The Natural Sciences and Engineering Council of Canada (NSERC) and the *Fonds Québécois de Recherche sur la Nature et les Technologies* (FQRNT) have supported this profoundly modified version of the P-2/P-6 algorithm published with the partial support of the "STEMM" project (Contract No.: ST-96–SC.301 of the Commission of the European Communities) as: Liem, T. C. and M. Gaudry, "PROBABILITY: The P-2 to P-6 programs for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT models with disaggregate data", Publication CRT-97–57, Centre de recherche sur les transports, Université de Montréal, 72 p., 1996 and 1997. Also Working Paper No. 9802, Bureau d'économie théorique et appliquée (BETA), Université Louis Pasteur, 1998. This document is complemented by a user guide for the program: Tran, C-L., and Gaudry, M., "The P-2/P-6 Program user guide with databases, examples and Tablex outputs", Publication AJD-112, Agora Jules Dupuit, Université de Montréal, April 2008.

PUBLICATION
AJD-111

PROBABILITY: The P-2 to P-6 estimation procedures for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT models with disaggregate data

Cong-Liem Tran[1] and Marc Gaudry[2]

[1] Agora Jules Dupuit, Université de Montréal
[2] Département de sciences économiques, Université de Montréal.
(B.P. 6128, Succursale centre-ville, Montréal, Qc, Canada H3C 3J7)

Publication AJD-111, Agora Jules Dupuit
Université de Montréal
April 2008 — Revised January 2009

ABSTRACT

The P-2/P-6 programs are designed for the maximum likelihood estimation of five discrete choice model families which all include the **Multinomial LOGIT** model as a special case, when the basic unit of observation is an individual and the choice between two or more alternatives is by nature a discrete event.

The P-2 program deals with the **LOGIT** family which is consistent with the Independence from Irrelevant Alternatives (IIA) axiom of choice theory (Luce, 1959). The LOGIT response curve is perfectly symmetric with "thin equal" tails, which may imply a systematic underestimation (or overestimation) at the tails of the distribution, if consumers will choose certain alternatives despite their apparent unattractiveness. Various directions have been tried to overcome the limitations of the IIA assumption and the "thin equal tails" property. The P-3/P-6 programs have been developed following these directions. The **S-DOGIT** and **G-DOGIT** families (P-3/P-4 programs) relax the IIA restriction by allowing the relative choice probabilities of some or all pairs of alternatives not to remain constant. Thicker tails can be fitted to the data to capture the specification error, the modeler's ignorance, the compulsive consumption or captivity to an alternative (S-DOGIT) or the similarity/dissimilarity between alternatives (G-DOGIT). Only the S-DOGIT response curve remains symmetric with respect to its inflection point. The **LIN-IPT-LOGIT** and the **BT-IPT-LOGIT** families (P-5/P-6 programs) totally remove the IIA restriction so that the relative choice probabilities of any pair of alternatives cannot remain constant. Asymmetric tails can be obtained to represent the fact that the consumers will react differently in the choice of each alternative due to a certain amount of stickiness or alternative-specificity.

In all five families, different Box-Cox transformations can be applied on variables to introduce flexible functional forms other than the usual linear and logarithmic forms of the independent variables: the Box-Cox parameters can be fixed or estimated jointly with the coefficients of the variables (except for the G-DOGIT family where only fixed Box-Cox parameters can be used). Program outputs include two types of elasticities (elasticity at sample means of the variables and weighted aggregate elasticity), marginal rates of substitution as well as forecasts obtained by simulation.

Keywords: Box-Cox Multinomial Logit, Discrete Choice Models, Disaggregate Data, Symmetric/Asymmetric Tails, Captivity, Similarity/Dissimilarity, Maximum Likelihood Estimation, Elasticities, Marginal Rates of Substitution, Simulated Forecast.

RÉSUMÉ

Les programmes P-2/P-6 sont destinés à l'estimation selon la méthode du maximum de vraisemblance de cinq familles de modèles de choix discret qui incluent toutes le modèle **LOGIT Multinomial** comme cas particulier, lorsque l'unité d'observation de base est un individu et que le choix entre deux ou plusieurs alternatives est par nature un événement discret.

Le programme P-2 traite de la famille du **LOGIT** qui est conforme à l'axiome de l'Indépendance des Alternatives Non Pertinentes (IANP) de la théorie du choix (Luce, 1959). La courbe de distribution du LOGIT est parfaitement symétrique avec des queues "minces et égales", ce qui peut impliquer une sous-estimation (ou sur-estimation) systématique dans les queues de la distribution, lorsque les consommateurs choisissent certaines alternatives malgré leur non-attirance apparente. Diverses directions ont été essayées pour surmonter les limitations de l'hypothèse IANP et de la propriété des queues "minces et égales". Les programmes P-3/P-6 ont été développés selon ces directions. Les familles du **S-DOGIT** et du **G-DOGIT** (programmes P-3/P-4) assouplissent la restriction IANP en permettant aux probabilités de choix relatives de certaines ou de toutes les paires d'alternatives de ne pas rester constantes. Des queues plus épaisses peuvent être ajustées aux données pour capter l'erreur de spécification, l'ignorance du modélisateur, la consommation forcée ou captivité à une alternative (S-DOGIT) ou la similarité/dissimilarité entre les alternatives (G-DOGIT). Seule la courbe de distribution du S-DOGIT demeure symétrique par rapport à son point d'inflexion. Les familles du **LIN-IPT-LOGIT** et du **BT-IPT-LOGIT** (programmes P-5/P-6) ne respectent plus du tout la restriction IANP, de telle sorte que les probabilités de choix relatives de n'importe quelle paire d'alternatives ne peuvent plus demeurer constantes. Des queues assymétriques peuvent être obtenues pour refléter le fait que les consommateurs réagiront différemment dans le choix de chaque alternative à cause d'une certaine inertie ou spécificité de l'alternative.

Dans toutes les cinq familles, différentes transformations de Box-Cox peuvent être appliquées à des variables indépendantes pour introduire des formes fonctionnelles flexibles autres que les formes usuelles, linéaire et logarithmique: les paramètres de type Box-Cox peuvent être fixés ou estimés conjointement avec les coefficients des variables (sauf pour la famille du G-DOGIT où seulement des paramètres fixés peuvent être utilisés). Le logiciel calcule notamment deux types d'élasticités (élasticité à la moyenne des variables et élasticité agrégée-pondérée), des taux marginaux de substitution, ainsi que des prévisions par simulation.

Mots-clés: Logit Multinomial avec Box-Cox, Modèles de Choix Discret, Données Désagrégées, Queues Symétriques/Assymétriques, Captivité, Similarité/Dissimilarité, Estimation du Maximum de Vraisemblance, Élasticités, Taux Marginaux de Substitution, Prévisions par Simulation.

Contents

List of Tables and Figures

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. 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). Existing programs to estimate the coefficients of the independent variables without Box-Cox (1964) transformations include 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).

The probability class considered here includes five discrete choice model families:

1. P-2: **LOGIT** family,
2. P-3: **S**tandard **DOGIT** (**S-DOGIT**) family,
3. P-4: **G**eneralized **DOGIT** (**G-DOGIT**) family,
4. P-5: **LIN**ear **I**nverse **P**ower **T**ransformation **LOGIT** (**LIN-IPT-L**) family,
5. P-6: **B**ox-**T**ukey **I**nverse **P**ower **T**ransformation **LOGIT** (**BT-IPT-L**) family,

which all contain the **MNL** model with the independent variables in linear form as a special case.

Due to its simple formulation, the **LOGIT** family is characterized among other things by the "symmetric-thin-equal-tails" of the probability response curve, leading to the systematic underestimation (or overestimation) of the choice probabilities in the left (or right) tail of the sigmoid curve, and hence cannot be appropriate in two situations frequently encountered:

i. Thicker tails, i.e. for both left and right tails, due to specification error, modeler's ignorance, compulsive consumption or captivity to modes.

ii. Asymmetric tails, i.e. thicker left tail and thinner right tail, due to a certain amount of stickiness or mode-specificity in the consumer's reactions to various modes.

Designed to handle the first case only, the **DOGIT** families were developed by Gaudry and Dagenais (1978, 1979). Only the **S-DOGIT** family was empirically studied by Gaudry and Wills (1979) with aggregate travel data and by Swait and Ben-Akiva (1987) in a discrete choice modeling context. In contrast, the **IPT-L** families were designed to solve both problems: the **LIN-IPT-L** family proposed by Gaudry (1981) was used in the estimation of educational production functions by Montmarquette and Mahseredjian (1985), and the new **BT-IPT-L** family programmed here by the present authors was originally developed for a comparative study (Gaudry 1989) where an intercity travel demand with aggregate Canadian data is estimated with all five families.

To model the flexible functional forms of the independent variables, an option for the Box-Cox (1964) transformations on these variables is also available. The Box-Cox parameters can be fixed or jointly estimated with the other parameters specific to each model family via the maximum likelihood method, except for the **G-DOGIT** family which can include only fixed Box-Cox transformations due to the greater number of parameters already involved.

1 STATISTICAL ESTIMATION

Although all the five discrete choice model families considered here are different in structure but are nonlinear in the parameters, notably due to the presence of the Box-Cox transformations on the independent variables, the common procedure used to estimate these families is the maximum likelihood method.

1.1 Maximum Likelihood Procedure

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:

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

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:

$$P_{in} = \exp\left(V_{in}\right) / \sum_{j \in C_n} \exp\left(V_{jn}\right)\,. \tag{4}$$

To provide a theoretical framework formulated in Gaudry (1989) within which the relationships among the five discrete choice model families can be examined, a general form of the choice probability P_{in} can be written as:

$$P_{in} = \frac{\mathcal{U}_{in}(\Pi)}{\Sigma_m \mathcal{U}_{mn}(\Pi)}\,,\ (i, m \in C_n) \tag{5}$$

where $\mathcal{U}_{in}(\Pi)$ is the overall utility (preference) associated with the ith alternative for an individual n, which depends on a set of parameters Π to be determined later, and P_{in} is continuous over the interval [0,1] and also satisfies the adding-up constraint $\Sigma_m P_{mn} = 1$.

The utility function $\mathcal{U}_{in}(\Pi)$ is assumed to be:

$$\mathcal{U}_{in}(\Pi) = \sum_j w_{ij}\left(\tilde{\theta}_{ij}\right) f_j\left(g_j\left(\exp\left(V_{jn}\right); \tilde{\phi}_j, \tilde{\mu}_j\right); \phi_j, \mu_j\right) \geq 0\ ,\ \ (i, j \in C_n) \tag{6}$$

where

— $w_{ij}\left(\tilde{\theta}_{ij}\right)$ is a weight function of the parameter $\tilde{\theta}_{ij}$ defined in terms of θ_{ij} as: $\tilde{\theta}_{ii} = 1 + \theta_{ii}$ if $i = j$ and $\tilde{\theta}_{ij} = \theta_{ij}$ if $i \neq j$, so that $w_{ij}\left(\tilde{\theta}_{ij}\right)$ can be implicitly considered as a function of θ_{ij};

— $f_j(\cdot)$ is an outer function of the parameters ϕ_j and μ_j applied to an inner function $g_j(\cdot)$;

— $g_j(\cdot)$ is a function of the parameters $\tilde{\phi}_j$ and $\tilde{\mu}_j$ applied to the exponential function of V_{jn};

— and V_{jn} is the representative utility component of mode$j$ specified as a linear combination of the Box-Cox transformations on the independent variables X_{jnk}'s with parameters λ_k's: $V_{jn} = \Sigma_k \beta_k X_{jnk}^{(\lambda_k)}$, $(k = 1, \ldots, K)$.

The column vector of parameters Π of the system includes all the parameters just defined: $\Pi = \left[\beta', \lambda', \theta', \phi', \mu', \tilde{\phi}', \tilde{\mu}'\right]'$, where all the parameters are denoted as column vectors, including θ which is a column-stacked vector of all the θ_{ij}'s : $\theta = [\theta_{11}, \ldots, \theta_{M1}, \theta_{12}, \ldots, \theta_{M2}, \ldots, \theta_{1M}, \ldots, \theta_{MM}]'$.

More explicitly, the two functions $g_j(\cdot)$ and $f_j(\cdot)$ in (6) are respectively the direct and inverse Box-Tukey transformations on a real variable or function Z_j defined as follows:

$$g_j\left(Z_j; \tilde{\phi}_j, \tilde{\mu}_j\right) = Z_j^{\left(\tilde{\phi}_j, \tilde{\mu}_j\right)} = \begin{cases} \frac{1}{\tilde{\phi}_j}\left[\left(Z_j + \tilde{\mu}_j\right)^{\tilde{\phi}_j} - 1\right] & \text{if } \tilde{\phi}_j \neq 0 \\ \ln\left(Z_j + \tilde{\mu}_j\right) & \text{if } \tilde{\phi}_j = 0 \end{cases} \tag{7}$$

where $Z_j = \exp\left(V_{jn}\right)$ and $Z_j + \tilde{\mu}_j > 0$,
and

$$f_j(Z_j; \phi_j, \mu_j) = Z_j^{(\phi_j, \mu_j)^{-1}} = \begin{cases} \left(\phi_j Z_j + 1\right)^{\frac{1}{\phi_j}} - \mu_j & \text{if } \phi_j \neq 0 \\ \exp\left(Z_j\right) - \mu_j & \text{if } \phi_j = 0 \end{cases} \tag{8}$$

where $Z_j = g_j(\cdot)$ and $\phi_j Z_j + 1 > 0$.

