, except that they can be included in $M - 1$ alternatives at most instead of $M - 1$ alternatives exactly since they do not have any special meaning like the constant terms.

Alternative-specific case

Using the first form of (22) as an example in which X_{1n4}, X_{2n5} and X_{3n6} represent the same socioeconomic variable ($X_{1n4} = X_{2n5}$=X_{3n6}) which is specified as alternative-specific ($\beta_4 \neq \beta_5 \neq \beta_6$), $P_n(1)$ can be rewritten in three equivalent forms:

$$P_n(1) = \begin{cases} \dfrac{\exp(\beta_1^*)\exp[(\beta_4-\beta_6)X_{1n4}]}{\exp(\beta_1^*)[(\beta_4-\beta_6)X_{1n4}] + \exp(\beta_2^*)[(\beta_5-\beta_6)X_{1n4}] + 1} \\ \dfrac{\exp(\beta_1^*)[(\beta_4-\beta_5)X_{1n4}]}{\exp(\beta_1^*)[(\beta_4-\beta_5)X_{1n4}] + \exp(\beta_2^*)\exp[(\beta_6-\beta_5)X_{1n4}]} \\ \dfrac{\exp(\beta_1^*)}{\exp(\beta_1^*) + \exp(\beta_2^*)[(\beta_5-\beta_4)X_{1n4}] + \exp[(\beta_6-\beta_4)X_{1n4}]} \end{cases} \tag{23}$$

where $\beta_1^* = \beta_1 - \beta_3$ and $\beta_2^* = \beta_2 - \beta_3$. From now on, we will use this notation for the differences of the constants in the first form of (22).

Similarly the second and third forms of (22) can each be expressed in three equivalent forms. Like in the case of the constant terms, the three forms of (23) can be viewed as if the socioeconomic variable is omitted in the third, second and first alternatives respectively. Hence the variable can be included in $M - 1$ alternatives at most. If it appears in $M - 1$ alternatives exactly like in (23), then each resulting coefficient of the variable is estimated as a difference between the coefficient in the alternative where the variable is included and the one in the alternative where the variable is omitted.

Generic case

Using the first form of (22) as an example in which X_{1n4}, X_{2n5} and X_{3n6} represent the same socioeconomic variable ($X_{1n4} = X_{2n5}$=X_{3n6}) but specified as generic ($\beta_4 = \beta_5 = \beta_6$), $P_n(1)$ can be rewritten as:

$$P_n(1) = \frac{\exp(\beta_1^*)\exp(\beta_4 X_{1n4})}{\exp(\beta_1^*)\exp(\beta_4 X_{1n4}) + \exp(\beta_2^*)\exp(\beta_4 X_{1n4}) + \exp(\beta_4 X_{1n4})} \tag{24}$$

wich reduces to:

$$P_n(1) = \frac{\exp(\beta_1^*)}{\exp(\beta_1^*) + \exp(\beta_2^*) + 1}. \tag{25}$$

It can be seen that the socioeconomic variable has no influence at all on the choice probability $P_n(1)$ if it is included in all M alternatives. Hence the variable must be specified in $M - 1$ alternatives at most.

2) **Standard Box-Cox MNL Model Type**

The specification of the socioeconomic variables in this model is similar to the one in the Standard Linear MNL Model Type, except that the variables, if they are strictly positive, are subject to the Box-Cox transformations.

Alternative-specific case

Rewriting for example the first form of (23) in which Box-Cox transformations with distinct parameters λ_4 and λ_5 are applied to the same socioeconomic variable which is specified as alternative-specific in the first and second alternatives respectively, the choice probability $P_n(1)$can be expressed as:

$$P_n(1) = \frac{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1n4}^{(\lambda_4)}\right]}{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1n4}^{(\lambda_4)}\right] + \exp(\beta_2^*)\exp\left[\beta_5^* X_{1n4}^{(\lambda_5)}\right] + 1} \tag{26}$$

where $\beta_4^* = \beta_4 - \beta_6$ and $\beta_5^* = \beta_5 - \beta_6$.

Clearly the coefficients β_4^* and β_5^* as well as the Box-Cox parameters λ_4 and λ_5 can all be estimated. Moreover, the two λ-parameters can be set equal without creating any identification problem for β_4^* and β_5^*.

Generic case

In (26), if the socioeconomic variable is specified as generic (β_4^*=β_5^*), then the program constrains the λ-parameters to be equal across the alternatives ($\lambda_4 = \lambda_5$), since it is not meaningful to transform a generic variable with two distinct λ-parameters.

3) **Generalized Box-Cox MNL Model Type**

Unlike the Standard Box-Cox MNL Model Type, the Generalized Box-Cox MNL Model Type allows the same socioeconomic variable to appear across all the alternatives provided that distinct Box-Cox transformations are specified across the alternatives.

Alternative-specific case

Using the first form of (22) as an example in which Box-Cox transformations with distinct parameters λ_4, λ_5 and λ_6 are applied to the alternative-specific variables X_{1n4}, X_{2n5} and X_{3n6} which represent the same socioeconomic variable ($X_{1n4} = X_{2n5} = X_{3n6}$), the choice probability $P_n(1)$ can be rewritten as:

$$\frac{\exp(\beta_1^*)\exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right]}{\exp(\beta_1^*)\exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right] + \exp(\beta_2^*)\exp\left[\beta_5 X_{1n4}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{1n4}^{(\lambda_6)}\right]}. \tag{27}$$

Provided that all the λ-parameters are distinct, there will be no identification problem for β_4, β_5 and β_6.

Generic case

In (27), if the socioeconomic variable is specified as generic ($\beta_4 = \beta_5 = \beta_6$), the common β-coefficient is identified only if distinct λ-parameters are specified across the alternatives. Since it is unmeaningful to use different Box-Cox transformations for a generic variable, this case is not allowed in the program.

c) Specification of the network variables

1) **Standard Linear MNL Model Type**

Unlike the constant terms and the socioeconomic variables, the network variables vary across the individuals but also across the alternatives. Thus they can be included in all alternatives such that a given network characteristic of the i-th mode is specified only in its own representative utility component V_{in} but not in the other V_{jn}'s ($i \neq j$).

Alternative-specific case

Using any of the three equivalent forms of (22) as an example in which the variables X_{1n4}, X_{2n5} and X_{3n6} represent the travel times associated with the three modes respectively ($X_{1n4} \neq X_{2n5} \neq X_{3n6}$) and specified as alternative-specific ($\beta_4 \neq \beta_5 \neq \beta_6$), it can be seen that the coefficients β_4, β_5 and β_6 can separately be estimated.

Generic case

The same is true if the variables are specified as generic ($\beta_4 = \beta_5 = \beta_6$).

2) **Standard Box-Cox MNL Model Type**

The specification of the network variables in this model type is similar to the one in the Standard Linear MNL Model Type, except that the variables X_{1n4}, X_{2n5} and X_{3n6}, if they are strictly positive, are subject to the Box-Cox transformations.

Alternative-specific case

Rewriting for example the first form of (22) in which Box-Cox transformations with distinct parameters λ_4, λ_5 and λ_6 are applied to the alternative-specific variables X_{1n4}, X_{2n5} and X_{3n6} representing the travel times associated with the three modes respectively ($X_{1n4} \neq X_{2n5} \neq X_{3n6}$), the choice probability $P_n(1)$ can be expressed as:

$$P_n(1) = \frac{\exp(\beta_1^*) \exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right]}{\exp(\beta_1^*) \exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right] + \exp(\beta_2^*) \exp\left[\beta_5 X_{2n5}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{3n6}^{(\lambda_6)}\right]}. \tag{28}$$

Clearly the coefficients β_4, β_5 and β_6 as well as the Box-Cox parameters λ_4, λ_5 and λ_6 can all be estimated. Moreover, equality constraints on some or all λ-parameters can be imposed without causing any identification problem for β_4, β_5 and β_6.

Generic case

In (28), if the variables X_{1n4}, X_{2n5} and X_{3n6} are specified as generic ($\beta_4 = \beta_5 = \beta_6$) with distinct λ-parameters or equality constraints on some or all λ-parameters, the common β-coefficient always remains identified. Since it is not meaningful to use different Box-Cox transformations for a generic variable, only the case with one common λ-parameter specified across the alternatives is allowed in the program.

3) **Generalized Box-Cox MNL Model Type**

Unlike the two other models, the Generalized Box-Cox MNL Model Type allows that a given network characteristic of an ith mode be specified not only in its own representative utility component V_{in} but also in some or all of the other V_{jn}'s ($i \neq j$).

Alternative-specific case

The alternative-specific case only refers to the case $\beta_4 \neq \beta_5 \neq \beta_6$ already considered in the Standard Linear and Box-Cox MNL Model Types. Two subcases must be distinguished for the specification of the λ-parameters associated with the network characteristic of an ith mode specified across the alternatives:

• Subcase I : if the network characteristic appears in $M-1$ V_{jn}'s at most including its own, this subcase is analogous to the alternative-specific case (26) of a socioeconomic variable in the Standard Box-Cox MNL Model Type, i.e. one or more λ-parameters can be specified across the alternatives.

• Subcase II : if the network characteristic appears in all V_{jn}'s, this subcase is similar to the alternative-specific case (27) of a socioeconomic variable in the Generalized Box-Cox MNL Model Type, i.e. only distinct λ-parameters should be specified across the alternatives so that the β-coefficients associated with the network characteristic across the alternatives are identified.