The direct Box-Tukey transformation (7) have been studied by Anscombe and Tukey (1954) and referred to in Tukey (1957), Box and Cox (1964), Zarembka (1974) and others: $\tilde{\phi}_j$ is a power parameter and $\tilde{\mu}_j$ is a location parameter used to shift the variable or function Z_j so that the expression $Z_j + \tilde{\mu}_j$ becomes strictly positive before being raised to the power $\tilde{\phi}_j$, since Z_j can be negative in a more general context than when Z_j has the form of $\exp\left(V_{jn}\right)$ considered here. The inverse Box-Tukey transformation (8) which is derived from the direct Box-Tukey transformation has a minus sign before the location parameter μ_j instead of a plus sign as it was originally formulated in Gaudry (1981) for the **LIN-IPT-L** family starting from a form of (Z) that subtracted this shift parameter — whence the abbreviations used by students: BC and BT for the direct forms and BCG and BTG for the inverse forms.

A special case of (7) without the location parameter is known as the Box-Cox transformation, which can be applied to an independent variable X_{jnk} specified in V_{jn} to allow for flexible functional forms including the linear and logarithmic forms traditionally used in regression models. Since the location parameter is not present, the variable X_{jnk} should be strictly positive so that $\left(X_{jnk}^{\lambda_k} - 1\right)/\lambda_k$ and $\ln\left(X_{jnk}\right)$ can be computed, where the power parameter λ_k is used instead of $\tilde{\phi}_j$ to indicate that it is associated with the kth variable in V_{jn}. In the compact notation of $V_{jn} = \Sigma_k \beta_k X_{jnk}^{(\lambda_k)}$, it is clearly

TABLE 1 Relationships among the five discrete choice model families.

Parameter Constraints in the Three Functions			**FAMILY**
Weight Function $w_{ij}\left(\tilde{\theta}_{ij}\right)$	Outer Function $f_j(.)$	Inner Function $g_j(.)$	Function $\mathcal{U}_{in}(\Pi) \geq 0$
$\theta_{ii} = \theta_{jj} = 0$	$\phi_j = \mu_j = 1$	$\tilde{\phi}_j = \tilde{\mu}_j = 1$	**LOGIT** $\exp(V_{in})$
$\theta_{i1} = \ldots = \theta_{iM} = \theta_i \geq 0$	as above	as above	**S-DOGIT** $\exp(V_{in}) + \theta_i \sum_j \exp(V_{jn})$
$\theta_{ii} = 0$ and $\theta_{ij} \geq 0$	as above	as above	**G-DOGIT** $\exp(V_{in}) + \sum_{j \neq i} \theta_{ij} \exp(V_{jn})$
$\theta_{ii} = \theta_{ij} = 0$	$\phi_j \geq 0$ and $\mu_j \leq 1$	$\tilde{\phi}_j = \tilde{\mu}_j = 1$	**LIN-IPT-L** $[\phi_i \exp(V_{in}) + 1]^{\frac{1}{\phi_i}} - \mu_i$
$\theta_{ii} = \theta_{ij} = 0$	$\phi_j = \mu_j = 0$	$\tilde{\mu}_j \geq 0$	**BT-IPT-L** $\exp\left\{\frac{[\exp(V_{in}) + \tilde{\mu}_i]^{\tilde{\phi}_i} - 1}{\tilde{\phi} i}\right\}$

understood that some of the independent variables such as the constant, the dummies and the ordinary variables not strictly positive, cannot be subject to the Box-Cox transformation.

Since the general form of $\mathcal{U}_{in}$ is not identified, different constraints on the "extra parameters", namely the θ, ϕ, μ, $\tilde{\phi}$ and $\tilde{\mu}$-parameters, should be imposed to yield an estimable form of $\mathcal{U}_{in}$ for each model family. Table 1 summarizes the different constraints on these extra parameters. For example, the **LOGIT** family can be obtained by setting every θ_{ii} and θ_{ij} equal to zero, and every ϕ_j, μ_j, $\tilde{\phi}_j$ and $\tilde{\mu}_j$ equal to one. The **S-DOGIT**, **G-DOGIT**, **LIN-IPT-L** and **BT-IPT-L** families all reduce to the **LOGIT** family by setting every $\theta_i = 0, \theta_{ij} = 0, \phi_i = \mu_i = 1$ and $\tilde{\phi}_i = \tilde{\mu}_i = 0$ respectively.

To insure that $\mathcal{U}_{in}$ is nonnegative over the whole range of $V_{in}(-\infty, +\infty)$, additional constraints on the extra parameters should be considered, namely $\theta_i \geq 0$ for **S-DOGIT**, $\theta_{ij} \geq 0$ for **G-DOGIT**, $\phi_i \geq 0$ and $1 - \mu_i \geq 0$ for **LIN-IPT-L**, and $\tilde{\mu}_i \geq 0$ for **BT-IPT-L**.

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:

$$\mathcal{L} = \prod_{n=1}^{N_1} P_{1n} \prod_{n=N_1+1}^{N_1+N_2} P_{2n} \ldots \prod_{n=N_{M-1}+1}^{N} P_{Mn}. \tag{9}$$

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:

$$\mathcal{L} = \prod_{n=1}^{N} \prod_{i \in C_n} P_{in}^{d_{in}}. \tag{10}$$

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

$$L = \ln \mathcal{L} = \sum_{n=1}^{N} \sum_{i \in C_n} d_{in} \ln P_{in}. \tag{11}$$

1.2 Computational Aspects

To maximize the log-likelihood function (11) with respect to the vector of parameters Π, the Davidon-Fletcher-Powell optimization method (Fletcher and Powell 1963) which requires the first derivatives of the function is used.

Let the column vector of parameters be $\Pi = [\beta', \psi', \lambda']'$ for the five model families:

1. β is a $(K \times 1)$ vector of β-coefficients.
2. ψ is a column vector of extra parameters which are present only in the last four families:
   a. **S-DOGIT**: ψ is an $(M \times 1)$ vector of θ_i's.
   b. **G-DOGIT**: ψ is an $[M(M-1) \times 1]$ vector of θ_{ij}'s.
   c. **LIN-IPT-L**: ψ is a $(2M \times 1)$ vector of ϕ_i's and μ_i's.
   d. **BT-IPT-L**: ψ is a $(2M \times 1)$ vector of $\tilde{\phi}_i$'s and $\tilde{\mu}_i$'s.
3. λ is a $(K \times 1)$ vector of Box-Cox parameters which can be estimated in all families, except for **G-DOGIT** which can include only fixed Box-Cox parameters due to the great number of estimated θ_{ij}'s.

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

The first derivatives of L with respect to Π can be written as:

$$\begin{aligned} \frac{\partial L}{\partial \Pi} &= \sum_n \sum_i d_{in} \frac{\partial \ln P_{in}}{\partial \Pi} \\ &= \sum_n \sum_i d_{in} \left(\frac{1}{\mathcal{U}_{in}} \frac{\partial \mathcal{U}_{in}}{\partial \Pi} - \frac{1}{\sum_m \mathcal{U}_{mn}} \sum_m \frac{\partial \mathcal{U}_{mn}}{\partial \Pi} \right) \quad . \end{aligned} \tag{12}$$

The first derivatives of a typical utility function $\mathcal{U}_{mn}$ with respect to the elements of Π are:

LOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial \beta_k} = \exp(V_{mn}) \, X_{mnk}^{(\lambda_k)} \tag{13}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \lambda_k} = \exp(V_{mn}) \, \beta_k \Delta_{mnk} \tag{14}$$

S-DOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial \beta_k} = \exp\left(V_{mn}\right) X_{mnk}^{(\lambda_k)} + \theta_m \sum_j \exp\left(V_{jn}\right) X_{jnk}^{(\lambda_k)} \tag{15}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \theta_r} = \begin{cases} 0 & \text{if } r \neq m \\ \sum\limits_j \exp\left(V_{jn}\right) & \text{if } r = m \end{cases} \tag{16}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \lambda_k} = \exp\left(V_{mn}\right) \beta_k \Delta_{mnk} + \theta_m \sum_j \exp\left(V_{jn}\right) \beta_k \Delta_{jnk} \tag{17}$$

G-DOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial \beta_k} = \exp\left(V_{mn}\right) X_{mnk}^{(\lambda_k)} + \sum_{j \neq m} \theta_{mj} \exp\left(V_{jn}\right) X_{jnk}^{(\lambda_k)} \tag{18}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \theta_{rs}} = \begin{cases} 0 & \text{if } r \neq m \\ \exp\left(V_{sn}\right) & \text{if } r = m \text{ and } r \neq s \end{cases} \tag{19}$$

LIN-IPT-L

$$\frac{\partial \mathcal{U}_{mn}}{\partial \beta_k} = A_{mn}^{\frac{1}{\phi_m}-1} \exp\left(V_{mn}\right) X_{mnk}^{(\lambda_k)} \tag{20}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \phi_r} = \begin{cases} 0 & \text{if } r \neq m \\ A_{mn}^{1/\phi_m}\left[\frac{\exp\left(V_{mn}\right)}{A_{mn}} - \ln A_{mn}^{1/\phi_m}\right]/\phi_m & \text{if } r = m \text{ and } \phi_m > 0 \\ -\frac{1}{2}\exp\left(\exp V_{mn}\right)\exp\left(2V_{mn}\right) & \text{if } r = m \text{ and } \phi_m = 0 \end{cases} \tag{21}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \mu_r} = \begin{cases} 0 & \text{if } r \neq m \\ -1 & \text{if } r = m \end{cases} \tag{22}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \lambda_k} = A_{mn}^{\frac{1}{\phi_m}-1} \exp\left(V_{mn}\right) \beta_k \Delta_{mnk} \tag{23}$$

where $A_{mn} = \phi_m \exp\left(V_{mn}\right) + 1$. Note that $\lim\limits_{\phi_m \to 0} A_{mn}^{\frac{1}{\phi_m}-1} = \exp\left(\exp V_{mn}\right)$ in the derivatives of $\mathcal{U}_{mn}$ with respect to β_k and λ_k.

BT-IPT-L

$$\frac{\partial \mathcal{U}_{mn}}{\partial \beta_k} = \mathcal{U}_{mn} B_{mn}^{\tilde{\phi}_m - 1} \exp\left(V_{mn}\right) X_{mnk}^{(\lambda_k)} \tag{24}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \tilde{\phi}_r} = \begin{cases} 0 & \text{if } r \neq m \\ \mathcal{U}_{mn}\left[B_{mn}^{\tilde{\phi}_m} \ln B_{mn} - B_{mn}^{(\tilde{\phi}_m)}\right] / \tilde{\phi}_m & \text{if } r = m \text{ and } \tilde{\phi}_m \neq 0 \\ \frac{1}{2}\mathcal{U}_{mn} \ln^2 B_{mn} & \text{if } r = m \text{ and } \tilde{\phi}_m = 0 \end{cases} \tag{25}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \tilde{\mu}_r} = \begin{cases} 0 & \text{if } r \neq m \\ \mathcal{U}_{mn} B_{mn}^{\tilde{\phi}_m - 1} & \text{if } r = m \end{cases} \tag{26}$$

$$\frac{\partial \mathcal{U}_{mn}}{\partial \lambda_k} = \mathcal{U}_{mn} B_{mn}^{\tilde{\phi}_m - 1} \exp\left(V_{mn}\right) \beta_k \Delta_{mnk} \tag{27}$$

where $B_{mn} = \exp\left(V_{mn}\right) + \tilde{\mu}_m$. Note that $\lim\limits_{\tilde{\phi}_m \to 0} \mathcal{U}_{mn} B_{mn}^{\tilde{\phi}_m - 1} = 1$ in the derivatives of $\mathcal{U}_{mn}$ with respect to β_k, $\tilde{\mu}_r$ and λ_k.

The derivative of $X_{mnk}^{(\lambda_k)}$ with respect to λ_k, denoted as Δ_{mnk} in (14), (17), (23) and (27), can be written as:

$$\Delta_{mnk} = \frac{\partial X_{mnk}^{(\lambda_k)}}{\partial \lambda_k} = \begin{cases} \dfrac{1}{\lambda_k}\left[X_{mnk}^{\lambda_k} \ln X_{mnk} - X_{mnk}^{(\lambda_k)}\right] & \text{if } \lambda_k \neq 0 \\ \dfrac{1}{2} \ln^2 X_{mnk} & \text{if } \lambda_k = 0 \end{cases}. \tag{28}$$

At the maximum point of the log-likelihood function L, the covariance matrix of the estimated parameters $\widehat{\Pi}$ is evaluated as the inverse of minus the matrix of second derivatives of L for the LOGIT family, and as the inverse of the matrix of cross products of the first derivatives of L for the other families (Berndt, Hall, Hall and Hausman, 1974), since it is easier to compute the first derivatives than the second derivatives:

LOGIT

$$\text{Var}\left(\widehat{\Pi}\right) = \left[-\frac{\partial^2 L}{\partial \Pi \partial \Pi'}\right]_{\Pi = \widehat{\Pi}}^{-1} \tag{29}$$

where

$$\frac{\partial^2 L}{\partial \beta_r \partial \beta_m} = -\sum_n \sum_i P_{in}\left[X_{inr}^{(\lambda_r)} - \sum_j X_{jnr}^{(\lambda_r)} P_{jn}\right]\left[X_{inm}^{(\lambda_m)} - \sum_j X_{jnm}^{(\lambda_m)} P_{jn}\right] \tag{30}$$

$$\frac{\partial^2 L}{\partial \lambda_r \partial \beta_m} = -\sum_n \sum_i \left\{X_{inm}^{(\lambda_m)} \frac{\partial P_{in}}{\partial \lambda_r} - \left[d_{in} - P_{in}\right] \frac{\partial X_{inm}^{(\lambda_m)}}{\partial \lambda_r}\right\} \tag{31}$$

and

$$\frac{\partial^2 L}{\partial \lambda_r \partial \lambda_m} = -\sum_n \sum_k \beta_k \sum_i \left\{\frac{\partial X_{ink}^{(\lambda_k)}}{\partial \lambda_m} \frac{\partial P_{in}}{\partial \lambda_r} - \left[d_{in} - P_{in}\right] \frac{\partial^2 X_{ink}^{(\lambda_k)}}{\partial \lambda_r \partial \lambda_m}\right\}. \tag{32}$$

The derivative $\partial P_{in}/\partial\lambda_r$ in (31) and (32) is computed as:

$$\frac{\partial P_{in}}{\partial\lambda_r} = P_{in}\sum_k \beta_k \left[\frac{\partial X_{ink}^{(\lambda_k)}}{\partial\lambda_r} - \sum_j P_{jn}\frac{\partial X_{jnk}^{(\lambda_k)}}{\partial\lambda_r}\right]. \tag{33}$$

The derivatives $\partial X_{inm}^{(\lambda_m)}/\partial\lambda_r$ in (31), $\partial X_{ink}^{(\lambda_k)}/\partial\lambda_m$ in (32), $\partial X_{ink}^{(\lambda_k)}/\partial\lambda_r$ and $\partial X_{jnk}^{(\lambda_k)}/\partial\lambda_r$ in (33) are computed as in (28).

Finally, the second derivative $\partial^2 X_{ink}^{(\lambda_k)}/\partial\lambda_r\partial\lambda_m$ in (32) is computed as:

$$\frac{\partial^2 X_{ink}^{(\lambda_k)}}{\partial\lambda_r\partial\lambda_m} = \begin{cases} 0 & \text{if } \lambda_k \neq \lambda_r \text{ or } \lambda_k \neq \lambda_m \\ \dfrac{1}{\lambda_k}\left[X_{ink}^{\lambda_k}\ln^2 X_{ink} - 2\dfrac{\partial X_{ink}^{(\lambda_k)}}{\partial\lambda_k}\right] & \text{if } \lambda_k = \lambda_r = \lambda_m \neq 0 \\ \dfrac{1}{3}\ln^3 X_{ink} & \text{if } \lambda_k = \lambda_r = \lambda_m = 0 \end{cases}. \tag{34}$$

OTHER FAMILIES

$$\mathrm{Var}\left(\widehat{\Pi}\right) = \left[\sum_n \frac{\partial L_n}{\partial\Pi}\frac{\partial L_n}{\partial\Pi'}\right]^{-1}_{\Pi=\widehat{\Pi}} \tag{35}$$

where $\partial L_n/\partial\Pi$ is a column vector of the first derivatives of the log-likelihood function L_n corresponding to an individual n with respect to the elements of Π:

$$\frac{\partial L_n}{\partial\Pi} = \sum_i d_{in}\frac{\partial\ln P_{in}}{\partial\Pi} = \sum_i d_{in}\left(\frac{1}{\mathcal{U}_{in}}\frac{\partial\mathcal{U}_{in}}{\partial\Pi} - \frac{1}{\sum_m \mathcal{U}_{mn}}\sum_m\frac{\partial\mathcal{U}_{mn}}{\partial\Pi}\right) \quad . \tag{36}$$

1.3 Families and their Model Types

Each model family includes different model types which depend on the specification of the independent variables in the representative utility components V_i's and the use of the Box-Cox transformations on the strictly positive variables.

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

1. The constant terms which have a value of one in each transportation mode where they appear and zero elsewhere;
2. 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;
3. 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.

Generally speaking, three model types can be estimated in each family:

1. The **Standard Linear Model Type** where 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_i 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_i, but not in the other V_i's $(i \neq j)$. The term "Linear" indicates that all the independent variables enter linearly without any Box-Cox transformation involved.
   For the **LOGIT** family, this model type corresponds to the Linear Classical Logit model commonly used in empirical studies where all the independent variables are generic, although this constraint is not necessary and can be relaxed. The term "Classical" is used to distinguish this model type from the Universal Logit of McFadden (1975), which is more general in the sense that each V_i can also include the network characteristics of some or all the other modes $j \neq i$ provided that sufficient nonlinearities are introduced on the X_{ik}'s and X_{jk}'s to prevent the multicollinearity problem which necessarily exists if the variables are all linear.
2. The **Standard Box-Cox Model Type** differs from the previous one only by the use of fixed or estimated Box-Cox transformations on the strictly positive independent variables. Note that it is equivalent to the Standard Linear Model Type if every λ-parameter is fixed at the value of one.
3. The **Generalized Box-Cox Model Type** (Gaudry, 1978) allows a more general specification of the independent variables than the Standard Box-Cox Model Type 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.

For the **IPT-L** families only, a fourth model type called the **Generalized Linear Model Type** can be considered due to the nonlinearity of the direct and inverse Box-Tukey transformations on the exponential functions of V_i's which allows a more general specification of the independent variables than the Standard Linear Model Type, i.e. like the Generalized Box-Cox Model Type but without the Box-Cox transformations on these variables.

Table 2 summarizes the different model types that can be estimated in each of the five families. For easy reference, symbolic names of the model types and the transformations on the independent variables are also given and will be used with the family name as follows: **FAMILY (Model Type – Transformation)**. Each of the first three families (**LOGIT**, **S-DOGIT**, **G-DOGIT**) includes three model types with their respective transformations, namely (S–LIN), (S–BC) and (G–BC), whereas each of the last two families (**LIN-IPT-L**, **BT-IPT-L**) includes four model types with their associated transformations, namely (S–LIN), (G–LIN), (S–BC) and (G–BC). For example, the Generalized Box-Cox **LOGIT** model can be referred to as **LOGIT**(G–BC), the Standard Linear **S-DOGIT** model as **S-DOGIT**(S–LIN), the Standard Box-Cox **G-DOGIT** model as **G-DOGIT**(S–BC), the Generalized Box-Cox **LIN-IPT-L** as **LIN-IPT-L**(G–BC), and the Standard Linear **BT-IPT-L** as **BT-IPT-L**(S–LIN).

All the model types in each family can be estimated if all the alternatives are available for every individual in the sample, but only the (S-LIN) and (S-BC) model types can be estimated if not all the alternatives are available for some individuals since a network characteristic of an ith alternative can become unavailable across the other alternatives, for the (G-LIN) and (G-BC) model types.

TABLE 2 Symbolic names for the model types and transformations in the five families.

FAMILY	Model Type	Transformation	Symbolic Name
LOGIT	Standard	LINear	(S-LIN)
	Standard	Box-Cox	(S-BC)
	Generalized	Box-Cox	(G-BC)
S-DOGIT	Standard	LINear	(S-LIN)
	Standard	Box-Cox	(S-BC)
	Generalized	Box-Cox	(G-BC)
G-DOGIT	Standard	LINear	(S-LIN)
	Standard	Box-Cox	(S-BC)
	Generalized	Box-Cox	(G-BC)
LIN-IPT-L	Standard	LINear	(S-LIN)
	Generalized	LINear	(G-LIN)
	Standard	Box-Cox	(S-BC)
	Generalized	Box-Cox	(G-BC)
BT-IPT-L	Standard	LINear	(S-LIN)
	Generalized	LINear	(G-LIN)
	Standard	Box-Cox	(S-BC)
	Generalized	Box-Cox	(G-BC)

2 LOGIT FAMILY

2.1 Standard Linear Model Type

This model type which is traditionally used in empirical studies for its computational ease, hence called the LINEAR CLASSICAL LOGIT model, corresponds to a set of fixed values of the extra parameters which reduces the function $\mathcal{U}_{in}(\Pi)$ to a single term $\exp(V_{in})$ where the representative utility component V_{in} only contains a linear specification of the independent variables X_{ink}'s and a set of β-coefficients to be estimated:

i. The parameter θ_{ij} is set at zero in the $w_{ij}\left(\tilde{\theta}_{ij}\right)$ function for every i and j so that $\mathcal{U}_{in}(\Pi)$ only depends on $f_i(\cdot)$, $g_i(\cdot)$ and $\exp(V_{in})$ but not on those associated with the other alternatives $j \neq i$.

ii. The parameters ϕ_i and μ_i are set at one in the $f_i(\cdot)$ function for every i, hence $f_i(\cdot)$ is a linear function of $g_i(\cdot)$.

iii. The parameters $\tilde{\phi}_i$ and $\tilde{\mu}_i$ are set at one in the $g_i(\cdot)$ function for every i, hence $g_i(\cdot)$ is a linear function of $\exp(V_{in})$.

iv. Every V_{in} includes:

   a. The independent variables X_{ink}'s in linear forms, i.e. without the Box-Cox transformations or other types of transformations inducing nonlinearities;
   b. Only the network characteristics of the ith alternative;
   c. One common β-coefficient for each network characteristic across the alternatives: this form of abstractness though frequently imposed can be relaxed if mode-specific effects are to be analyzed.

This model is characterized by the following properties:

1. **Additive separability of utility**

Due to (i), (ii), (iii) and (iv.b), this model type implies that:

(a) The cross elasticities of demand are uniform: the percentage changes in the shares of the alternatives j with respect to a percentage change in a network characteristic of an ith alternative $(i \neq j)$ are all equal in accordance with Luce's Independence from Irrelevant Alternatives (IIA) axiom (Luce 1959) which states that "if a set of alternative choices exists, then the relative probability of choice among any two alternatives is unaffected by the removal (or addition) of any set of other alternatives".

(b) The alternatives behave as substitutes only, but not as complements.

(c) Only the differences in the utility components, $V_{mn} - V_{in} = \sum_k \beta_k(X_{mnk} - X_{ink})$, hence the differences in the characteristics, matter but not the absolute levels of the V's or X's since the utility is ordinal, not cardinal. Adding a constant c in all V_{mn}'s including V_{in} leaves P_{in} unchanged, although the level of every V_{mn} will shift by the same amount: $P^*_{in} = \exp(V_{in}+c)/\sum_m \exp(V_{mn}+c) = 1/\sum_m \exp(V_{mn}+c-V_{in}-c) = 1/\sum_m \exp(V_{mn}-V_{in}) = P_{in}$. Multiplying all V_{mn}'s including V_{in} by a positive factor d affects P_{in}: $P^*_{in} = \exp(dV_{in})/\sum_m \exp(dV_{mn}) = [\exp(V_{in})]^d/\sum_m [\exp(V_{mn})]^d = 1/\sum_m [\exp(V_{mn}-V_{in})]^d = 1/\sum_m [\exp(V_{mn}-V_{in})]^d \neq P_{in}.$ This means that the function $\sum_m [\exp(V_{mn})]^d$ is homogeneous of degree $d > 0$.

2. **Symmetry of response curve**

(d) Two response curves should be considered :

(d.1) The response curve $P_i(V_i)$ in function of V_i: this curve is invariant with respect to the model types in the family since the effect of an independent variable X_{ik} on V_i which is different according to the model types is not explicity taken into account. This curve is nondecreasing and symmetric with respect to its inflection point $\tilde{V}_i$ since the slope $\partial P_i(V_i)/\partial V_i = P_i(V_i)[1 - P_i(V_i)]$ is nonnegative and has the same value when computed at two symmetric points of V_i around $\tilde{V}_i$.

(d.2) The response curve $P_i(X_{ik})$ which indicates that the curve $P_i(V_i)$ is now considered as an explicit function of an independent variable X_{ik}, since V_i is a function of X_{ik}.
The slope of the curve $P_i(X_{ik})$ for LOGIT (S-LIN) can be written as:

$$\frac{\partial P_i(X_{ik})}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\beta_k \tag{37}$$

which shows that $P_i(X_{ik})$ remains symmetric, but can be nondecreasing or nonincreasing if β_k is positive or negative respectively. Note that this slope plays a key role since it is used to compute the elasticity of $P_i(X_{ik})$ with respect to X_{ik} (see Section 5).

3. **Zero probability limits**

(e) The thin-equal-tails property: the choice probability $P_i(V_i)$ tends to zero (one) if V_i tends to $-\infty(\infty)$ given any set of fixed values of the other V_m's. This built-in property prevents the flexibility to fit thicker tails which can reflect the specification error, the modeler's ignorance, the compulsive consumption or captivity of consumers to specific alternatives as in the DOGIT families, or asymmetric tails which can represent a certain amount of stickiness or mode-specificity in the consumer's reactions to various modes as in the IPT-L families.