Furthermore, each subcase has its own alternative-specific and generic specifications, i.e. the β-coefficients associated with the characteristic of a same mode specified across the alternatives can all be distinct or equal respectively:

• Alternative-specific Subcase I

Modifying the alternative-specific case (28) in the Standard Box-Cox MNL Model Type to illustrate the subcase I, the different V_{jn}'s involving in the choice probability $P_n(1)$ can be represented as follows:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1n4}^{(\lambda_4)} & 0 & 0 & 0 & X_{2n5}^{(\lambda_8)} & X_{3n6}^{(\lambda_9)} \\ 0 & 1 & 0 & X_{2n5}^{(\lambda_5)} & 0 & X_{1n4}^{(\lambda_7)} & 0 & 0 \\ 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_6)} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_9 \end{bmatrix} \tag{29}$$

where the characteristic of an ith mode appears in its own mode and one of the other modes, i.e. X_{1n4} is specified in modes 1 and 2, X_{2n5} in modes 2 and 1, and X_{3n6} in modes 3 and 1. Other specifications of the variables across the alternatives are possible provided that the characteristic of an ith mode appears in $M-1$ alternatives at most including its own.

The subcase I illustrated above corresponds to the "alternative-specific subcase":

- for X_{1n4} : $\beta_4 \neq \beta_7$ and $\lambda_4 \neq \lambda_7$,
- for X_{2n5} : $\beta_5 \neq \beta_8$ and $\lambda_5 \neq \lambda_8$,
- for X_{3n6} : $\beta_6 \neq \beta_9$ and $\lambda_6 \neq \lambda_9$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

Since it is difficult to estimate this subcase with all distinct λ-parameters, different equality constraints can be imposed on these parameters. Here are some examples:

- all the λ-parameters may be set equal.
- $\lambda_4 = \lambda_8 = \lambda_9, \lambda_5 = \lambda_7$ and distinct λ_6.
- $\lambda_4 = \lambda_5 = \lambda_6$ and distinct λ_7, λ_8 and λ_9.

• Generic Subcase I

The "generic subcase" can also be specified as follows:

- for X_{1n4} : $\beta_4 = \beta_7$ and $\lambda_4 = \lambda_7$,
- for X_{2n5} : $\beta_5 = \beta_8$ and $\lambda_5 = \lambda_8$,
- for X_{3n6} : $\beta_6 = \beta_9$ and $\lambda_6 = \lambda_9$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

• Alternative-Specific Subcase II

This subcase is in fact a generalization of the subcase I since each representative utility component V_{jn} now includes the network characteristics of all the modes with distinct β-coefficients and λ-parameters in each alternative and also across all the alternatives:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1n4}^{(\lambda_4)} & 0 & 0 & 0 & 0 & X_{2n5}^{(\lambda_9)} & 0 & X_{3n6}^{(\lambda_{11})} & 0 \\ 0 & 1 & 0 & X_{2n5}^{(\lambda_5)} & 0 & X_{1n4}^{(\lambda_7)} & 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_{12})} \\ 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_6)} & 0 & X_{1n4}^{(\lambda_8)} & 0 & X_{2n5}^{(\lambda_{10})} & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_{12} \end{bmatrix} \tag{30}$$

This subcase which corresponds to the “alternative-specific subcase” is difficult to estimate in practice due to the great number of parameters involved:

- for X_{1n4} : $\beta_4 \neq \beta_7 \neq \beta_8$ and $\lambda_4 \neq \lambda_7 \neq \lambda_8$,
- for X_{2n5} : $\beta_5 \neq \beta_9 \neq \beta_{10}$ and $\lambda_5 \neq \lambda_9 \neq \lambda_{10}$,
- for X_{3n6} : $\beta_6 \neq \beta_{11} \neq \beta_{12}$ and $\lambda_6 \neq \lambda_{11} \neq \lambda_{12}$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

One way to reduce this number is to impose equality constraints on the λ-parameters associated with each alternative:

- for V_{1n} : $\lambda_4 = \lambda_9 = \lambda_{11}$,
- for V_{2n} : $\lambda_5 = \lambda_7 = \lambda_{12}$,
- for V_{3n} : $\lambda_6 = \lambda_8 = \lambda_{10}$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

- Generic Subcase II

The “generic subcase” which considerably reduces the number of parameters is unfortunately not identified since the characteristic of a same mode, e.g. X_{1n4} with one common λ-parameter ($\lambda_4 = \lambda_7 = \lambda_8$), cannot be specified across all the alternatives:

- for X_{1n4} : $\beta_4 = \beta_7 = \beta_8$ and $\lambda_4 = \lambda_7 = \lambda_8$,
- for X_{2n5} : $\beta_5 = \beta_9 = \beta_{10}$ and $\lambda_5 = \lambda_9 = \lambda_{10}$,
- for X_{3n6} : $\beta_6 = \beta_{11} = \beta_{12}$ and $\lambda_6 = \lambda_{11} = \lambda_{12}$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

Generic case

This case can also be analyzed in terms of the two subcases.

- Subcase I

This subcase is similar to (29) except that $\beta_4 = \beta_5 = \beta_6$ and $\lambda_4 = \lambda_5 = \lambda_6$. All the β-coefficients and λ-parameters remain identified for both the “alternative-specific and generic subcases”, but the “generic subcase” is not allowed in the program since it leads to the estimation of only one β-coefficient and one λ-parameter common to all the variables X_{1n4}, X_{2n5} and X_{3n6} appearing in the system: $\beta_4 = \beta_5 = \cdots = \beta_9$ and $\lambda_4 = \lambda_5 = \cdots = \lambda_9$.

- Subcase II

This subcase is also similar to (30) except that $\beta_4 = \beta_5 = \beta_6$ and $\lambda_4 = \lambda_5 = \lambda_6$. Only the “alternative-specific subcase” is identified but not the “generic subcase”.

The main specifications of the different types of variables for the three model types is summarized in Table 1.

TABLE 1 Specifications of the independent variables for the three model types.

TYPE OF VARIABLE	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
CONSTANT	always in $M-1$ alternatives exactly.	not allowed.
SOCIOECONOMIC		
Standard Linear	in $M-1$ alternatives at most.	same as alternative-specific case
Standard Box-Cox	in $M-1$ alternatives at most with one or more λ-parameters specified across the alternatives.	same as the alternative-specific case, but with one common λ-parameter specified across the alternatives.
Generalized Box-Cox	in all alternatives with distinct λ-parameters specified across the alternatives.	not allowed.
NETWORK		
Standard Linear	the characteristic of an ith mode appears only in its own V_{in}, but not in some or all of the other V_{jn}'s $(i \neq j)$.	same as the alternative-specific case.
Standard Box-Cox	same as above with one or more λ-parameters specified across the alternatives.	same as the alternative-specific case, but with one common λ-parameter specified across the alternatives.
Generalized Box-Cox	the characteristics of an ith mode appears not only in its own V_{in} but also in the other V_{jn}'s: • Subcase I (in some V_{jn}'s): the alternative-specific and generic subcases are allowed. • Subcase II (in all V_{jn}'s): only the alternative-specific subcase is allowed.	same as the specific-case: • Subcase I: only the alternative-specific subcase is allowed. • Subcase II: same as the alternative-specific subcase.

d) Multicollinearity among the variables

If some of the independent variables transformed or not by Box-Cox transformations are perfectly collinear, the model cannot be estimated. The correlation matrix and the variance-decomposition proportions in Section 2.3 can be used to detect any dependencies among the variables.

If an attribute X_{ink} can only take on a positive value for people who actually choose alternative i when $d_{in} = 1$ and a zero value when $d_{in} = 0$, and if for the portion of the sample corresponding to $d_{in} = 1$, X_{ink} does not have sufficient variability, or is a constant in the extreme case (perfect collinearity with d_{in}), then the associated β-coefficient is unbounded, i.e. must increase indefinitely to let $P_n(i)$ tend towards one.

e) Miscellaneous problems

There still remains other problems which are hard to detect or to handle:

i. Insufficient variability in the data samples which contain alternatives chosen by only a very small group of individuals, although the model is correctly specified: if in a given alternative, the number of parameters associated with the alternative-specific variables is greater than the number of individuals who actually chose that alternative, then it is quite impossible to estimate those parameters.

ii. Numerical underflows and overflows associated with the exponential function which is the most frequently computed function in this type of model, and also with the Box-Cox transformations which involve the logarithmic and power functions, even when the independent variables are scaled.

iii. Unlike the case of a multinomial logit model without any estimated Box-Cox transformations, which has a unique maximum likelihood solution for the β-coefficients conditional on the fixed λ-parameters if any, the model which estimates jointly the β-coefficients and the λ-parameters can admit multiple maxima. So it is advised to proceed prudently by trying different sets of starting values for the parameters and by using the scaling option provided in the program for the independent variables, when one particular solution seems doubtful.

2 ESTIMATION RESULTS

2.1 Elasticity and Value of Time

a) Weighted aggregate elasticity

Using the approach known as the sample enumeration method (Hensher and Johnson 1981) which considers a random sample of the population as “representative” of the entire population, the weighted aggregate elasticity of the aggregate probability of choosing alternative i, $\bar{P}(i)$, with respect to any variable X_{jk} is computed as follows:

$$E_{X_{jk}}^{\bar{P}(i)} = \sum_n P_n(i) E_{X_{jnk}}^{P_n(i)} / \sum_n P_n(i) \tag{31}$$

where $E_{X_{jnk}}^{P_n(i)}$ is the point elasticity of choosing alternative i for individual n with respect to the kth variable which describes the utility of the jth alternative for individual n:

$$E_{X_{jnk}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial X_{jnk}} \frac{X_{jnk}}{P_n(i)}. \tag{32}$$

The weighted aggregate elasticity (31) can be viewed as evaluating (32) for each individual n, then aggregating by weighting each point elasticity by the individual's estimated choice probability.

• Case where the variable appears in one alternative only

In this case, the point elasticity (32) can be expressed as:

$$E_{X_{jnk}}^{P_n(i)} = \beta_k[\delta_{ij} - P_n(j)]X_{jnk}^{\lambda_k} \tag{33}$$

where $\delta_{ij} = 1$, if $i = j$ (direct elasticity) and $\delta_{ij} = 0$, if $i \neq j$ (cross elasticity). Since the cross elasticities do not depend at all on $P_n(i)$, they are uniform, i.e. equal across the alternatives $i \neq j$, which is one of the implications of the **Independence from Irrelevant Alternatives** (**IIA**) property which states that for a particular individual, the ratio of the choice probabilities for any pair of alternatives remains the same irrespective of the total number M of choices considered.