4. **Underidentification of mode-specific effects**

Due to (ii), (iii) and (iv.a) which constrain respectively the $f_i(\cdot)$ and $g_i(\cdot)$ functions and the independent variables X_{itk}'s to be linear, there is an underidentification of the alternative-specific β-coefficients for:

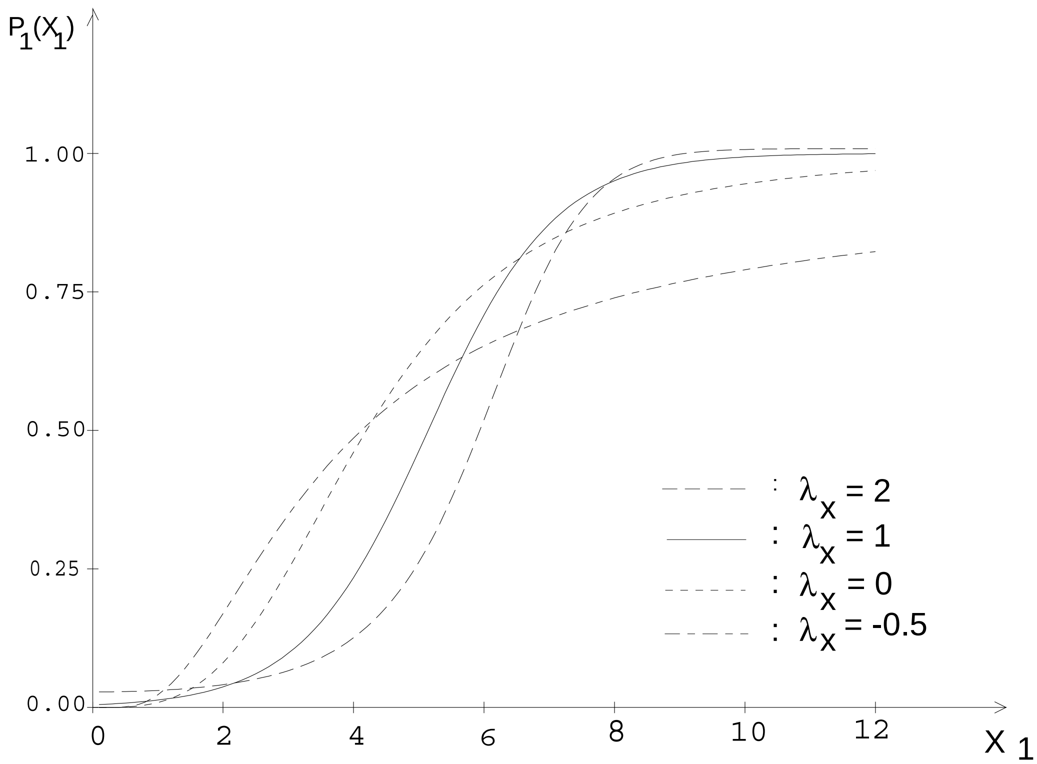
(f) A socioeconomic variable if it appears in all alternatives, as well as a network characteristic of an ith alternative if it is included in all alternatives. For the same reason, two (or more) components of a same network characteristic, say the travel fares by modes i and j, cannot be specified at the same time in all alternatives, if one wants to include in each V_i not only the network characteristics of the ith alternative but also those of all the other alternatives to obtain a system which is very similar to those found in microeconomics where all the commodities prices appear in all demand functions.

(g) The constant terms if they appear in all alternatives.

2.2 Standard Box-Cox Model Type

The STANDARD BOX-COX LOGIT only differs from the LINEAR CLASSICAL LOGIT by modifying the assumption (iv.a) to allow for the use of fixed or estimated Box-Cox transformations on the strictly positive independent variables X_{ink}'s included in each V_{in} : $V_{in} = \sum_k \beta_k X_{ink}^{(\lambda_k)}$.

Figure 1 LOGIT(S–BC) with different λ-values



This implies that:

(d.2) The response curve $P_i(X_{ik})$ is asymmetric with respect to its inflection point $\tilde{X}_{ik}$. The slope of the curve $P_i(X_{ik})$ for LOGIT (S-BC) can be written as:

$$\frac{\partial P_i(X_{ik})}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\beta_k X_{ik}^{\lambda_k - 1} \tag{38}$$

which indicates that $P_i(X_{ik})$ can be nondecreasing or nonincreasing due to the effect on V_i of a unit change in X_{ik}, $\partial V_i/\partial X_{ik}$, which depends not only on the sign of β_k, but also on the level of X_{ik}, if λ_k is different from unity:

- If λ_k is less than unity, then the power $\lambda_k - 1$ of X_{ik} is negative: smaller levels of X_{ik} have greater increasing (decreasing) effects on V_i if β_k is positive (negative).
- If λ_k is equal to unity, i.e. X_{ik} is in a linear form, the effect on V_i is constant but can be positive or negative since it depends only on β_k. In this case only, the response curve $P_i(X_{ik})$ remains symmetric with respect to its inflection point.
- If λ_k is greater than unity, then the power $\lambda_k - 1$ is positive: greater levels of X_{ik} have greater increasing (decreasing) effects on V_i if β_k is positive (negative).

Figure 1 shows the response curve $P_1(X_1)$ for different values of the λ-parameter associated with X_1, given a positive value of β_1.

2.3 Generalized Box-Cox Model Type

The GENERALIZED BOX-COX LOGIT allows a more general specification of the independent variables than the STANDARD BOX-COX LOGIT by relaxing the assumption (iv.b) so that every V_{in} can include not only the network characteristics of its own alternative, but also those associated with some or all alternatives $j \neq i$.

The model can be viewed as a particular estimable form of McFadden's UNIVERSAL LOGIT (1975) which suggests that a network characteristic of a given alternative can be specified at the same time in the other alternatives provided that some unspecified, yet different, functional forms of the variable should be used across the alternatives in the general case where the same variable appears in all the alternatives, for the identification of the associated β-coefficients. Thus distinct Box-Cox parameters should be specified for the same variable if it appears across all the alternatives. This type of specification as far as the network variables are concerned is very similar to a demand equation system where each equation is a function of all the commodity prices in the context of consumer demand theory.

The main difficulty resides in the estimation of all the Box-Cox parameters when the number of alternatives is large and when several network characteristics of a particular mode are jointly specified in all the alternatives.

This general specification of the independent variable implies that:

(d.2) The response curve $P_i(X_{\cdot k})_L$ is asymmetric with respect to its inflection point $\tilde{X}_{\cdot k}$, where $X_{\cdot k}$ represents the kth variable which appears in the ith alternative and also in some or all of the alternatives $j \neq i$.

The slope of the curve can be written as:

$$\frac{\partial P_i(X_{\cdot k})_L}{\partial X_{\cdot k}} = \frac{\partial P_i(V_i)_L}{\partial V_i}\frac{\partial V_i}{\partial X_{\cdot k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_L}{\partial V_j}\frac{\partial V_j}{\partial X_{\cdot k}} \tag{39}$$

where the probability curve of an ith alternative, $P_i(\cdot)$, is considered in turn as a function of V_i and of each $V_j (j \neq i)$. Note that V_i is also a function of $X_{\cdot k}$ whereas each V_j can be or not a function of $X_{\cdot k}$: $\partial V_i / \partial X_{\cdot k} = \beta_k^i X_{\cdot k}^{\lambda_k^i - 1}$ and $\partial V_j / \partial X_{\cdot k} = \beta_k^j X_{\cdot k}^{\lambda_k^j - 1}$ if V_j is a function of $X_{\cdot k}$, and zero otherwise.

Since the slope of $P_i(X_{\cdot k})_L$ has two components — the first corresponds to the variation of $P_i(X_{\cdot k})_L$ due to the variation of $X_{\cdot k}$ in the ith alternative alone, and the second to the variation of $P_i(X_{\cdot k})$ due to the variation of $X_{\cdot k}$ in the other alternatives $j \neq i$ — and that the sign of each component depends on the individual signs of the β-coefficients in the different alternatives where $X_{\cdot k}$ appears, the total effect on $P_i(X_{\cdot k})$ due to the change of $X_{\cdot k}$ in all the alternatives is difficult to predict.

2.4 Specification of the Constant Terms

Since the constant terms have a special meaning and that they can never be transformed by the Box-Cox transformations in the Standard and Generalized Box-Cox LOGIT Model Types, the same specification is used for the three model types: 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 LOGIT 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 observation t can be written in matrix form as:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} X_{1n1} & 0 & 0 & X_{1n4} & 0 & 0 \\ 0 & X_{2n2} & 0 & 0 & X_{2n5} & 0 \\ 0 & 0 & X_{3n3} & 0 & 0 & X_{3n6} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{bmatrix} \tag{40}$$

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_{1n} can be expressed without explicitly writing them as follows:

$$P_{1n} = \frac{\exp(V_{1n})}{\sum_{j=1}^{3} \exp(V_{jn})} = \frac{\exp(\beta_1 + \beta_4 X_{1n4})}{\exp(\beta_1 + \beta_4 X_{1n4}) + \exp(\beta_2 + \beta_5 X_{2n5}) + \exp(\beta_3 + \beta_6 X_{3n6})}. \tag{41}$$

This expression can be represented in three equivalent forms:

$$P_{1n} = \begin{cases} \dfrac{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n4})}{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n4}) + \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2n5}) + \exp(\beta_6 X_{3n6})} \\ \dfrac{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1n4})}{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1n4}) + \exp(\beta_5 X_{2n5}) + \exp(\beta_3 - \beta_2)\exp(\beta_6 X_{3n6})} \\ \dfrac{\exp(\beta_4 X_{1n4})}{\exp(\beta_4 X_{1n4}) + \exp(\beta_2 - \beta_1)\exp(\beta_5 X_{2n5}) + \exp(\beta_3 - \beta_1)\exp(\beta_6 X_{3n6})} \end{cases} \tag{42}$$

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 (42), is irrelevant due to the special meaning of the differences of the constants.

2.5 Specification of the Socioeconomic Variables

1) Standard Linear LOGIT Model

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 (42) 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_{1n} can be rewritten in three equivalent forms:

$$P_{1n} = \begin{cases} \dfrac{\exp(\beta_1^*)\exp[(\beta_4-\beta_6)X_{1n4}]}{\exp(\beta_1^*)\exp[(\beta_4-\beta_6)X_{1n4}] + \exp(\beta_2^*)\exp[(\beta_5-\beta_6)X_{1n4}] + 1} \\ \dfrac{\exp(\beta_1^*)\exp[(\beta_4-\beta_5)X_{1n4}]}{\exp(\beta_1^*)\exp[(\beta_4-\beta_5)X_{1n4}] + \exp(\beta_2^*) + \exp[(\beta_6-\beta_5)X_{1n4}]} \\ \dfrac{\exp(\beta_1^*)}{\exp(\beta_1^*) + \exp(\beta_2^*)\exp[(\beta_5-\beta_4)X_{1n4}] + \exp[(\beta_6-\beta_4)X_{1n4}]} \end{cases} \tag{43}$$

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 (32).

Similarly the second and third forms of (42) can each be expressed in three equivalent forms. Like in the case of the constant terms, the three forms of (43) 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 (43), 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 (42) as an example in which X_{1n4}, X_{2n5}andX_{3n6} represent the same socioeconomic variable $(X_{1n4} = X_{2n5} = X_{3n6})$ but specified as generic $(\beta_4 = \beta_5 = \beta_6)$, P_{1n} can be rewritten as:

$$P_{1n} = \frac{\exp(\beta_1^*)\exp[\beta_4 X_{1n4}]}{\exp(\beta_1^*)\exp[\beta_4 X_{1n4}] + \exp(\beta_2^*)\exp[\beta_4 X_{1n4}] + \exp[\beta_4 X_{1n4}]} \tag{44}$$

which reduces to:

$$P_{1n} = \frac{\exp(\beta_1^*)}{\exp(\beta_1^*) + \exp(\beta_2^*) + 1} \tag{45}$$

It can be seen that the socioeconomic variable has no influence at all on the mode share P_{1n} if it is included in all M alternatives. Hence the variable must be specified in $M-1$ alternatives at most.

2) Standard Box-Cox LOGIT Model

The specification of the socioeconomic variables in this model is similar to the one in the Standard Linear LOGIT Model, 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 (43) 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 mode share P_{1n} can be expressed as:

$$P_{1n} = \frac{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1n4}^{(\lambda_4)}\right]}{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1n4}^{(\lambda_4)}\right] + \exp(\beta_2^*)\exp\left[\beta_5^* X_{1n4}^{(\lambda_5)}\right] + 1} \tag{46}$$

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 (46), if the socioeconomic variable is specified as generic $(\beta_4^* = \beta_5^*)$, 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 LOGIT Model

Unlike the Standard Box-Cox LOGIT Model, the Generalized Box-Cox LOGIT Model 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 (42) 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 mode share P_{1n} can be rewritten as:

$$P_{1n} = \frac{\exp(\beta_1^*) \exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right]}{\exp(\beta_1^*) \exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right] + \exp(\beta_2^*) \exp\left[\beta_5 X_{1n4}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{1n4}^{(\lambda_6)}\right]}. \tag{47}$$

Provided that all the λ-parameters are distinct, there will be no identification problem for β_4, β_5andβ_6.

Generic case

In (47), 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.

2.6 Specification of the Network Variables

1) Standard Linear LOGIT Model

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_{it} but not in the other V_{jt}'s $(j \neq i)$

Alternative-specific case

Using any of the three equivalent forms of (42) as an example in which the variables X_{1n4}, X_{2n5}andX_{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 LOGIT Model

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

$\boxed{\textbf{Alternative-specific case}}$

Rewriting for example the first form of (42) 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 mode share P_{1n} can be expressed as:

$$P_{1n} = \frac{\exp\left(\beta_1^*\right)\exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right]}{\exp\left(\beta_1^*\right)\exp\left[\beta_4 X_{1n4}^{(\lambda_4)}\right] + \exp\left(\beta_2^*\right)\exp\left[\beta_5 X_{2n5}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{3n6}^{(\lambda_6)}\right]}. \tag{48}$$

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.

$\boxed{\textbf{Generic case}}$

In (48), 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 LOGIT Model

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

$\boxed{\textbf{Alternative-specific case}}$

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 LOGIT Models. 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 $M - 1 \quad V_{jn}$'s at most including its own, this subcase is analogous to the alternative-specific case (46) of a socioeconomic variable in the Standard Box-Cox LOGIT Model, 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 (47) of a socioeconomic variable in the Generalized Box-Cox LOGIT Model, 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.

- Subcase I

Modifying the alternative-specific case (48) in the Standard Box-Cox LOGIT Model to illustrate the subcase I, the different V_{jn}' s involving in the mode share P_{1n} can be represented as follows:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1n4}^{(\lambda_4)} & 0 & 0 & 0 & X_{2n5}^{(\lambda_8)} & X_{3n6}^{(\lambda_9)} \\ 0 & 1 & 0 & X_{2n5}^{(\lambda_5)} & 0 & X_{1n4}^{(\lambda_7)} & 0 & 0 \\ 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_6)} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_9 \end{bmatrix} \tag{49}$$

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 in practice it is difficult to estimate this subcase with all the 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.

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 $\lambda_4 \neq \lambda_5 \neq \lambda_6$.
- 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:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1n4}^{(\lambda_4)} & 0 & 0 & 0 & 0 & X_{2n5}^{(\lambda_9)} & 0 & X_{3n6}^{(\lambda_{11})} & 0 \\ 0 & 1 & 0 & X_{2n5}^{(\lambda_5)} & 0 & X_{1n4}^{(\lambda_7)} & 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_{12})} \\ 0 & 0 & 0 & 0 & X_{3n6}^{(\lambda_6)} & 0 & X_{1n4}^{(\lambda_8)} & 0 & X_{2n5}^{(\lambda_{10})} & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_{12} \end{bmatrix} \tag{50}$$

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 $\lambda_4 \neq \lambda_5 \neq \lambda_6$.

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 (49) 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: $\beta_4 = \beta_5 = \ldots = \beta_9$ and $\lambda_4 = \lambda_5 = \ldots = \lambda_9$.

- Subcase II

This subcase is also similar to (50) 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".

A summary of the main specifications for the different types of variables in the three models of the **LOGIT** family is given in the following table.

TABLE 3 Specifications of the variables in the LOGIT family.

TYPE OF VARIABLE	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
CONSTANT	always in $M-1$ alternatives at most	not allowed.
SOCIOECONOMIC		
LOGIT (S-LIN)	in $M-1$ alternatives at most.	same as the alternative-specific case.
LOGIT (S-BC)	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.
LOGIT (G-BC)	in all alternatives with distinct λ-parameters specified across the alternatives.	not allowed.
NETWORK		
LOGIT (S-LIN)	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 $(j \neq i)$.	same as the alternative-specific case.
LOGIT (S-BC)	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.
LOGIT (G-BC)	the characteristic 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 alternative-specific case: • Subcase I : only the alternative-specific is allowed. • Subcase II : same as the alternative-specific subcase.

3 DOGIT FAMILIES

3.1 Standard Linear Model Types

The Standard Linear Model Types of the **DOGIT** families are obtained by relaxing only the assumption (i) of the LINEAR CLASSICAL LOGIT to allow the estimation of the θ-parameters in the weight functions $w_{ij}\left(\tilde{\theta}_{ij}\right)$ as follows:

(i.a) **S-DOGIT**: for every alternative i, the $\theta_{i1},\ldots,\theta_{iM}$ are constrained to be equal, say to a common value θ_i, which implies that only one extra parameter is added per alternative:

$$P_i(V_i) = \frac{\exp\left(V_i\right) + \theta_i \sum\limits_j \exp\left(V_j\right)}{\sum\limits_m \left[\exp\left(V_m\right) + \theta_m \sum\limits_h \exp\left(V_h\right)\right]}, (\theta_i \geq 0), \tag{51}$$

(i.b) **G-DOGIT**: for every alternative i, θ_{ii} is normalized at zero, which implies that only M(M-1) θ-parameters, $\theta_{ij}(i \neq j)$, can be identified in the whole system:

$$P_i(V_i) = \frac{\exp\left(V_i\right) + \sum\limits_{j\neq i} \theta_{ij} \exp\left(V_j\right)}{\sum\limits_m \left[\exp\left(V_m\right) + \sum\limits_{h\neq m} \theta_{mh} \exp\left(V_h\right)\right]}, (\theta_{ij} \geq 0). \tag{52}$$

Both model types include the LINEAR CLASSICAL LOGIT as a special case by setting all the θ_i's and θ_{ij}'s at zero for S-DOGIT and G-DOGIT respectively, but the former is not nested in the latter due to the equality constraints on the θ-parameters in the S-DOGIT family, except for the binary case where the two model types are exactly equivalent although the types of parametrization are different. Hence for this case one can indifferently estimate the S-DOGIT(S–LIN) or G-DOGIT(S–LIN) to obtain the same fit of the data: the estimated shares P_{it}'s, as well as the log-likelihood function value, will remain the same, except for the three parameters namely the alternative-specific constant and the two θ-parameters.

In contrast with the LINEAR CLASSICAL LOGIT, the DOGIT families introduce the following new properties:

a. The cross elasticities of demand are not uniform since the relative choice probabilities of any two alternatives now depend on the characteristics of all the alternatives. This added property avoids or "dodges" the IIA restriction, hence the name DOGIT, to allow for more flexible patterns of crossalternative substitution.

b. Whereas the S-DOGIT family restricts the alternatives to behave as substitutes only as in the LINEAR CLASSICAL LOGIT, the G-DOGIT family also allows complements.

(d) The probability curve of the S-DOGIT family, $P_i(V_i)_{SD}$, remains symmetric with respect to its inflection point although it will be less steep than in the LINEAR CLASSICAL LOGIT, since $P_i(V_i)_{SD}$ is just a linear transformation of the symmetric probability curve of the LINEAR CLASSICAL LOGIT, $P_i(V_i)_L$, given a set of fixed θ-values:

$$\begin{aligned} P_i(V_i)_{SD} &= \frac{\exp\left(V_i\right) + \theta_i \sum\limits_j \exp\left(V_j\right)}{\sum\limits_m \exp\left(V_m\right) + \sum\limits_m \theta_m \sum\limits_h \exp\left(V_h\right)} \\ &= \frac{P_i(V_i)_L + \theta_i}{1 + \sum\limits_m \theta_m} = \frac{1}{1+\sum\limits_m \theta_m} P_i(V_i)_L + \frac{\theta_i}{1+\sum\limits_m \theta_m} \end{aligned} \tag{53}$$

where $P_i(V_i)_L = \exp(V_i) / \sum_m \exp(V_m)$.

Hence, for a given set of θ-values, the slope curve of $P_i(V_i)_{SD}$ is also symmetric since it is in constant proportion to the slope curve of $P_i(V_i)_L$ over the whole range $-\infty < V_i < +\infty$:

$$\frac{\partial P_i(V_i)_{SD}}{\partial V_i} = \frac{1}{1 + \sum_m \theta_m} \frac{\partial P_i(V_i)_L}{\partial V_i}, \tag{54}$$

which clearly indicates that it is less steep than the slope of $P_i(V_i)_L$ since $\sum_m \theta_m$ is always positive. Note that $\partial P_i(V_i)_L / \partial V_i$ which is equal to $(1 - P_i(V_i)_L) P_i(V_i)_L$ is symmetric with respect to the inflection point of the probability curve $P_i(V_i)_L$.

Similarly, the response curve $P_i(X_{ik})_{SD}$ is also symmetric with respect to its inflection point since its slope is in constant proportion to the slope of $P_i(V_i)_L$:

$$\frac{\partial P_i(X_{ik})_{SD}}{\partial X_{ik}} = \beta_k \frac{\partial P_i(V_i)_{SD}}{\partial V_i} = \frac{\beta_k}{1 + \sum_m \theta_m} \frac{\partial P_i(V_i)_L}{\partial V_i}. \tag{55}$$

This curve is nondecreasing or increasing if β_k is positive or negative respectively, and less steep than the curve $P_i(X_{ik})_L$ of the LOGIT (S-LIN) model type since $\sum_m \theta_m$ is always positive.

Three economic interpretations can be given to the expression (53) if every θ_m is constrained to be nonnegative:

- In terms of analogies with Stone's (1954) approach to expenditure systems, the second term in (53), $\theta_i / \left(1 + \sum_m \theta_m\right)$, represents the share of the consumer's income to satisfy his compulsive or basic needs for the ith commodity irrespective of its price, and the factor in the first term, $1/\left(1 + \sum_m \theta_m\right)$, corresponds to the share of discretionary income left over after all the basic needs are satisfied.
- In the context of discrete choice models with probabilistic choice set generation based on the assumption that the modeler can no longer specify correctly the decision-maker's true choice set — the generation process has been first studied by Manski (1977) — Ben-Akiva (1977) has given another interpretation of the θ-parameters: the second term in (53) represents the probability of captivity to alternative i, and the factor in the first term corresponds to the probability of the consumer being free to choose from the full choice set of alternatives. More recently, Swait and Ben-Akiva (1987) have extended this notion to a more general model called the Parametrized Logit Captivity (PLC) model by replacing every θ_m with the function $\exp\left(\tilde{V}_m\right)$, where $\tilde{V}_m$ is a linear combination of the variables which are thought to explain the captivity of alternative m and which can partially or totally be chosen from the set of variables included in V_m. Unlike the S-DOGIT family, the PLC model also allows complements.
- In marketing science, customer captivity or loyalty to a product, i.e. the bias that people tend to buy the same product as they bought before, has been recently considered by Colombo and Morrison (1989). Their model postulates a simple form of dependency between the previous and present purchase periods, and is not a random utility model like the S-DOGIT(S–LIN) model which considers only information about an individual's present choices. Using Markov chain theory, Bordley (1990) showed that given a fixed probability distribution over an individual's previous purchases, the choice probabilities associated with the present period in the Colombo-Morrison model are just the dogit probabilities, and that the equilibrium or asymptotic choice

probabilities, i.e. the ones which will arise if the choice process is continually repeated, are not in general the dogit probabilities, except under certain strong conditions they reduce to the logit, and under somewhat weaker conditions to the dogit. Furthermore, exploiting the fact that the Colombo-Morrison model can be reinterpreted as a conditional logit model that allows for varying degrees of loyalty across the population (Bordley 1989), he similarly deduced that the dogit model can be used even if buyers are not perfectly captive to one choice.

On the contrary, the probability curve of the G-DOGIT family, $P_i(V_i)_{GD}$, is not symmetric with respect to its inflection point since it depends nonlinearly on the choice probabilities of all the alternatives of the LINEAR CLASSICAL LOGIT:

$$\begin{aligned} P_i(V_i)_{GD} &= \frac{\exp(V_i) + \sum_{j \neq i} \theta_{ij} \exp(V_j)}{\sum_m \exp(V_m) + \sum_m \sum_{h \neq m} \theta_{mh} \exp(V_h)} \\ &= \frac{P_i(V_i)_L + \sum_{j \neq i} \theta_{ij} P_j(V_j)_L}{1 + \sum_m \sum_{h \neq m} \theta_{mh} P_h(V_h)_L} \end{aligned} \tag{56}$$

Hence, for a given set of θ-values, the slope of $P_i(V_i)_{GD}$ is not in constant proportion to the slope of $P_i(V_i)_L$ as in the S-DOGIT case:

$$\frac{\partial P_i(V_i)_{GD}}{\partial V_i} = \left[\frac{1 - P_i(V_i)_{GD} \sum_{m \neq i} \theta_{mi}}{1 + \sum_m \sum_{h \neq m} \theta_{mh} P_h(V_h)_L} \right] \frac{\partial P_i(V_i)_L}{\partial V_i}. \tag{57}$$

which indicate that the slope of $P_i(V_i)_{GD}$ is less steep than the slope of $P_i(V_i)_L$ since the quotient in the square brackets is always less than unity.

To ensure that the probability curve $P_i(V_i)_{GD}$ is monotonically nondecreasing over the whole range of V_i, i.e. $\partial P_i(V_i)_{GD}/\partial V_i$ is always nonnegative, since the denominator in the factor of $\partial P_i(V_i)_L/\partial V_i$ is always greater than or equal to one, the numerator should be positive, which implies that $P_i(V_i)_{GD} < 1/\sum_{m \neq i} \theta_{mi}$. To check this inequality we need to consider only the right tail of $P_i(V_i)_{GD}$ which is the upper bound of the probability curve. Since the limit of $P_i(V_i)_{GD}$ as V_i tends towards to infinity is: $\lim_{V_i \to \infty} P_i(V_i)_{GD} = 1/\left(1 + \sum_{m \neq i} \theta_{mi}\right)$, the inequality holds for the upper bound.

Similarly, the probability curve $P_i(X_{ik})_{GD}$ is not symmetric with respect to it inflection point since its slope is not in constant proportion to the slope of $P_i(V_i)_L$:

$$\frac{\partial P_i(X_{ik})_{GD}}{\partial X_{ik}} = \beta_k \frac{\partial P_i(V_i)_{GD}}{\partial V_i} = \beta_k \left[\frac{1 - P_i(V_i)_{GD} \sum_{m \neq i} \theta_{mi}}{1 + \sum_m \sum_{h \neq m} \theta_{mh} P_h(V_h)_L} \right] \frac{\partial P_i(V_i)_L}{\partial V_i}. \tag{58}$$

This curve is nondecreasing or nonincreasing if β_k is positive or negative respectively, and less steep than the curve $P_i(X_{ik})_L$ of LOGIT (S-LIN) since the quotient in the square brackets is always less than unity.