The weighted aggregate elasticity (31) can be written as:

$$E_{X_{jk}}^{\bar{P}(i)} = \beta_k \sum_n P_n(i)[\delta_{ij} - P_n(j)]X_{jnk}^{\lambda_k} / \sum_n P_n(i) \tag{34}$$

which depends on both $P_n(i)$ and $P_n(j)$ for the cross elasticities so that they need not be uniform at the aggregate level (Ben-Akiva and Lerman 1985).

Two cases for the sign of the weighted aggregate elasticity can be distinguished:

i) Case with Box-Cox ($X_{jnk} > 0$): the direct elasticity has the same sign as β_k and the cross elasticity has the opposite sign, since the sum in the numerator of (34) is always positive for the direct elasticity and always negative for the cross elasticity;

ii) Case without Box-Cox ($-\infty < X_{jnk} < +\infty$), i.e. omitting λ_k in the formula (34) for X_{jnk}:

— If X_{jnk} is positive for all individuals in the sample, this subcase is identical to case (i);

— If X_{jnk} is negative for some individuals in the sample, the signs of the direct and cross elasticities will depend not only on the sign of β_k, but also on the individual signs of each X_{jnk} in the sample so that the sum in the numerator of (34) can be positive or negative;

— If X_{jnk} is negative for all individuals in the sample, this subcase is the opposite of case (i).

•Case where the variable appears in several alternatives

This case corresponds to the following specifications:

— A socioeconomic variable which may appear in several alternatives;

— A network characteristic of an ith mode which may be specified not only in its own alternative but also in some or all of the other alternatives with one or more λ-parameters across the alternatives, as allowed in the Generalized Box-Cox MNL Model Type.

The point elasticity (32) can now be expressed as:

$$E_{jnk}^{P_n(i)} = \left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_h P_n(h)\frac{\partial V_{hn}}{\partial X_{jnk}}\right] X_{jnk} \tag{35}$$

where

$$
\frac{\partial V_{hn}}{\partial X_{jnk}} = \begin{cases} 0 & \text{if the variable is not present in alternative } h \neq j, \\ \beta_m X_{hnm}^{\lambda_m - 1} & \text{if the variable is alternative-specific } (m \neq k) \\ & \text{and appears in both alternatives } h \text{ and } j \ (h \neq j), \\ \beta_k X_{hnk}^{\lambda_k - 1} & \text{if the variable is generic and appears in both} \\ & \text{alternatives } h \text{ and } j \ (h \neq j), \\ \beta_k X_{jnk}^{\lambda_k - 1} & \text{if the variable is either generic or alternative-specific } (h = j). \end{cases} \tag{36}
$$

Note that in the explicit expression of this derivative, each V_{hn} is assumed to include only one X_{hnk} if any. In the case where the variable appears as two independent variables with distinct functional forms in the same alternative h, for example “income” and “income squared”, the program will fully take into account these special cases, if the fixed Box-Cox transformation with $\lambda = 2$ is used on the variable “income” to specify the variable “income squared”. The program can also handle the more general case where the variable is allowed to be present across several alternatives in each of which it may appear as two or more independent variables with different functional forms.

The weighted aggregate elasticity (31) can now be written as:

$$
E_{X_{jk}}^{\bar{P}(i)} = \sum_n P_n(i) \left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_h P_n(h) \frac{\partial V_{hn}}{\partial X_{jnk}} \right] X_{jnk} / \sum_n P_n(i) \tag{37}
$$

which depends on $P_n(i)$ as well as the other $P_n(h)$’s for the cross elasticities so that they need not be uniform at the aggregate level.

In general, the signs of the direct and cross elasticities will depend not only on the sign of β_k associated with V_{in}, but also on the signs of the β-coefficients in the other V_{hn}’s as well as on the sign of each X_{jnk} if the variable is negative for some or all individuals in the sample for the case without Box-Cox.

- **Components of the weighted aggregate elasticity**

Two components of the weighted aggregate elasticity (31) are also computed:

1. The weighted aggregate change of the aggregate probability $\bar{P}(i)$ relative to a percent change of X_{jk}:

$$
C_{X_{jk}}^{\bar{P}(i)} = \sum_n P_n(i) \left[\frac{\partial P_n(i)}{\partial X_{jnk}} X_{jnk} \right] / \sum_n P_n(i) \tag{38}
$$

2. The weighted aggregate derivative of the aggregate probability $\bar{P}(i)$ with respect to X_{jk}:

$$
D_{X_{jk}}^{\bar{P}(i)} = \sum_n P_n(i) \left[\frac{\partial P_n(i)}{\partial X_{jnk}} \right] / \sum_n P_n(i), \tag{39}
$$

where $\partial P_n(i)/\partial X_{jnk} = P_n(i)[\partial V_{in}/\partial X_{jnk} - \Sigma_h P_n(h)\partial V_{hn}/\partial X_{jnk}]$.

b) Elasticity at the sample means of the independent variables

The point elasticities corresponding to the two cases (33) and (35) can be computed at the sample means of all the independent variables, $\bar{X}$, which are assumed to be the characteristics of an "average" individual.

To compute the sample means of the independent variables, two cases must be considered:

1. Every individual in the sample has a full choice set: the sample means of the socioeconomic and network variables are computed from the whole sample, since in this case the observations on the network variables are always available for every alternative.
2. Some or all individuals in the sample have not a full choice set: the sample means of the socioeconomic variables are computed as before from the whole sample, since the social and economic attributes of an individual are always available independently of the fact that the individual has only a subset of available alternatives, but the sample means of the network variables can only be computed from the subset of available alternatives of the individual.

• **Case where the variable appears in one alternative only**

$$E^{P_n(i)}_{X_{jnk}}|_{\bar{X}} = \beta_k[\delta_{ij} - P_n(j)]X^{\lambda_k}_{jnk}|_{\bar{X}} \ . \tag{40}$$

Like for the weighted aggregate elasticity, two cases for the sign of the elasticity at the sample means can be distinguished:

i. Case with Box-Cox ($X_{jnk} > 0$): the direct elasticity has the same sign as β_k and the cross elasticity has the opposite sign since the expression in square brackets as well as $X^{\lambda_k}_{jnk}$ evaluated at the sample means are always positive for the direct elasticity, whereas the expression is negative for the cross elasticity;

ii. Case without Box-Cox ($-\infty < X_{jnk} < +\infty$), i.e. omitting λ_k for X_{jnk} in (40):

— If X_{jnk} is positive for all individuals in the sample, then the sample mean $\bar{X}_{jk}$ is positive: this subcase is identical to case (i);

— If X_{jnk} is negative for some individuals in the sample, then the signs of the direct and cross elasticities will depend not only on the sign of β_k, but also on the sign of $\bar{X}_{jk}$ which can be negative or positive;

— If X_{jnk} is negative for all individuals in the sample, then $\bar{X}_{jk}$ is negative and this subcase is the opposite of case (i).

•**Case where the variable appears in several alternatives**

$$E^{P_n(i)}_{X_{jnk}}|_{\bar{X}} = \left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_n P_n(h)\frac{\partial V_{hn}}{\partial X_{jnk}}\right] X_{jnk}|_{\bar{X}} \ . \tag{41}$$

In general, the signs of the direct and cross elasticities will depend not only on the sign of β_k associated with V_{in}, but also on the signs of the β-coefficients in the other V_{hn}'s as well as on the sign of $\bar{X}_{jk}$ if the variable is negative for some or all individuals in the sample for the case without Box-Cox.

•**Components of the elasticity at the sample means: probability points and partial derivative.**

Two components of the elasticity evaluated at the sample means are also given:

1. The change of the probability $P_n(i)$ relative to a percent change of X_{jnk}:

$$C_{X_{jnk}}^{P_n(i)}|_{\bar{X}} = \frac{\partial P_n(i)}{\partial X_{jnk}} X_{jnk}|_{\bar{X}}, \tag{42}$$

2. The derivative of the probability $P_n(i)$ with respect to X_{jnk}:

$$D_{X_{jnk}}^{P_n(i)}|_{\bar{X}} = \frac{\partial P_n(i)}{\partial X_{jnk}}|_{\bar{X}}, \tag{43}$$

c) Elasticities and probability points corrected for a quasi-dummy or a real dummy

The concept of elasticity strictly implies that an independent variable X_{jk} is of continuous type. Since all the independent variables specified in a model may not be necessarily continuous, a classification of the variables into four categories is needed to allow for the correction of the two types of elasticity (weighted aggregate and at the sample means of the independent variables) when the variable is a discrete variable such as a real dummy, or a nonnegative continuous variable which can be dichotomized as having an effect if it is positive and no effect if it is null, and hence can be called for convenience a “quasi-dummy”:

Category 0: Continuous variable strictly positive.

Category 1: Any continuous variable which is not strictly positive.

Category 2: Quasi-dummy, i.e. a continuous variable containing only positive and null values, for example the number of cars divided by the number of driver’s licenses in the family.

Category 3: Real dummy which has only two values: 0 and a positive constant, for example a binary (0–1) variable such as sex, marital status, etc. or a dummy associated with a quasi-dummy if the latter is allowed to be transformed by Box-Cox but only for the positive values of the variable.