One interpretation of the parameters θ_{ij}'s can be given in terms of the similarity / dissimilarity measures. In fact, all models forms which try to incorporate these measures into the LOGIT model are variants of Restle's (1961) form which is based on the notion that the variation among the choice alternatives will influence the choice process, i.e. the attributes that display little or no variability among the alternatives do not affect the choice process.
In general, the class of choice models using these measures can be expressed as:

$$P_i(V_i) = \frac{R_i \exp(V_i)}{\sum_m R_m \exp(V_m)} \tag{59}$$

where R_i is a positive measure of the degree of similarity / dissimilarity between alternative i and all other choice alternatives.

The models previously considered by Batsell (1981), Meyer and Eagle (1981, 1982), Borgers and Timmermans (1984), and Cascetta et al. (1996) all use a form of R_i which is based on an average degree of similarity / dissimilarity between the attributes X_{ik}'s and the other X_{jk}'s across the alternatives.

For the G-DOGIT family, the measure R_i can be derived as:

$$R_i = 1 + \sum_{j \neq i} \theta_{ij} \exp(V_j - V_i) \tag{60}$$

which is not an average form since both V_i and V_j vary across the observations.

A shift in the overall V_i value resulting from the similarity / dissimilarity of the attributes across the alternatives can be best seen if $P_i(V_i)$ is rewritten under the following form:

$$P_i(V_i) = \frac{\exp(V_i + \ln R_i)}{\sum_m \exp(V_m + \ln R_m)} \tag{61}$$

where $\ln R_i$ can now be viewed as the total shift in V_i resulting from the similarity / dissimilarity of the attributes X_{ik}'s to the other X_{jk}'s across the alternatives.

For expositional simplicity, assuming that all the attributes X_{ik}'s and X_{jk}'s are generic, except for the alternative-specific constants β_{i0} and β_{j0}, the total shift can be written as:

$$\begin{aligned} \ln R_i &= \ln \left[1 + \sum_{j \neq i} \theta_{ij} \exp(V_j - V_i) \right] \\ &= \ln \left\{ 1 + \sum_{j \neq i} \theta_{ij} \exp(\beta_{j0} - \beta_{i0}) \prod_k \exp\left[\beta_k (X_{jk} - X_{ik})\right] \right\} \\ &= \ln \left\{ 1 + \sum_{j \neq i} \theta_{ij} \exp(\beta_{j0} - \beta_{i0}) \prod_k \left[\exp(X_{jk} - X_{ik})\right]^{\beta_k} \right\} \end{aligned} \tag{62}$$

where β_{j0} and β_{i0} are the original constants associated with the jth and ith alternatives, before normalizing one of the M constants in the system.

For a given alternative j, if every attribute X_{jk} is identical to $X_{ik} (k = 1, K)$ in the extreme case, then $\prod_k \left[\exp(X_{jk} - X_{ik})\right]^{\beta_k}$ is equal to unity independently of the values of the β_k's, and the partial shift in V_i resulting from this perfect similarity is equal to $\ln\left[1 + \theta_{ij} \exp(\beta_{j0} - \beta_{i0})\right]$. In this respect, the G-DOGIT completely differs from the previous models due to the fact that these models consider an effect on V_i only if there is a dissimilarity among the attributes and none in the similarity case. In other words, it has an additional feature implying that there is still a shift in V_i even if the attributes X_{jk}'s are similar to the X_{ik}'s. In fact, the acoustic and psychological experiments by Clarke (1957) and Coombs (1958) respectively have shown that when two stimuli are similar, there is a finite probability of confusion between them, i.e. the probability of attaining a state of indifference is not zero, unlike the LOGIT model which assumes that this probability is zero even if the differences in the utilities are very small. Models which attempt to incorporate this concept of thresholds of indifference — the "minimum perceivable differences" among the utilities — into the LOGIT model, have been considered by Krishnan (1977) and Lioukas (1984).

On the contrary, if the X_{ik}'s are dissimilar to the X_{jk}'s, then the partial shift in V_i can be greater or smaller than $\ln\left[1+\theta_{ij}\exp\left(\beta_{j0}-\beta_{i0}\right)\right]$ depending on the combination of the cases where each X_{jk} is greater or smaller than X_{ik} with the sign and the magnitude of β_k.

The thick-unequal tails can be fitted for both families.

S-DOGIT

According to (53), the thickness of the left tail of the probability curve $P_i(V_i)_{SD}$ is given by:

$$\lim_{V_i\to-\infty} P_i(V_i)_{SD} = \frac{1}{1+\sum\limits_m \theta_m}\ \lim_{V_i\to-\infty} P_i(V_i)_L + \frac{\theta_i}{1+\sum\limits_m \theta_m}. \tag{63}$$

Since $\lim\limits_{V_i\to-\infty} P_i(V_i)_L$ is equal to zero, the expression above reduces to:

$$\lim_{V_i\to-\infty} P_i(V_i)_{SD} = \frac{\theta_i}{1+\sum\limits_m \theta_m}, \tag{64}$$

which represents the probability of captivity to alternative$i$.

The thickness of the right tail can be expressed as:

$$1-\lim_{V_i\to\infty} P_i(V_i)_{SD} = 1-\left[\frac{1}{1+\sum\limits_m \theta_m}\lim_{V_i\to\infty} P_i(V_i)_L + \frac{\theta_i}{1+\sum\limits_m \theta_m}\right]. \tag{65}$$

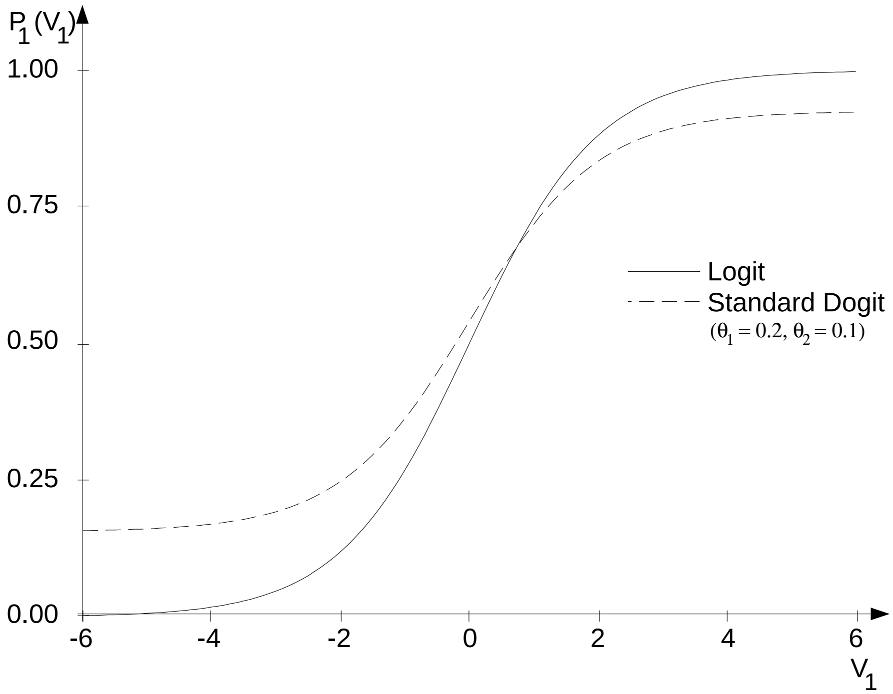
Since $\lim\limits_{V_i\to\infty} P_i(V_i)_L$ is equal to one, the expression above reduces to:

$$1-\lim_{V_i\to\infty} P_i(V_i)_{SD} = 1-\frac{1+\theta_i}{1+\sum\limits_m \theta_m} = \frac{\sum\limits_{h\neq i}\theta_h}{1+\sum\limits_m \theta_m}, \tag{66}$$

which represents the sum of the left tails of all the other probability curves $P_h(V_h)_{SD}$'s $(h\neq i)$, i.e. the sum of the probabilities of captivity to all the other alternatives h.

Hence the two tails are of unequal thickness and independent of the choice probabilities of the alternatives but the probability curve $P_i(V_i)_{SD}$ remains symmetric with respect to its inflection point. Figure 2 shows the probability curves $P_1(V_1)_L$ and $P_1(V_1)_{SD}$ given $V_2=0$ with $(\theta_1,\theta_2)=(0.2,0.1)$ for the binary case $(M=2)$. Note that each curve is invariant with respect to the model types of the family since it depends on the V_m's only.

Figure 2 LOGIT(family) versus S-DOGIT(family).



G-DOGIT

According to (56), the thickness of the left tail of the probability curve $P_i(V_i)_{GD}$ is given by:

$$\lim_{V_i \to -\infty} P_i(V_i)_{GD} = \frac{\lim\limits_{V_i \to -\infty} \left[P_i(V_i)_L + \sum\limits_{j \neq i} \theta_{ij} P_j(V_j)_L \right]}{\lim\limits_{V_i \to -\infty} \left[1 + \sum\limits_m \sum\limits_{h \neq m} \theta_{mh} P_h(V_h)_L \right]} \tag{67}$$

$$= \frac{\sum\limits_{j \neq i} \theta_{ij} P_j(V_j)_L}{1 + \sum\limits_m \sum\limits_{h \neq m, i} \theta_{mh} P_h(V_h)_L},$$

since $\lim\limits_{V_i \to -\infty} P_i(V_i)_L = 0.$

Hence, in general, the left tail is of variable thickness since it depends on the choice probabilities of all the other alternatives $j \neq i$, except for the limit case where one of the $P_j(V_j)_L$'s, say $P_r(V_r)_L$, is allowed to tend towards unity as V_r tends towards infinity, while all the other choice probabilities $P_j(V_j)_L$'s ($j \neq r$) tend towards to zero due to the constraint $\sum\limits_m P_m(V_m)_L = 1$. The left tail is then independent of the choice probabilities:

$$\lim_{\substack{V_i, V_{j \neq r} \to -\infty \\ V_r \to \infty}} P_i(V_i)_{GD} = \frac{\theta_{ir}}{1 + \sum\limits_{m \neq r} \theta_{mr}}. \tag{68}$$

The thickness of the right tail can be expressed as:

$$1-\lim_{V_i\to\infty}P_i(V_i)_{GD}=1-\frac{\lim\limits_{V_i\to\infty}\left[P_i(V_i)_L+\sum\limits_{j\neq i}\theta_{ij}P_j(V_j)_L\right]}{\lim\limits_{V_i\to\infty}\left[1+\sum\limits_{m}\sum\limits_{h\neq m}\theta_{mh}P_h(V_h)_L\right]}. \tag{69}$$

Due to the constraint $\sum\limits_{m}P_m(V_m)_L=1$, if $P_i(V_i)_L$ tends towards unity as V_i tends towards infinity, then all the other choice probabilities $P_j(V_j)_L$'s $(j\neq i)$ tend towards zero. Hence the thickness of the right tail reduces to:

$$1-\lim_{V_i\to\infty}P_i(V_i)_{GD}=1-\frac{1}{1+\sum\limits_{m\neq i}\theta_{mi}}=\frac{\sum\limits_{m\neq i}\theta_{mi}}{1+\sum\limits_{m\neq i}\theta_{mi}}, \tag{70}$$

which represents the sum of the left tails of all the other $P_m(V_m)_{GD}$'s $(m\neq i)$ in the limit case (68). This property can be clearly seen with an example for three alternatives $(M=3)$. The thickness of the right tail of $P_1(V_1)_{GD}$ can be written as:

$$\begin{aligned}1-\lim_{V_i\to\infty}P_1(V_1)_{GD}&=\frac{\theta_{21}+\theta_{31}}{1+\theta_{21}+\theta_{31}}=\frac{\theta_{21}}{1+\theta_{21}+\theta_{31}}+\frac{\theta_{31}}{1+\theta_{21}+\theta_{31}}\\&=\lim_{\substack{V_2,V_3\to-\infty\\V_1\to\infty}}P_2(V_2)_{GD}+\lim_{\substack{V_3,V_2\to-\infty\\V_1\to\infty}}P_3(V_3)_{GD}.\end{aligned} \tag{71}$$

Hence, the two tails of the probability curve $P_i(V_i)_{GD}$ are of unequal thickness, but only the right tail (70) is independent of the choice probabilities whereas the left tail (67) depends in general on the choice probabilities of all the other alternatives $j\neq i$, unless one of the V_j's tends towards infinity as in (68). Note that in the binary case $(M=2)$ where the S-DOGIT and G-DOGIT families are equivalent, the left tail is also independent of the choice probabilities:

$$\lim_{V_i\to\infty}P_i(V_i)_{GD}=\frac{\theta_{ij}}{1+\theta_{ij}}(i\neq j \text{ and } i,j=1,2). \tag{72}$$

3.2 Standard Box-Cox Model Types

The Standard Box-Cox Model Types of the DOGIT families only differ from the Standard Linear Model Types by relaxing the assumption (iv.a) to allow for the use of the Box-Cox transformations on the strictly positive independent variables X_{itk}'s included in each V_{it}: $V_{it}=\sum\limits_{k}\beta_k X_{itk}^{(\lambda_k)}$.

For the **S-DOGIT** family, the Box-Cox transformations can be fixed or estimated, but for the **G-DOGIT** family they can only be fixed due to the greater number of $\theta-$parameters involved.