The first category is used to regroup all the continuous variables which can be transformed by Box-Cox since they are strictly positive. The second category includes any continuous variable not strictly positive, hence not transformed by Box-Cox. The third category is mainly used to indicate a continuous variable which is nonnegative but for which the two types of elasticity (weighted aggregate and at the sample means) are corrected for the portion of the positive values of the variable, if one is interested in evaluating the impact of the variable when the effect is present only. Strictly speaking, the variable cannot be transformed by Box-Cox, but if a Box-Cox transformation is desired, then only the portion of the positive values of the variable is subject to the transformation. Hence an associated real dummy ,which has a value of zero if the quasi-dummy is null and a value of one if the quasi-dummy is positive, must be created and specified in the same alternative where the quasi-dummy appears. Finally, the last category includes any real dummy, including the dummy associated with a quasi-dummy, for which the two types of elasticity are corrected for the portion of the positive values of the real dummy. Note that all the variables in this category cannot be transformed by Box-Cox.

The elasticity with respect to a quasi-dummy or a real dummy X_{jk} is corrected for the portion of the positive values of the variable as follows:

1.A. Weighted aggregate elasticity:

$$\tilde{E}_{X_{jk}}^{P_n(i)} = E_{X_{jk}}^{\bar{P}(i)} \frac{\Sigma_n P_n(i)}{\Sigma_{n, X_{jnk}>0} P_n(i)}, \tag{44–A}$$

where $E^{\bar{P}(i)}_{X_{jk}}$ is given in (31) and $\Sigma_{n,X_{jnk>0}} P_n(i)$ represents the sum of the probabilities of choosing alternative i for the individuals such that X_{jnk} is positive only.

1.B. Weighted aggregate change in probability point:

$$\tilde{C}^{\bar{P}(i)}_{X_{jk}} = C^{\bar{P}(i)}_{X_{jk}} \frac{\Sigma_n P_n(i)}{\Sigma_{n,X_{jnk>0}} P_n(i)}. \tag{44–B}$$

2.A. Elasticity at the sample means:

$$\tilde{E}^{P_n(i)}_{X_{jnk}}|_{\bar{X}} = E^{P_n(i)}_{X_{jnk}}|_{\bar{X}} \frac{\bar{X}^+_{jk}}{\bar{X}_{jk}}, \tag{45–A}$$

where $E^{P_n(i)}_{X_{jnk}}|_{\bar{X}}$ is given in (39), $\bar{X}_{jk}$ is the sample mean computed from the whole sample for a socioeconomic variable and from the subset of observations available in alternative j for a network variable, and $\bar{X}^+_{jk}$ is the sample mean computed from the positive observations in the whole sample for a socioeconomic variable and from the subset of positive observations available in alternative j for a network variable.

2.B. Change in probability points at the sample means:

$$\tilde{C}^{P_n(i)}_{X_{jnk}}|_{\bar{X}} = C^{P_n(i)}_{X_{jnk}}|_{\bar{X}} \frac{\bar{X}^+_{jk}}{\bar{X}_{jk}}. \tag{45–B}$$

d) Value of time

In consumer demand theory, the marginal rate of substitution (MRS) of a given variable X_{ik} for any other variable X_{ir} associated with the utility component V_i is defined as the implicit tradeoff between X_{ik} and X_{ir} at a constant level of utility V_i:

$$\mathrm{MRS}^{V_i}_{X_{ik},X_{ir}} = \frac{\partial V_i/\partial X_{ir}}{\partial V_i/\partial X_{ik}} \tag{48}$$

where the subscript n referring to the individual has been dropped from V_{in}, X_{ink} and X_{inr} for simplicity.

In transportation demand analysis, one MRS type of interest is computed for the travel cost, say $C_i = X_{ik}$, and one of the different types of time (e.g. waiting time, access time, in-vehicle time), say $T_i = X_{ir}$, by mode i. In this case, the marginal rate of substitution $\mathrm{MRS}^{V_i}_{C_i,T_i}$ is called the **Value of Time** (VOT), which represents the implicit price that the consumer is willing to pay for a reduction of his time by mode i without affecting his level of utility V_i:

$$\mathrm{VOT}^{V_i}_{C_i,T_i} = \frac{\partial V_i/\partial T_i}{\partial V_i/\partial C_i} \tag{49}$$

For notational compactness, $\mathrm{VOT}^{V_i}_{C_i,T_i}$ will be denoted as VOT^i_{ii} where the superscript i is associated with V and the first and second subscripts ii with C and T respectively.

• **Standard Linear MNL Model Type**

Each V_i includes only its own C_i and T_i in linear form:

$$\mathrm{VOT}^i_{ii} = \frac{\beta^i_{Ti}}{\beta^i_{C_i}} \tag{50}$$

where $\beta^i_{T_i}$ and $\beta^i_{C_i}$ are the coefficients associated with T_i and C_i respectively in V_i.

- **Standard Box-Cox MNL Model Type**

Each V_i includes only its own T_i and C_i with a Box-Cox transformation on each variable:

$$\text{VOT}^{\,i}_{ii} = \frac{\beta^i_{T_i} T_i^{\lambda^i_{T_i}-1}}{\beta^i_{C_i} C_i^{\lambda^i_{C_i}-1}} \tag{51}$$

where $\lambda^i_{T_i}$ and $\lambda^i_{C_i}$ are the Box-Cox parameters associated with T_i and C_i respectively in V_i.

- **Generalized Box-Cox MNL Model Type**

In this model, each V_i can include not only its own T_i and C_i but also the other C_j's and T_j's of one or more alternatives $j \neq i$ with a Box-Cox transformation on each variable. To be completely general, we will assume that each V_i includes all the cost and time variables of the M alternatives so that a matrix VOT$^{\,i}$ of M^2 values of time resulting from the pairwise combinations of the cost and time variables can be defined for each V_i:

$$\text{VOT}^{\,i} = \left[\text{VOT}^{\,i}_{jh}\right] = \begin{bmatrix} \text{VOT}^{\,i}_{11} & \text{VOT}^{\,i}_{12} & \cdots & \text{VOT}^{\,i}_{1M} \\ \text{VOT}^{\,i}_{21} & \text{VOT}^{\,i}_{22} & \cdots & \text{VOT}^{\,i}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}^{\,i}_{M1} & \text{VOT}^{\,i}_{M2} & \cdots & \text{VOT}^{\,i}_{MM} \end{bmatrix} \tag{52}$$

where a typical element VOT$^{\,i}_{jh}$ represents the MRS of C_j for T_h at a constant level of utility V_i:

$$\text{VOT}^{\,i}_{jh} = \frac{\partial V_i / \partial T_h}{\partial V_i / \partial C_j} = \frac{\beta^i_{T_h} T_h^{\lambda^i_{T_h}-1}}{\beta^i_{C_j} C_j^{\lambda^i_{C_h}-1}}. \tag{53}$$

These VOT$^{\,i}_{jh}$'s are unique since each VOT$^{\,i}_{jh}$ is equal to VOT$^{\,m}_{jh}$, $\forall m \neq i$. In other words, they are invariant with respect to the utility component chosen V_i. Thus, instead of computing a matrix VOT$^{\,i}$ for each V_i, it is sufficient to calculate for the whole model a unique matrix VOT where each row j associated with the cost variable C_j and the different time variables T_h's is now computed with respect to V_j, i.e. for each V_j, the own cost variable C_j is used to compute the MRS of C_j for all the time variables T_h's appearing in the same alternative j:

$$\text{VOT} = \left[\text{VOT}^{\,j}_{jh}\right] = \begin{bmatrix} \text{VOT}^{\,1}_{11} & \text{VOT}^{\,1}_{12} & \cdots & \text{VOT}^{\,1}_{1M} \\ \text{VOT}^{\,2}_{21} & \text{VOT}^{\,2}_{22} & \cdots & \text{VOT}^{\,2}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}^{\,M}_{M1} & \text{VOT}^{\,M}_{M2} & \cdots & \text{VOT}^{\,M}_{MM} \end{bmatrix} \tag{54}$$

where a typical element is VOT$^{\,j}_{jh} = (\partial V_j / \partial T_h)/(\partial V_j / \partial C_j) = \left(\beta^j_{T_h} T_h^{\lambda^j_{T_h}-1}\right) / \left(\beta^j_{C_j} C_j^{\lambda^j_{C_j}-1}\right)$.

In this form, the diagonal elements of VOT can be called the *own* or *direct* VOT's since only the own cost and time variables of a given alternative are involved, and the off-diagonal elements the *cross* VOT's otherwise. Note that the form also includes as special cases the Standard Linear and Box-Cox MNL model types, where the direct VOT's are defined and already given in (50) and (51) respectively, but the cross VOT's do not exist since for every alternative j, only the own cost and time variables, C_j and T_j, are specified in V_j.

• **Computation of** VOT_{jh}^{j} **in terms of the choice probability** $P(j)$

In the program, for computational convenience, the VOT_{jh}^{j}'s in (52) are evaluated in terms of $P(j)$ instead of V_j, since the derivatives of $P(j)$ with respect to C_j and T_j as well as the other C_h's and T_h's $(h \neq j)$ appearing in V_j are already available when computing the elasticities and their components.

We will show the equivalence of the procedure for the Generalized Box-Cox MNL model type which includes as special cases the Standard Linear and Box-Cox MNL model types.

Let V_j be a function of all the cost and time variables of the M alternatives:

$$V_j = V_j(C, T, etc.) \tag{55}$$

where $C = [C_1, \ldots, C_M]$ and $T = [T_1, \ldots, T_M]$.

For expositional purposes, every individual n in the sample is assumed to have a full choice set so that every individual choice probability $P(j)$ is a function of all the V_m's:

$$P(j) = P_j(V) = \exp(V_j) / \sum_{m=1}^{M} \exp(V_m) \tag{56}$$

where $V = [V_1, \ldots, V_M]$ and the subscript n is omitted from P and V.

Analogous to the matrix VOT in (52), the matrix $\widetilde{\text{VOT}}$ computed in terms of the choice probabilities $P(j)$'s can be defined as $\widetilde{\text{VOT}} = \left[\widetilde{\text{VOT}}_{jh}^{j}\right]$ where a typical element is:

$$\widetilde{\text{VOT}}_{jh}^{j} = \frac{\partial P(j)/\partial T_h}{\partial P(j)/\partial C_j}. \tag{57}$$

Using the chain rule for functions of functions, the derivative of $P(j)$ with respect to T_h in the numerator of (55) can be written as:

$$\begin{aligned} \frac{\partial P(j)}{\partial T_h} &= \frac{\partial P(j)}{\partial V_1}\frac{\partial V_1}{\partial T_h} + \frac{\partial P(j)}{\partial V_2}\frac{\partial V_2}{\partial T_h} + \ldots + \frac{\partial P(j)}{\partial V_M}\frac{\partial V_M}{\partial T_h} = \sum_{i=1}^{M} \frac{\partial P(j)}{\partial V_i}\frac{\partial V_i}{\partial T_h} \\ &= \frac{\partial V_j}{\partial T_h} \sum_{i=1}^{M} \frac{\partial P(j)}{\partial V_i}\frac{\partial V_i}{\partial T_h}\frac{\partial T_h}{\partial V_j} = \frac{\partial V_j}{\partial T_h} \sum_{i=1}^{M} \frac{\partial P(j)}{\partial V_i}\frac{\partial V_i}{\partial V_j}. \end{aligned} \tag{58}$$

Since $\partial V_i / \partial V_j$ is equal to one if $i = j$ and zero if $i \neq j$, the expression above reduces to:

$$\frac{\partial P(j)}{\partial T_h} = \frac{\partial V_j}{\partial T_h}\frac{\partial P(j)}{\partial V_j}. \tag{59}$$

Similarly, the derivative of $P(j)$ with respect to C_j in the denominator of (55) can be shown to be:

$$\frac{\partial P(j)}{\partial C_j} = \frac{\partial V_j}{\partial C_j}\frac{\partial P(j)}{\partial V_j}. \tag{60}$$

Clearly, $\partial P(j)/\partial T_h$ and $\partial P(j)/\partial C_j$ are different from $\partial V_j/\partial T_h$ and $\partial V_j/\partial C_j$ respectively, but only by a same factor $\partial P(j)/\partial V_j$, so that $\widetilde{\text{VOT}}_{jh}^{j}$ is identical to VOT_{jh}^{j} in (52).

For the Standard Linear and Box-Cox models where each V_j is exclusively a function of C_j and T_j, only the diagonal elements $(j = h)$ of $\widetilde{\text{VOT}}$ are defined:

$$\widetilde{\text{VOT}}^j_{jj} = \frac{\partial P(j)/\partial T_j}{\partial P(j)/\partial C_j} = \frac{\partial P(j)/\partial V_j}{\partial P(j)/\partial V_j}\frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \text{ VOT}^j_{jj}. \tag{61}$$

- **Weighted aggregate value of time**

Similar to the weighted aggregate elasticity $E^{\bar{P}(i)}_{X_{jk}}$, the weighted aggregate value of time $\widetilde{\text{VOT}}^{\bar{P}(j)}_{jh}$ is defined as:

$$\widetilde{\text{VOT}}^{\bar{P}(j)}_{jh} = \sum_n P_n(j)\widetilde{\text{VOT}}^j_{jh.n}/\sum_n P_n(j) \tag{62}$$

where $\widetilde{\text{VOT}}^j_{jh.n}$ is the value of time associated with an individual n:

$$\widetilde{\text{VOT}}^j_{jh.n} = \frac{\partial P_n(j)/\partial T_{hn}}{\partial P_n(j)/\partial C_{hn}} = \frac{\beta^j_{T_h} T_{hn}^{\lambda^j_{T_h}-1}}{\beta^j_{C_h} C_{jn}^{\lambda^j_{C_j}-1}} \tag{63}$$

Note that for the Standard Linear model type $\left(j = h \text{ and } \lambda^j_{C_j} = \lambda^j_{T_j} = 1\right)$, every $\widetilde{\text{VOT}}^j_{jj}$ reduces to the ratio $\beta^j_{T_j}/\beta^j_{C_j}$ so that the weighted aggregate $\widetilde{\text{VOT}}^{\bar{P}(j)}_{jj}$ is also identical to this ratio.

- **Value of time at the sample means of the independent variables**

The value of time $\widetilde{\text{VOT}}^j_{jh}$ can be computed at the sample means of all the independent variables $\bar{X}$ as:

$$\widetilde{\text{VOT}}^j_{jh}|_{\bar{X}} = \frac{\partial P(j)/\partial T_h}{\partial P(j)/\partial C_j}|_{\bar{X}} = \frac{\beta^j_{T_h} \bar{T}_h^{\lambda^j_{T_h}-1}}{\beta^j_{C_j} \bar{C}_j^{\lambda^j_{C_j}-1}}. \tag{64}$$

In contrast to the elasticity and its components computed at the sample means for a given independent variable X_{jk}, which depend on the sample means of all the independent variables due to the presence of $P(j)$ if X_{jk} only appears in alternative j as well as the presence of the other choice probabilities if X_{jk} also appears across the alternatives, the value of time at the sample means $\widetilde{\text{VOT}}^j_{jh}|_{\bar{X}}$ associated with two given variables C_j and T_h depends only on:

i) Their respective sample means $\bar{C}_j$ and $\bar{T}_h$ for the Standard $(j = h)$ and Generalized Box-Cox model types $\left(\lambda^j_{C_j}, \lambda^j_{T_h} \neq 1\right)$.

ii) The ratio of their β-coefficients, $\beta^j_{T_h}/\beta^j_{C_j}$, for the Standard Linear model type ($j = h$ and $\lambda^j_{C_j} = \lambda^j_{T_j} = 1$) , in which case the value of time at the sample means $\widetilde{\text{VOT}}^j_{jj}|_{\bar{X}}$ is also identical to the weighted aggregate analog $\widetilde{\text{VOT}}^{\bar{P}(j)}_{jj}$.

2.2 Prediction Success Table

The predictive capability of the model can be measured by a set of prediction tests computed in a prediction success table due to McFadden (1979).

TABLE 2 Prediction success table.

ALTERNATIVES	PREDICTED ALTERNATIVES $(j = 1, \ldots, M)$	OBSERVED COUNT	OBSERVED SHARE
ACTUAL ALTERNATIVES $(i = 1, \ldots, M)$	$N_{11} \ \ldots \ N_{1M}$ $N_{21} \ \ldots \ N_{2M}$ $\vdots \quad \vdots$ $\vdots \quad \vdots$ $N_{M1} \ \ldots \ N_{MM}$	$\Sigma_j N_{1j}$ $\Sigma_j N_{2j}$ $\vdots$ $\vdots$ $\Sigma_j N_{Mj}$	$\Sigma_j N_{1j}/N$ $\Sigma_j N_{2j}/N$ $\vdots$ $\vdots$ $\Sigma_j N_{Mj}/N$
PREDICTED COUNT	$\Sigma_i N_{i1} \ \ldots \ \Sigma_i N_{iM}$	$\Sigma_i \Sigma_j N_{ij} = N$	1
PREDICTED SHARE	$\Sigma_i N_{i1}/N \ \ldots \ \Sigma_i N_{iM}/N$	1	
P.S.P.	$N_{i1}/\Sigma_i N_{i1} \ \ldots \ N_{MM}/\Sigma_i N_{iM}$	OVERALL P.S.P. $\Sigma_j N_{jj}/N$	
P.S.I.	$\frac{N_{11}}{\Sigma_i N_{i1}} - \frac{\Sigma_i N_{i1}}{N} \ \ldots \ \frac{N_{MM}}{\Sigma_i N_{iM}} - \frac{\Sigma_i N_{iM}}{N}$	OVERALL P.S.I. $\Sigma_j \left[\frac{N_{ij}}{N} - \left(\frac{\Sigma_i N_{ij}}{N} \right)^2 \right]$	
P.E.P.S.	$\frac{\Sigma_i N_{i1} - \Sigma_j N_{1j}}{N} \ \ldots \ \frac{\Sigma_i N_{iM} - \Sigma_j N_{Mj}}{N}$		

In the upper left corner of the table, each N_{ij} represents the expected number of individuals who chose alternative i but had been predicted to choose alternative j:

$$N_{ij} = \sum_n d_{in} P_n(j). \tag{65}$$

Each column sum over i of the N_{ij}'s gives the *predicted count* which represents the total number of individuals predicted to choose alternative j in the sample. Similarly, each row sum over j of the N_{ij}'s gives the *observed count* which denotes the total number of individuals observed to select alternative i in the sample. The predicted and observed counts are used to compute respectively the *predicted* and *observed shares*. Note that by definition both the row sum of the predicted shares and the column sum of the observed shares are always equal to one.

The *proportion successfully predicted* (P.S.P.) indicates the proportion of the individuals who actually choose an alternative and who are expected to choose that alternative. Since the P.S.P. for each alternative varies with the corresponding aggregate share, the predictive capability of the model can be measured against the market-share hypothesis by computing the difference between the P.S.P. and the predicted share, which gives the *prediction success index* (P.S.I.) for each alternative.

The overall P.S.I. which is generally non negative can be computed by the sum of each P.S.I. weighted by the corresponding predicted share over $j = 1, \ldots, M$:

$$\sigma = \sum_j \left(\sum_i \frac{N_{ij}}{N} \right) \left(\frac{N_{jj}}{\Sigma_i N_{ij}} - \sum_i \frac{N_{ij}}{N} \right) = \sum_j \left[\frac{N_{jj}}{N} - \left(\sum_i \frac{N_{ij}}{N} \right)^2 \right]. \tag{66}$$

When the overall P.S.P. is equal to one, $\Sigma_j N_{jj}/N = 1$, i.e. $\Sigma_j N_{jj} = N$, so that the model perfectly predicts, the overall P.S.I. has the following maximum value:

$$\max\,(\sigma) = \sum_j \frac{N_{jj}}{N} - \sum_j \left(\sum_i \frac{N_{ij}}{N} \right)^2 = 1 - \sum_j \left(\sum_i \frac{N_{ij}}{N} \right)^2 \tag{67}$$

where $\Sigma_j (\Sigma_i N_{ij}/N)^2$ is the sum of squares of the predicted shares. Note that σ can be normalized so as to have a maximum value of 1 if desired.