The use of the Box-Cox transformation on an independent variable X_{ik} implies that:

(d.2) The response curve $P_i(X_{ik})_{SD}$ for the (S-BC) model type, is no longer symmetric as for the (S-LIN) model type since its slope is not in constant proportion to the slope of the curve $P_i(V_i)_L$:

$$\frac{\partial P_i(X_{ik})_{SD}}{\partial X_{ik}}=\beta_k X_{ik}^{\lambda_k-1}\frac{\partial P_i(V_i)_{SD}}{\partial V_i}=\frac{\beta_k X_{ik}^{\lambda_k-1}}{1+\sum\limits_{m}\theta_m}\frac{\partial P_i(V_i)_L}{\partial V_i}. \tag{73}$$

But like for the (S-LIN) model type, the curve is nondecreasing or nonincreasing if β_k is positive or negative respectively, and less steep than the curve $P_i(X_{ik})_L$ of the LOGIT (S-BC) model type since $\sum_m \theta_m$ is always positive.

The response curve $P_i(X_{ik})_{GD}$ for the (S-BC) model type remains asymmetric as for the (S-LIN) model type since its slope is not in constant proportion to the slope of the curve $P_i(V_i)_L$:

$$\frac{\partial P_i(X_{ik})_{GD}}{\partial X_{ik}} = \beta_k X_{ik}^{\lambda_k - 1} \frac{\partial P_i(V_i)_{GD}}{\partial V_i} = \beta_k X_{ik}^{\lambda_k - 1} \left[\frac{1 - P_i(V_i)_{GD} \sum_{m \neq i} \theta_{mi}}{1 + \sum_m \sum_{h \neq m} \theta_{mh} P_h(V_h)_L} \right] \frac{\partial P_i(V_i)_L}{\partial V_i} \quad (74)$$

Like for the (S-LIN) model type, the curve is nondecreasing or nonincreasing if β_k is positive or negative respectively, and less steep than the curve $P_i(X_{ik})_L$ of the LOGIT (S-BC) model type since the quotient in the square brackets is always less than unity.

3.3 Generalized Box-Cox Model Types

Like the LOGIT (G-BC) model type, the S-DOGIT (G-BC) and G-DOGIT (G-BC) model types allow a more general specification of the independent variables by relaxing in the (S-BC) model type of each family the assumption (iv.b) so that every V_{it} can include not only the network characteristics of its own alternative, but also those associated with some or all alternatives $j \neq i$.

Due to the presence of the extra parameters, θ_i in the S-DOGIT family and $\theta_{ij}(i \neq j)$ in the G-DOGIT family, these model types are still harder to estimate than the LOGIT (G-BC) model type which is already difficult to estimate due to the great number of λ-parameters.

The general specification of the independent variables implies that :

(d.2) The response curves $P_i(X_{\cdot k})_{SD}$ and $P_i(X_{\cdot k})_{GD}$ are asymmetric with respect to their inflection points, where $X_{\cdot k}$ represents the kth independent variable which appears in the ith alternative and also in some or all of the alternatives $j \neq i$.
Their slopes can be written as:

$$\begin{aligned} \frac{\partial P_i(X_{\cdot k})_{SD}}{\partial X_{\cdot k}} &= \frac{\partial P_i(V_i)_{SD}}{\partial V_i} \frac{\partial V_i}{X_{\cdot k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_{SD}}{\partial V_j} \frac{\partial V_j}{\partial X_{\cdot k}} \\ \frac{\partial P_i(X_{\cdot k})_{GD}}{\partial X_{\cdot k}} &= \frac{\partial P_i(V_i)_{GD}}{\partial V_i} \frac{\partial V_i}{X_{\cdot k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_{GD}}{\partial V_j} \frac{\partial V_j}{\partial X_{\cdot k}} \end{aligned} \quad (75)$$

where the probability curve of an ith alternative for each family, $P_i(\cdot)$, is considered in turn as a function of V_i and of each $V_j (j \neq i)$. Note that V_i is a function of $X_{\cdot k}$ whereas each V_j can be or not a function of $X_{\cdot k}$: $\partial V_i / \partial X_{\cdot k}$ is equal to $\beta_k^i X_{\cdot k}^{\lambda_k^i - 1}$ and $\partial V_j / \partial X_{\cdot k}$ is equal to $\beta_k^j X_{\cdot k}^{\lambda_k^i - 1}$ if V_j is a function of $X_{\cdot k}$, or zero otherwise.
Like for the LOGIT (G-BC) model type, these slopes can be nonnegative or nonpositive depending on the combination of the signs of the β-coefficients of $X_{\cdot k}$ in the different alternatives.

3.4 Specification of the Independent Variables

Since the numerator and the denominator in each probability $P_i(\cdot)$ of the S-DOGIT and G-DOGIT families are just different linear combinations of the exponentials of the V_j's $(j = 1, \ldots, M)$, the specification of the independent variables, i.e. the constant terms, the socioeconomic and network variables, for the different model types of the LOGIT family still holds for the S-DOGIT and G-DOGIT families (see Table 3).

4 IPT-L FAMILIES

4.1 Standard Linear Model Types

The Standard Linear Model Types of the **IPT-L** families are obtained by relaxing in the LINEAR CLASSICAL LOGIT:

(ii.a) **LIN-IPT-L**: the assumption (ii) to allow the estimation of the ϕ_i and μ_i-parameters in the outer function $f_i(\cdot)$, given the values of the $\tilde{\phi}_i$ and $\tilde{\mu}_i$-parameters fixed at one in the inner function $g_i(\cdot)$ for every alternative i:

$$P_i(V_i) = \frac{[\phi_i \exp(V_i) + 1]^{1/\phi_i} - \mu_i}{\sum_m \left\{[\phi_m \exp(V_m) + 1]^{1/\phi_m} - \mu_m\right\}}, (\phi_i \geq 0 \text{ and } \mu_i \leq 1), \tag{76}$$

(iii.a) **BT-IPT-L**: the assumption (iii) to allow the estimation of the $\tilde{\phi}_i$ and $\tilde{\mu}_i$-parameters in the inner function $g_i(\cdot)$, given the values of the ϕ_i and μ_i-parameters fixed at zero in the outer function $f_i(\cdot)$ for every alternative i:

$$P_i(V_i) = \frac{\exp\left\{\dfrac{[\exp(V_i) + \tilde{\mu}_i]^{\tilde{\phi}_i} - 1}{\tilde{\phi}_i}\right\}}{\sum_m \exp\left\{\dfrac{[\exp(V_i) + \tilde{\mu}_m]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m}\right\}}, (\tilde{\mu}_i \geq 0). \tag{77}$$

LIN-IPT-L

In contrast with the LINEAR CLASSICAL LOGIT, the symmetry property **(d)** and the thin-equal-tails property **(e)** no longer hold:

(d.1) The probability curve of the LIN-IPT-L family, $P_i(V_i)_{LIL}$, is not symmetric with respect to its inflection point, since its slope is not in constant proportion to the slope of the symmetric LINEAR LOGIT curve $P_i(V_i)_L$ over the whole range $-\infty < V_i < +\infty$:

$$\frac{\partial P_i(V_i)_{LIL}}{\partial V_i} = A_i^{\frac{1}{\phi_i}-1} \left[\frac{\sum_m \exp(V_m)}{\sum_m \left(A_m^{1/\phi_m} - \mu_m\right)}\right]^2 \left[\frac{\sum_{m \neq i} \left(A_m^{1/\phi_m} - \mu_m\right)}{\sum_{m \neq i} \exp(V_m)}\right] \frac{\partial P_i(V_i)_L}{\partial V_i}. \tag{78}$$

For a given set of ϕ and μ-values, the first two factors of $\partial P_i(V_i)_L / \partial V_i$ are highly nonlinear in V_i, hence cannot be constant over the whole range of V_i.

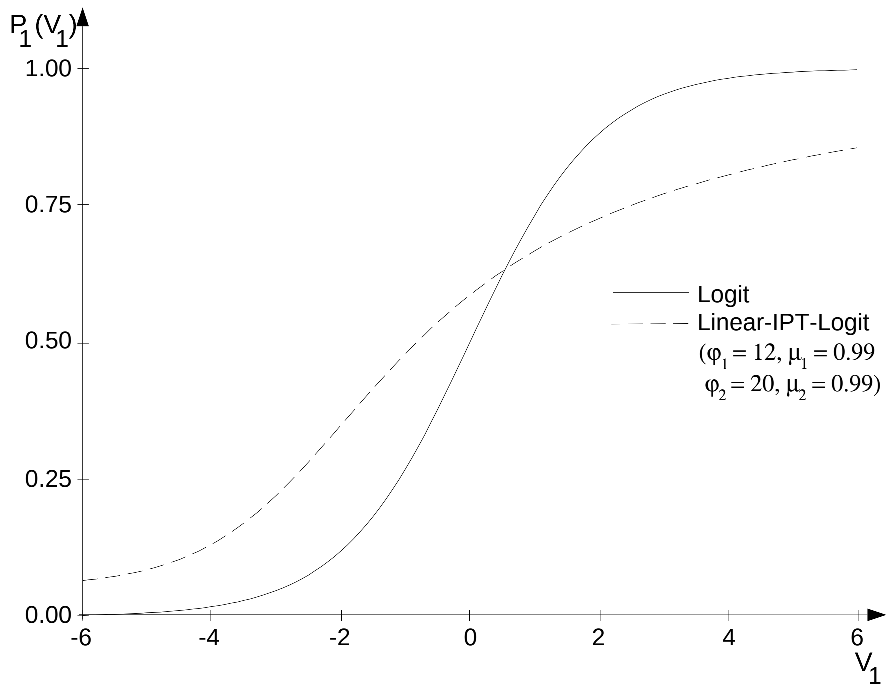
(d.2) The probability curve $P_i(X_{ik})_{LIL}$ in function of an independent variable X_{ik} is also asymmetric with respect to its inflection point since its slope is just in constant proportion to the slope of the asymmetric probability curve $P_i(V_i)_{LIL}$:

$$\frac{\partial P_i(X_{ik})_{LIL}}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i} \frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i} \beta_k \tag{79}$$

which can be nonnegative or nonpositive if β_k is positive or negative respectively.

(e) In general, the two tails of the probability curve $P_i(V_i)_{LIL}$ are asymmetric, with a left tail which may or not disappear as V_i tends towards $-\infty$ in contrast to the right tail which always vanishes as V_i tends towards $+\infty$. Figure 3 shows the probability curves $P_1(V_1)_L$ and $P_1(V_1)_{LIL}$ given $V_2 = 0$ with $(\phi_1, \mu_1) = (12, 0.99)$ and $(\phi_2, \mu_2) = (20, 0.99)$ for the binary case. Note that each curve is invariant with respect to the model types of the family since it depends on the V_m's only.

Figure 3 LOGIT(family) versus LIN-IPT-L(family)



- **Left tail**

  The left tail can be written as:

$$\lim_{V_i \to -\infty} P_i(V_i)_{LIL} = \frac{1 - \mu_i}{(1 - \mu_i) + \sum\limits_{m \neq i} \left\{ [\phi_m \exp(V_m) + 1]^{1/\phi_m} - \mu_m \right\}}. \tag{80}$$

  Two distinct cases must be considered:

  - Every $V_m (m \neq i)$ has a finite value or some V_m's can tend towards $-\infty$, but no single V_m is allowed to tend towards $+\infty$: the left tail will not disappear as V_i tends towards $-\infty$, except if μ_i is equal to one.
  - One of the V_m's $(m \neq i)$ is allowed to tend towards $+\infty$: the left tail will disappear like in the LINEAR CLASSICAL LOGIT, although at a different rate as indicated by the slope $\partial P_i(V_i)_{LIL} / \partial V_i$.

- **Right tail**

  Due to the constraint $\phi_i \geq 0$, the right tail always vanishes as V_i tends towards $+\infty$ like in the LINEAR CLASSICAL LOGIT, although at a different rate:

$$\lim_{V_i \to +\infty} P_i(V_i)_{LIL} = \frac{\lim\limits_{V_i \to +\infty} \dfrac{\partial}{\partial V_i} \left\{ [\phi_i \exp(V_i) + 1]^{1/\phi_i} - \mu_i \right\}}{\lim\limits_{V_i \to +\infty} \dfrac{\partial}{\partial V_i} \left\{ \sum\limits_m \left\{ [\phi_m \exp(V_m) + 1]^{1/\phi_m} - \mu_m \right\} \right\}} = 1. \tag{81}$$

• **Interpretation of ϕ_i and μ_i**

The role of μ_i can be best understood if every ϕ_i is set equal to 1:

$$
\begin{aligned}
P_i(V_i)_{LIL} &= \frac{\exp(V_i) + (1-\mu_i)}{\sum\limits_m [\exp(V_m) + (1-\mu_m)]} \\
&= \frac{R_i \exp(V_i)}{\sum\limits_m R_m \exp(V_m)} = \frac{\exp(V_i + \ln R_i)}{\sum\limits_m \exp(V_m + \ln R_m)}
\end{aligned}
\tag{82}
$$

where $R_i = 1 + (1-\mu_i)\exp(-V_i)$.