Finally, the *proportional error in prediction share* (P.E.P.S.) is computed as the difference between the predicted and observed shares for each alternative.

2.3 Other Useful Statistics

a) Correlation matrix

For each group of independent variables specified in each alternative, a correlation matrix is computed in terms of:

i. The original variables at the beginning of the program,

ii. and the Box-Cox transformed variables at the maximum point of the log-likelihood function, if there is any fixed or estimated Box-Cox transformation involved.

b) Variance-decomposition proportions

To detect the presence of multiple linear dependencies among the original (or Box-Cox transformed) independent variables, the spectral decomposition of $X'X$ is used for each group of independent variables specified in each alternative (Judge et al. 1985). This method is similar to the singular value decomposition of the matrix X (Belsley et al. 1980).

Let X_i be the group of independent variables in alternative i and K_i the associated number of independent variables. The spectral decomposition of $X_i'X_i$ is defined as:

$$X_i'X_i = \sum_{k=1}^{K_i} \gamma_k p_k p_k' \tag{68}$$

where p_k is the $(K_i \times 1)$ eigenvector associated with the kth eigenvalue γ_k of $X_i'X_i$ and the columns of X_i are scaled to unit length but not centered around their sample means, because centering obscures any linear dependency that involves the constant.

Belsley et al. use a set of condition indexes which is a generalization of the concept of the condition number of a matrix to detect the presence of near dependencies among the columns of X_i:

i. Condition number of X_i: $\kappa(X_i) = (\gamma_{\max}/\gamma_{\min})^{\frac{1}{2}}$, where $\gamma_{\max}$ and $\gamma_{\min}$ are respectively the greatest and smallest eigenvalues of the γ_k's. This number measures the sensivity of b to changes in $X_i'X_i$ or $X_i'Y_i$ in linear systems represented by the normal equations $X_i'X_ib = X_i'Y_i$;

ii. Condition indexes: $\eta_k = (\gamma_{\max}/\gamma_k)^{\frac{1}{2}}, k = 1, \ldots, K_i$. Note that if γ_k is equal to $\gamma_{\min}$, then η_k has a maximum value which corresponds to the condition number of X_i: $\eta_{\max} = \kappa(X_i)$.

To determine which variables are involved in each near dependency, a decomposition of the variance of $b_k (k = 1, \ldots, K_i)$ is performed:

$$\text{Var}\ (b_k) = \sigma^2 \sum_{m=1}^{K_i} \left(p_{km}^2/\gamma_m\right). \tag{69}$$

The proportion of Var(b_k) associated with any $\gamma_r (r = 1, \ldots, K_i)$ is then computed:

$$\Pi_{rk} = \frac{p_{kr}^2/\gamma_r}{\Sigma_{m=1}^{K_i}\left(p_{km}^2/\gamma_m\right)} \tag{70}$$

These results can be summarized in a table of variance-decomposition proportions where the elements in each column are reordered according to the increasing values of the γ_r's and the sum of the proportions in each column is equal to one.

TABLE 3 Eigenvalues of $X_i' X_i$, Condition Indexes of X_i and Proportions of Var(b_k).

EIGENVALUE	CONDITION INDEX	VARIANCE-DECOMPOSITION PROPORTIONS			
		Var(b_1)	Var(b_2)	...	Var(b_{K_i})
$\gamma_1(\gamma_{\min})$	$\eta_1 = \kappa(X_i)$	Π_{11}	Π_{12}	...	Π_{1K_i}
γ_2	η_2	Π_{21}	Π_{22}	...	Π_{2K_i}
.	.	.	.	...	.
.	.	.	.	...	.
.	.	.	.	...	.
$\gamma_{K_i}(\gamma_{\max})$	$\eta_{K_i} = 1$	Π_{K_i1}	Π_{K_i2}	...	$\Pi_{K_iK_i}$

The following rules of thumb can be used to detect the presence of near dependencies:

i. High values of the condition indexes ($\eta > 30$) signal the existence of near dependencies, while high associated Π's in excess of 0.5 indicate which variable is involved in the collinear relations;

ii. When a given variable is involved in several relations, its proportions can be individually small across the high η's. In this case, the sum of these proportions which is in excess of 0.5 can also diagnose variable involvement.

c) t-statistics

In the table of results, two types of t-statistics for the estimated β-coefficients are given:

i. **Unconditional** t-statistics computed with the standard errors given by the square roots of the diagonal elements of (13). Since these statistics are *not invariant* to changes of measurement units in X if estimated Box-Cox transformations are involved, another type of t-statistics is computed in the second row for this case. Note that when some λ-parameters are fixed, the statistics are implicitly conditional on their fixed values.

ii. T-statistics **conditional** on the estimated λ-values and the fixed λ-values if any, computed with the standard errors given by the square roots of the diagonal elements of the covariance matrix:

$$\text{Var}\left(\widehat{\beta} \mid \lambda\right) = \left[-\frac{\partial^2 L}{\partial\beta\,\partial\beta'}\right]^{-1}_{\Pi=\widehat{\Pi}}. \tag{71}$$

These statistics are *invariant* to changes of measurement units in X. Note that when there is no estimated Box-Cox transformations, the second row is not given since it is identical to the first row.

For the estimated λ-parameters, two types of t-statistics are also given:

i. **Unconditional** t-statistics computed under H_0: $\lambda_k = 0$;

ii. **Unconditional** t-statistics computed under H_0: $\lambda_k = 1$.

The estimated λ-values and their t-statistics are *invariant* to changes of measurement units in X. Note that if there are also fixed Box-Cox transformations involved, these statistics are implicitly conditional on the fixed λ-values.

d) Other statistics

i. Log-likelihood with constants only (L_c): value of the log-likelihood function when all the β-coefficients are equal to zero, except for the alternative-specific constants which are estimated.

ii. Log-likelihood at convergence (L_m): maximum value of the log-likelihood function evaluated at $\widehat{\Pi}$.

iii. Likelihood ratio test: $-2(L_c - L_m)$ distributed as chi-squared under the null hypothesis that all the β-coefficients (except for alternative-specific constants) are truly zero, with K^* degrees of freedom where K^* is equal to the number of estimated β and λ-parameters in the model.

iv. Sum of squared residuals: $\Sigma_n \Sigma_i \left[d_{in} - \widehat{P}_n(i)\right]^2$, where $\widehat{P}_n(i)$ is the choice probability of the ith alternative for an individual n estimated at $\widehat{\Pi}$.

v. Rho-squared: $1 - (L_m / L_c)$, which is a goodness-of-fit statistic originally proposed by McFadden(1974).

vi. Rho-squared bar: it is the rho-squared adjusted for K^* degrees of freedom. More appropriate than the rho-squared for comparing different models, this statistic can be computed according to three types of measure:

- Akaike information criterion (Akaike 1974): $1 - [(L_m - K^*)/L_c]$;
- Horowitz (1982, 1983): $1 - \left[\left(L_m - \frac{1}{2}K^*\right)/L_c\right]$;
- Hensher and Johnson (1981): $1 - \left(\frac{L_m}{J^* - K^*} / \frac{L_c}{J^*}\right)$, where $J^* = \Sigma_n(J_n - 1)$ and J_n is the number of alternatives available in an individual's choice set C_n.

vii. Percent right: $(N^*/N)100$, where N is the number of individuals in the data sample and N^* is the number of individuals for whom the alternative actually chosen i such that $d_{in} = 1$ is correctly predicted, i.e. when $\widehat{P}_n(i) = \max_{j \in C_n}\left[\widehat{P}_n(j)\right]$. This goodness of fit measure often given in many computer programs for discrete choice models can mask poor goodness of fit and hence is of little use. A more proper procedure for evaluating model performance is to use the log-likelihood value, the rho-squared and the rho-squared bar, or the overall P.S.I. (Prediction Success Index σ, computed in the Prediction Success Table).

3 SPECIAL OPTIONS

3.1 Estimation with Choice-Based Samples

Three main types of sampling process are currently used in the estimation of discrete choice models:

i. Random sampling: the individuals are selected randomly from a home survey conducted within a given geographical area.

ii. Stratified sampling: the population is first divided into subsets based on one or more exogenous variables, then a random sample is drawn at a different rate from each group. This procedure is also called the *exogenous sampling process*.

iii. Choice-based sampling: the population is first classified into subsets based on the mode choices, then a random sample is drawn from each group. Since the mode choice is an endogenous variable, this procedure is also known as the *endogenous sampling process*.

Mixed sampling procedures can be used among these three types. Due to the larger costs of data collection, the household interviews are sometimes updated by small-scale choice-based surveys which are less expensive, when additional information is needed to obtain more precise estimates for the infrequently chosen alternatives already collected in the home surveys. The combined sample is then called an *enriched sample*. A purely choice-based sample can also be supplemented by the addition of a random sample based on the stratification by the exogenous variables. The resulting sample is called a *supplemented sample*.

Manski and McFadden (1981) made a formal analysis of the general stratified sampling process which includes both the exogenous and endogenous sampling processes as special cases. In the class of generalized choice-based samples, there are several estimation methods depending on the choice of estimators and the design of samples (Cosslett 1981). One estimation procedure which is easily implemented in the computer program by modifying the standard log-likelihood function (9) is the **Weighted Exogenous Sampling Maximum Likelihood** (**WESML**) method developed by Manski and Lerman (1977) who showed that the usual maximum likelihood method for random samples would yield inconsistent estimates when applied to choice-based samples. The **WESML** estimator which is consistent and asymptotically normal is obtained by weighting each observation's contribution to the log-likelihood function with known positive constants $w(i) = Q(i)/H(i)$, where $Q(i)$ is the proportion of the population choosing alternative i and $H(i)$ is the analogous proportion for the choice-based sample. Note that in this case, the derivatives (10) and (11) of the log-likelihood function, the weighted aggregate elasticity (31) and the expected number N_{ij} in the prediction success table are also weighted by $w(i)$ for each observation n. For random samples, the weights $w(i)$ are all equal to 1 since $Q(i)$ is identical to $H(i)$.

3.2 Estimation of Joint-Choice Models

The **MNL** model was first used in single-choice cases where the choice set usually consists of various modes (Lisco 1967, Quarmby 1967, Domencich et al. 1968, Lave 1969, Quandt 1970). More recently, it has been extended to analyze discrete choice situations where the set of feasible alternatives includes joint choices which involve a discrete choice (e.g. mode) and an additional discrete choice (e.g. destination) and/or other choices such as trip frequency, automobile ownership, parking location, residential location, workplace, etc. for passenger demand models (Lerman 1976, Westin and Gillen 1978, Swait et al. 1984), or shipment size, firm location, etc. for freight demand models (McFadden and Winston 1981, Chiang et al. 1981). The joint-choice models are more realistic than the basic single-choice models since the former can capture the interactions between the mode-choice decision and the choice of destination and/or the other choices which are all treated as endogenous, whereas the latter are in fact models of mode choice conditional on the choice of destination and/or the other choices which are considered as exogenous.

As an example, we will use a joint mode and destination choice model to illustrate how it can be estimated in the **MNL** context. The estimation of this type of model can be found in the report by Cambridge Systematics Europe (1984).

Assume that the choice set C_n includes as elements (or alternatives) the mode m and destination d combinations for individual n. The total utility of each element $(m, d) \in C_n$ can be expressed as:

$$U_{md} = \widetilde{V}_m + \widetilde{V}_d + \widetilde{V}_{md} + \epsilon_{md} \tag{72}$$

where $\widetilde{V}_m$ (or $\widetilde{V}_d$) denotes the systematic or representative utility component associated with all elements of C_n using mode m (or destination d), $\widetilde{V}_{md}$ is the remaining systematic utility component specific to the combination (m, d) and ϵ_{md} is the unobserved, random utility component. Note that the subscript n referring to the individual is omitted from the total utility and all its components.

If the errors ϵ_{md}'s are independently and identically distributed according to the Type 1 extreme value (or Gumbel) distribution, the joint-choice probability has exactly the **MNL** form (4):

$$P_n(m, d) = \frac{\exp\left(\widetilde{V}_m + \widetilde{V}_d + \widetilde{V}_{md}\right)}{\sum\limits_{(m',d') \in C_n} \exp\left(\widetilde{V}_{m'} + \widetilde{V}_{d'} + \widetilde{V}_{m'd'}\right)} \tag{73}$$

where $\widetilde{V}_m$ includes the variables which vary only across the modes (e.g. a bus mode constant defined as "1" if alternative i contains the bus mode and "0" otherwise), $\widetilde{V}_d$ includes the variables which vary only across the destinations (e.g. a CBD constant defined as "1" if alternative i contains the central business district as its destination and "0" otherwise), and $\widetilde{V}_{md}$ includes the variables which vary across both the mode and destination, such as the waiting time, the in-vehicle time or the fare for a combination (m, d).

3.3 Disaggregate Forecasting Tests

The primary goal of discrete-choice modeling is to develop a forecasting capability both at the disaggregate and aggregate levels. The aggregate forecasting methods which concern the prediction of the expected aggregate behavior of a population, since the forecast of some aggregate demand is needed for planning strategies, are based on different aggregation procedures designed to reduce the aggregation errors created when aggregate forecasts are derived from disaggregate models. Among the five general types of aggregation procedures first studied by Koppelman (1975) and also analyzed in Koppelman and Ben-Akiva (1977), the simplest approach consists in approximating the forecast aggregate choice probability by its disaggregate analog evaluated at the sample means of the explanatory variables, which are assumed to be the characteristics of an "average" individual representative of the entire population, and the most flexible procedure, known as the sample enumeration method, is based on the average of the choice probabilities across the individuals in the sample which is assumed to be representative of the whole population. Since the choice probability is a nonlinear function of an individual's characteristics, the average of this function is not equal to the function computed at the sample means of the variables: the difference between the predictions obtained by the sample enumeration method and the average individual procedure is termed as the aggregation bias which constitutes one source of aggregation errors.

The sample enumeration and the average individual procedures have been applied respectively in the computation of the weighted aggregate elasticities and the point elasticities evaluated at the sample means given in Section 2.1. Since the aggregate elasticities are mainly used for incremental forecasts of the effects due to policy changes and the point elasticities are valid only for small variations about the sample means, an option available in the program allows to compute the disaggregate predictions which are needed in the policy forecasting tests by which the sensitivity of the model can be checked with respect to the significant changes in policy during the model development process. The forecasting accuracy depends on the availability of prior data which are collected on the variables before and after the real changes in policy are observed.

With the parameter estimates obtained from the first sample, i.e. before changes, the predicted choice probability associated with each alternative $i \in C_n$ is computed for each individual n in the second sample:

$$\widehat{P}_n(i) = \exp\left(\widehat{V}_{in}\right) / \sum_j \exp\left(\widehat{V}_{jn}\right) \tag{74}$$

where $\widehat{V}_{jn} = \Sigma_k \widehat{\beta}_k X_{jnk}^{(\widehat{\lambda}_k)}$ and the $\widehat{\beta}_k$'s and $\widehat{\lambda}_k$'s are the parameter values estimated from the first sample.

If the forecasts seem unmeaningful due to model specification or data measurement errors, the model can be respecified with data correction if any, and the forecasting test is reevaluated. The whole process can be repeated several times until a specification of the model which produces satisfactory forecasts is found.

A model's forecasting error can be characterized by different types of measures such as the average forecasting error, the root-mean-square forecasting error and the confidence interval for the forecast. Whereas the first two measures are only overall indexes of the forecasting accuracy, the confidence interval indicates the range in which the true forecast value is likely to be found. Horowitz (1979) has analyzed three methods for estimating the confidence intervals of the choice probabilities in an **MNL** model involving just the β-coefficients. Since the choice probabilities are nonlinear functions of the coefficients, only asymptotic confidence regions can be derived. The first method which is based directly on the asymptotic sampling distribution of the choice probabilities does not produce a joint rectangular confidence region for these probabilities and therefore is useful mainly for testing hypotheses about them. Moreover, it cannot be extended to functions of choice probabilities such as aggregate market shares and changes in choice probabilities due to changes in explanatory variables. In contrast, both other methods, namely the linear approximation method and the nonlinear programming method, lead to joint rectangular confidence regions and can be extended to functions of choice probabilities. Although the linear approximation method can yield erroneous results, it is generally preferred to the nonlinear programming method due to its computational simplicity. If the linear approximation is accurate, the associated confidence region is smaller than the one obtained with the other method, unless the choice set either is very large or contains only two alternatives.

Due to the above considerations, in the program, the confidence intervals for the choice probabilities are computed by the linear approximation method in which each predicted choice probability (72) is linearized by a first order Taylor-series expansion about $\widehat{\Pi}$:

$$\widehat{P}_n(i) = P_n(i) + \left(\widehat{\Pi} - \Pi\right)' \frac{\partial P_n(i)}{\partial \Pi} \tag{75}$$

where $\widehat{\Pi}$ is the column vector containing the β and λ-values estimated from the first sample, and $P_n(i)$ and Π are respectively the true choice probability value and the column vector of the true β and λ-values.

The typical elements of the column vector of derivatives $\partial P_n(i)/\partial\Pi$ in (73) are given by:

$$\frac{\partial P_n(i)}{\partial \beta_m} = P_n(i)\left[X_{inm}^{(\lambda_m)} - \sum_j P_n(j) X_{jnm}^{(\lambda_m)}\right] \tag{76}$$

and

$$\frac{\partial P_n(i)}{\partial \lambda_m} = P_n(i)\sum_k \beta_k \left[\frac{\partial X_{ink}^{(\lambda_m)}}{\partial \lambda_m} - \sum_j P_n(j)\frac{\partial X_{jnk}^{(\lambda_k)}}{\partial \lambda_m}\right] \tag{77}$$

where the derivatives $\partial X_{ink}^{(\lambda_k)}/\partial\lambda_m$ and $\partial X_{jnk}^{(\lambda_k)}/\partial\lambda_m$ are computed as in (12). Note that the general formula (75) includes both the cases where the λ_k's are constrained to be equal in the same alternative and/or across the alternatives.

Since $\widehat{\Pi}$ is asymptotically normal with mean Π and covariance matrix Var($\widehat{\Pi}$) as given in (13), $\widehat{P}_n(i)$ in its linearized form (73) is asymptotically normal with mean $P_n(i)$ and variance:

$$\text{Var}\left[\widehat{P}_n(i)\right] = \frac{\partial P_n(i)}{\partial \Pi'}\ \text{Var}\ \left(\widehat{\Pi}\right)\frac{\partial P_n(i)}{\partial \Pi}. \tag{78}$$

The numerical value of $\text{Var}\left[\widehat{P}_n(i)\right]$ can be obtained by substituting $\widehat{\Pi}$ for Π in the derivative $\partial P_n(i)/\partial\Pi$ and it is used to compute an asymptotic $100(1-\epsilon)$ percent confidence interval for $P_n(i)$:

$$\widehat{P}_n(i) - Z_{\epsilon/2}\sqrt{\text{Var}\left[\widehat{P}_n(i)\right]} \le P_n(i) \le \widehat{P}_n(i) + Z_{\epsilon/2}\sqrt{\text{Var}\left[\widehat{P}_n(i)\right]} \tag{79}$$

where $Z_{\epsilon/2}$ is the $(1-\epsilon/2)$ percentile of the standard normal distribution. In the program, a value of 0.05 is chosen for ϵ to give a 95 percent confidence interval for $P_n(i)$ and hence the value of $Z_{0.025}$ is equal to 1.96.

4 BIBLIOGRAPHY FOR THE PROBABILITY CLASS

1. **Akaike, H.** (1973), “Information Theory and an Extension of the Maximum Likelihood Principle", in B.N. Petrov and F. Csaski (eds.), *Second International Symposium on Information Theory*, Akademiai Kaido, Budapest, 267–281.
2. **Amemiya, T.** (1981), “Qualitative Response Models: A Survey", *Journal of Economic Literature*, 19(4), 1483–1536.
3. **Anscombe, F.J. and J.W. Tukey** (1954), “The Criticism of Transformations", Paper presented before the American Statistical Association and Biometric Society, Montreal.
4. **Belsley, D., E. Kuh and R.E. Welsh** (1980), *Regression Diagnostics*, John Wiley, New York.
5. **Ben-Akiva, M. and S.R. Lerman** (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass
6. **Berkson, J.** (1944), “Applications of the Logistic Function to Bioassay", *Journal of the American Statistical Association*, 39, 357–365.
7. **Bock, R.D. and L.V. Jones** (1968), *The Measurement and Prediction of Judgement and Choice*, Holden-Day, San Francisco.
8. **Box, G.P. and D.R. Cox** (1964), “An Analysis of Transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211–243.
9. **Box, G.P. and P.W. Tidwell** (1962), “Transformations of the Independent Variables", *Technometrics*, 4, 531–550.
10. **Cambridge Systematics Europe** (1984), *Estimation and Application of Disaggregate Models of Mode and Destination Choice*, Draft Report prepared for Direction des Etudes Générales, Régie Autonome des Transports Parisiens, Paris.
11. **Chiang, Y.S., P. Roberts and M. Ben-Akiva** (1981), “A Short-Run Freight Demand Model: The Joint Choice of Mode and Shipment Size", Paper presented at the 60th meeting of the Transportation Research Board.
12. **Cosslett, S.R.** (1981), “Efficient Estimation of Discrete Choice Models", in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass.
13. **Cox, D.R.** (1970), *Analysis of Binary Data*, Methuen, London.
14. **Cragg, J.G. and T.E. Lisco** (1968), *PROLO-User Manual*, Department of Economics, University of Chicago (mimeo).
15. **Crittle, F.J. and L.W. Johnson** (1980), *Basic Logit (BLOGIT) – Technical Manual*, ATM no.9, Australian Road Research Board
16. **De Donnea, F.X.** (1971), *The Determinants of Transport Mode Choice in Dutch Cities*, Rotterdam University Press, Rotterdam, The Netherlands.
17. **Domencich, T., G. Kraft and J. Valette** (1968), “Estimation of Urban Passenger Travel Behavior: An Econometric Demand Model", *Highway Research Board Record*, 238, 64–78.
18. **Domencich, T. and D. McFadden** (1975), *Urban Travel Demand: A Behavioral Analysis*, North-Holland, Amsterdam.
19. **Finney, D.J.** (1971), *Probit Analysis*, Third Edition , Cambridge University Press, Cambridge, England.

20. **Fletcher, R. and M.J.D. Powell** (1963), “A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, 6, 163–168.

21. **Gaudry, M.J.I.** (1978), “A Generalized Logit mode choice model”, Publication #98, Centre de Recherche sur les Transports, Université de Montréal.

22. **Gronau, R.** (1970), *The Value of Time in Passenger Transportation: The Demand for Air Travel*, National Bureau of Economic Research, New York.

23. **Hensher, D.A.** (1979), *MLOGPROT User’s Manual*, School of Economic and Financial Studies, Macquarie University, Sidney.

24. **Hensher, D.A. and L.W. Johnson** (1981), *Applied Discrete-Choice Modelling*, Croom Helm, London.

25. **Horowitz, J.L.** (1979), “Confidence Intervals for Choice Probabilities of the Multinomial Logit Model”, *Transportation Research Record*, 728, 23–29.

26. **Horowitz, J.L.** (1982), “An Evaluation of the Usefulness of Two Standard Goodness-of-Fit Indicators for Comparing Non-Nested Random Utility Models",*Transportation Research Record*, 874, 19–25.

27. **Horowitz, J.L.** (1983), “Statistical Comparison of Non-Nested Probabilistic Choice Models", *Transportation Science*, 17, 319–350.

28. **Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.C. Lee** (1985), *The Theory and Practice of Econometrics*, Second Edition, John Wiley, New York.

29. **Koppelman, F.** (1975), *Travel Prediction with Models of Individualistic Choice Behavior*, Ph.D. dissertation, Department of Civil Engineering, MIT, Cambridge, Mass.

30. **Koppelman, F. and M. Ben-Akiva** (1977), “Aggregate Forecasting with Disaggregate Travel Demand Models Using Normally Available Data”, *Proceedings of the World Conference on Transport Research*, Rotterdam, The Netherlands.

31. **Lave, C.A.** (1969), “A Behahioral Approach to Modal Split Forecasting", *Transportation Research*, 3, 463–480.

32. **Lerman, S.R.** (1976), “Location, Housing, Auto Ownership and Mode to Work: A Joint-Choice Model", *Transportation Research Record*, no. 610, 6–11.

33. **Lerman, S.R.** (1982), *Micro Data Analyzer: the MDA User’s Manual*, Release 1.0, Cambridge Systematics, Inc.

34. **Lisco, T.E.** (1967), *The Value of Commuters’ Travel Time: A Study in Urban Transportation*, Ph.D. Dissertation, Department of Economics, University of Chicago.

35. **Luce, R.D. and P. Suppes** (1965), “Preference, Utility and Subjective Probability", in R.D. Luce, R. Bush and E. Galanter (eds.), *Handbook of Mathematical Psychology*, Wiley, New York.

36. **Luce, R.D.** (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.

37. **Manski, C.F.** (1981), “Structural Models of Discrete Data", *Sociological Methodology*, 58–109.

38. **Manski, C.F. and S.R. Lerman** (1977), “The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45, 1977–1988.

39. **Manski, C.F. and D. McFadden** (1981), “Alternative Estimators and Sample Designs for Discrete Choice Analysis", in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass.

40. **Marschak, J.** (1960), “Binary Choice Constraints on Random Utility Indicators", in K. Arrow (ed.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford University Press,

Calif.

41. **McFadden, D.** (1974), “Conditional Logit Analysis of Qualitative Choice Behavior", in P.C. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 105–142.
42. **McFadden, D.** (1975), “On Independence, Structure and Simultaneity in Transportation Demand Analysis”, Working Paper 7511, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.
43. **McFadden, D.** (1979), “Quantitative Methods for Analysis Travel Behaviour of Individuals: Some Recent Developments", in D.A. Hensher and P.R. Stopher (eds.), *Behavioural Travel Modelling*, Croom Helm, London.
44. **McFadden, D.** (1982), “Econometric Analysis of Qualitative Response Models", Working Paper, Department of Economics, MIT, Cambridge, Mass.
45. **McFadden, D., J. Berkman, D. Browstone and G.M. Duncan** (1979), *QUAIL 3.0 User’s Manual*, Department of Economics, University of California, Berkeley.
46. **McFadden, D. and C. Winston** (1981), “Joint Estimation of Discrete and Continuous Choices in Freight Transportation", Paper presented at the 1981 Meeting of the Econometrics Society.
47. **McGillivray, R.G.** (1972), “Binary Choice of Urban Transport Mode in the San Francisco Region", *Econometrica*, 40, 827–848.
48. **Ortúzar, J. de D.** (1986), “The Cultural and Temporal Transferability of Discrete Choice Disaggregate Modal Split Models", Paper presented to the 1986 World Conference on Transport Research, Vancouver, May 1986, Department of Transport Engineering, Pontificia Universidad Católica de Chile, Santiago.
49. **Quandt, R.** (1970), *The Demand for Travel*, D.C. Heath, Lexington, Mass.
50. **Quarmby, D.A.** (1967), “Choice of Travel Mode for the Journey to Work: Some Findings", *Journal of Transportation and Economic Policy*, 1, 273–314.
51. **Spitzer, J.J.** (1978), “A Monte Carlo Investigation of the Box-Cox Transformation in Small Samples", *Journal of the American Statistical Association*, 73, 488–495.
52. **Stopher, P.R.** (1969), “A Probability Model of Travel Mode Choice for the Work Journey", *Highway Research Record*, 283, 57–65.
53. **Swait, J., V. Kozel, R. Barros and M. Ben-Akiva** (1984), “A Model System of Individual Travel Behavior for a Brazilian City", *Transportation Policy and Decision Making*, 2, 451–480.
54. **Talvitie, A.D.** (1972), “Comparison of Probabilistic Modal Choice Models: Estimation Methods and System Inputs", *Highway Research Record*, 392, 111–120.
55. **Thurstone, L.** (1927), “A Law of Comparative Judgement", *Psychological Review*, 34, 273–286.
56. **Tukey, J.W.** (1957), “On the Comparative Anatomy of Transformations", *Annals of Mathematical Statistics*, 28, 602–632.
57. **Tversky, A.** (1972), “Elimination by Aspects: A Theory of Choice", *Psychological Review*, 79, 281–299.
58. **Warner, S.L.** (1962), *Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice*, Northwestern University Press, Evanston, Ill.
59. **Watson, P.L.** (1974), *The Value of Time: Behavioral Models of Modal Choice*, D.C. Heath, Lexington, Mass.
60. **Westin, R. and D. Gillen** (1978), “Parking Location and Transit Demand", *Journal of Econometrics*,

8, 75–101.

61. **Wigner, M.F.** (1973), “Disaggregate Modal Choice Models of Downtown Trips in the Chicago Region", *Highway Research Record*, 446, 49–65.

62. **Zarembka, P.C.** (1974), “Transformation of Variables in Econometrics", in P.C. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 81–104.