Although this form of $P_i(V_i)_{LIL}$ is analogous to that of $P_i(V_i)_{GD}$, R_i cannot be interpreted as a similarity / dissimilarity measure between the attributes X_{ik}'s and the other X_{jk}'s across the alternatives, since it is only a function of the attributes of its own alternative. But the total shift in V_i represented by $\ln R_i$ can be clearly seen through this form:

1. In the absence of information on V_i, i.e. $V_i = 0$, the shift is just equal to $\ln(2-\mu_i)$ which is always nonnegative due to the constraint $\mu_i \leq 1$. In this respect, μ_i can be interpreted as the "*a priori* ignorance" level of the modeler, or as a captivity level, which is gradually modified as he introduces some information on V_i.
2. For positive values of V_i, the shift will be less than $\ln(2-\mu_i)$ and will disappear as V_i tends towards ∞, whereas for negative values of V_i, it will be greater than $\ln(2-\mu_i)$ and will not disappear as V_i tends towards $-\infty$, unless one of the other V_j's tends towards ∞. In this sense, the shift can be interpreted as the portion left unexplained that the alternative specificity represented by V_i fails to capture.

The role of ϕ_i is to allow different functional forms for the term $\exp(V_i)$ including the linear form which corresponds to $\phi_i = 1$, due to errors of specification of the representative utility functions or incomplete descriptions of consumer characteristics.

Since the function $[\phi_i \exp(V_i) + 1]^{1/\phi_i}$ is inversely proportional to ϕ_i due to the dominant inverse power of ϕ_i relatively to the multiplicative factor ϕ_i in front of $\exp(V_i)$, smaller positive values of ϕ_i will have a greater effect on $\exp(V_i)$, whereas very large positive values of ϕ_i will reduce $\exp(V_i)$ almost to a constant. As a limit case, if ϕ_i tends towards zero, the function tends to the limit value $\exp[\exp(V_i)]$, which is clearly much higher than $[\exp(V_i) + 1]$ which corresponds to $\phi_i = 1$, for positive values of V_i, and much less than $[\exp(V_i) + 1]$ for negative values of V_i. As a consequence, μ_i will have a much greater effect for the negative portion of V_i than for the positive portion for which $\exp[\exp(V_i)]$ will clearly dominate μ_i, even for relatively small positive values of V_i such as 8. Inversely for large positive values of ϕ_i, μ_i will have a more even effect on both portions of negative and positive values of V_i, but the influence of V_i, hence the attributes of the alternative, will be greatly diminished, i.e. large fluctuations in V_i will not affect too much the level of the function $[\phi_i \exp(V_i) + 1]^{1/\phi_i}$.

BT-IPT-L

Like the LIN-IPT-L family, the symmetric property (**d**) and the thin-equal-tails property (**e**) in the LINEAR CLASSICAL LOGIT no longer hold for the BT-IPT-L family:

(d.1) The probability curve $P_i(V_i)_{BIL}$ in function of V_i is not symmetric with respect to its inflection point, since its slope is not in constant proportion to the slope of the symmetric LINEAR LOGIT curve $P_i(V_i)_L$:

$$\frac{\partial P_i(V_i)_{BIL}}{\partial V_i} = B_i^{\tilde{\phi}_i-1}\exp\left[B_i^{(\tilde{\phi}_i)}\right]\left[\frac{\sum\limits_m \exp(V_m)}{\sum\limits_m \exp\left[B_m^{(\tilde{\phi}_m)}\right]}\right]^2\left[\frac{\sum\limits_{m\neq i}\exp\left[B_m^{(\tilde{\phi}_m)}\right]}{\sum\limits_{m\neq i}\exp(V_m)}\right]\frac{\partial P_i(V_i)_L}{\partial V_i} \qquad (83)$$

which indicates that for a given set of $\tilde{\phi}$ and $\tilde{\mu}-$values, hence cannot be constant over the whole range $-\infty < V_i < +\infty$.

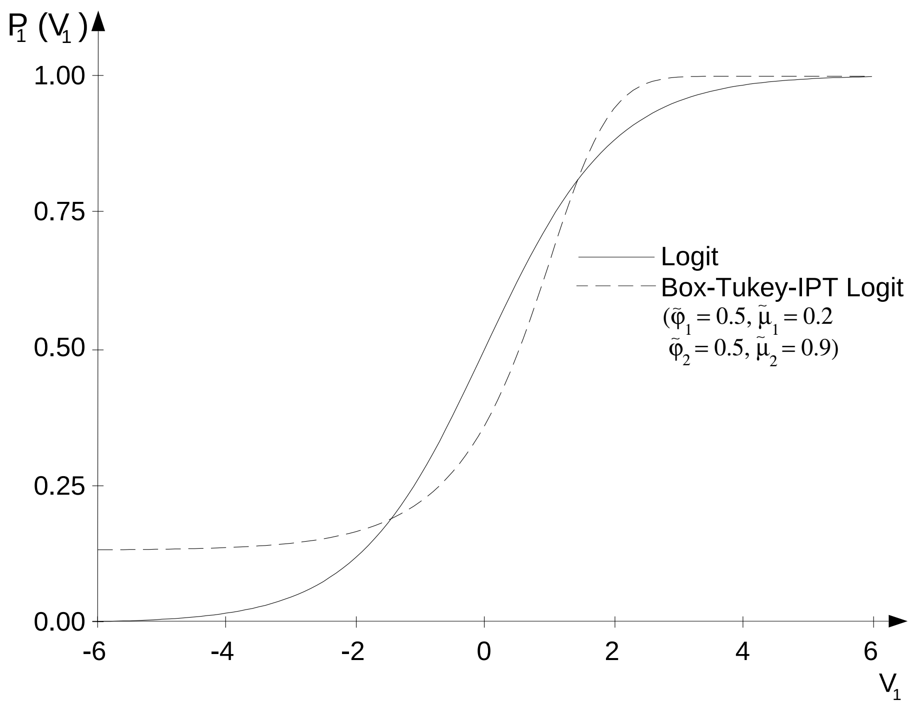
(d.2) The probability curve $P_i(X_{ik})_{BIL}$ in function of X_{ik} is also asymmetric with respect to its inflection point for the same reason. Its slope can be written as:

$$\frac{\partial P_i(X_{ik})_{BIL}}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\beta_k \qquad (84)$$

which can be nonnegative or nonpositive if β_k is positive or negative respectively.

(e) In general, the two tails of the probability curve $P_i(V_i)_{BIL}$ are asymmetric with a left (or right) tail which may or not disappear as V_i tends towards $-\infty$(or $+\infty$). Figure 4 shows the probability curves $P_1(V_1)_L$ and $P_1(V_1)_{BIL}$ given $V_2 = 0$ with $\left(\tilde{\phi}_1, \tilde{\mu}_1\right) = (0.5, 0.2)$ and $\left(\tilde{\phi}_2, \tilde{\mu}_2\right) = (0.5, 0.9)$ for the binary case. Note that each curve is invariant with respect to the model types of the family since it depends on the V_m's only.

Figure 4 LOGIT(family) versus BT-IPT-L(family)



- **Left tail**

  The left tail can be written as:

$$\lim_{V_i\to-\infty} P_i(V_i)_{BIL} = \frac{\exp\left(\dfrac{\tilde{\mu}_i^{\tilde{\phi}_i}-1}{\tilde{\phi}_i}\right)}{\exp\left(\dfrac{\tilde{\mu}_i^{\tilde{\phi}_i}-1}{\tilde{\phi}_i}\right) + \sum\limits_{m\neq i}\left\{\dfrac{\left[\exp(V_m)+\tilde{\mu}_m\right]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}}. \qquad (85)$$

Two distinct cases must be considered:

- Every $V_m (m \neq i)$ has a finite value or some V_m's can tend towards $-\infty$, but no single V_m is allowed to tend towards $+\infty$: the left tail will not disappear as V_i tends towards $-\infty$ for finite values of $\tilde{\phi}_i$ and $\tilde{\mu}_i$.
- One of the V_m's $(m \neq i)$ is allowed to tend towards $+\infty$ with $\tilde{\phi}_m > 0$: the left tail will disappear like in the LINEAR CLASSICAL LOGIT, although at a different rate as indicated by the slope $\partial P_i(V_i)_{BIL}/\partial V_i$.

- **Right tail**
  Depending on the sign of $\tilde{\phi}_i$, two cases have to be considered:

  - If $\tilde{\phi}_i$ is positive or zero, the right tail will vanish as V_i tends towards $+\infty$ like in the LINEAR CLASSICAL LOGIT, although at a different rate:

$$\lim_{V_i \to +\infty} P_i(V_i)_{BIL} = \frac{\displaystyle\lim_{V_i \to +\infty} \frac{\partial}{\partial V_i} \left\{ \exp \left[\frac{(\exp(V_i) + \tilde{\mu}_j)^{\tilde{\phi}_i} - 1}{\tilde{\phi}_i} \right] \right\}}{\displaystyle\lim_{V_i \to +\infty} \frac{\partial}{\partial V_i} \left\{ \sum_m \exp \left[\frac{(\exp(V_m) + \tilde{\mu}_m)^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right] \right\}} = 1. \tag{86}$$

  - If $\tilde{\phi}_i$ is negative, the right tail can be written as:

$$\lim_{V_i \to +\infty} P_i(V_i)_{BIL} = \frac{\exp\left(-\dfrac{1}{\tilde{\phi}_i}\right)}{\exp\left(-\dfrac{1}{\tilde{\phi}_i}\right) + \displaystyle\sum_{m \neq i} \exp\left\{ \frac{[\exp(V_m) + \tilde{\mu}_m]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right\}} \tag{87}$$

    which indicates that the right tail will not disappear for the finite values of $\tilde{\phi}_i$, unless $\tilde{\phi}_i$ is allowed to tend towards $-\infty$.

- **Interpretation of $\tilde{\phi}_i$ and $\tilde{\mu}_i$.**

Analogously to the LIN-IPT-L family, the role of $\tilde{\mu}_i$ can be best understood if every $\tilde{\phi}_i$ is set equal to 0:

$$\begin{aligned} P_i(V_i)_{BIL} &= \frac{\exp\{\ln[\exp(V_i) + \tilde{\mu}_i]\}}{\sum_m \exp\{\ln[\exp(V_m) + \tilde{\mu}_m]\}} = \frac{\exp(V_i) + \tilde{\mu}_i}{\sum_m [\exp(V_m) + \tilde{\mu}_m]} \\ &= \frac{R_i \exp(V_i)}{\sum_m R_m \exp(V_m)} = \frac{\exp(V_i + \ln R_i)}{\sum_m \exp(V_m + \ln R_m)} \end{aligned} \tag{88}$$

where $R_i = 1 + \tilde{\mu}_i \exp(-V_i)$.

In this special case, the BT-IPT-L and LIN-IPT-L families are exactly equivalent since $\tilde{\mu}_i$ is equal to $(1 - \mu_i)$. The total shift in V_i represented by $\ln R_i$ can be similarly distinguished:

1. In the absence of information on V_i, i.e. $V_i = 0$, the shift is just equal to $\ln(1 + \tilde{\mu}_i)$ which is always nonnegative due to the constraint $\tilde{\mu}_i \geq 0$. In this respect, $\tilde{\mu}_i$ can be similarly interpreted as in LIN-IPT-L: it represents the "*a priori* ignorance" of the modeler or a captivity level which is gradually modified as he incorporates some information on V_i.
2. For positive values of V_i, the shift will be less than $\ln(1 + \tilde{\mu}_i)$ whereas for negative values of V_i, it will be greater than $\ln(1 + \tilde{\mu}_i)$. In this sense, the shift can be interpreted as the portion left unexplained that the alternative specificity represented by V_i fails to capture.

But unlike the LIN-IPT-L family, the role of $\tilde{\phi}_i$ is to allow different functional forms for the term $[\exp(V_i) + \tilde{\mu}_i]$ instead of $\exp(V_i)$ alone, in order to take into account the errors of specification of the representative utility functions or incomplete descriptions of consumer characteristics.

4.2 Generalized Linear Model Types

These model types are more general than the STANDARD LINEAR model types. Due to the nonlinear transformations on all the exponentials of V_m's $(m = 1, \ldots, M)$ in each probability function $P_i(\cdot)$, namely the inverse and direct Box-Tukey transformations for LIN-IPT-L and BT-IPT-L respectively, different specification contraints on the independent variables of the LINEAR CLASSICAL LOGIT can be relaxed in these model types without creating any identification problem for the β-coefficients, provided that at least $M-1$ ϕ-parameters of LIN-IPT-L be different from 1, and at least $M-1$ $\tilde{\phi}$-parameters of BT-IPT-L be different from 0:

1. The assumption **(iv.b)** is modified so that a network characteristic of an ith alternative can now be specified in a linear form not only in its own alternative, but also across some or all alternatives.
2. The underidentification problem of the mode-specific effects **(f)** and **(g)** which exists in the LINEAR CLASSICAL LOGIT can be dealt with by specifying in all alternatives:

   (f) A socioeconomic variable or a network characteristic of an ith alternative in a linear form.
   (g) The alternative-specific constants.

The main implication of (1) and (2) is that the LINEAR CLASSICAL LOGIT is not nested in these model types, since for the β-coefficients to be identified, the ϕ-parameters cannot be all equal to 1 for LIN-IPT-L and the $\tilde{\phi}$-parameters cannot be all equal to 0 for BT-IPT-L, even though these parameter contraints can be relaxed for the special case where all the socioeconomic variables and the network characteristics of an ith alternative appear only in $M-1$ alternatives at most and that the alternative-specific constants appear in $M-1$ alternatives exactly, since by definition a network characteristic of an ith alternative can only be specified in its own alternative in the LINEAR CLASSICAL LOGIT. Table 4 summarizes the special contraints on the extra parameters for the Generalized Linear Model Types of LIN-IPT-L and BT-IPT-L required for the estimation of the β-coefficients.

Note in particular that for assumption 2.g under which the alternative-specific constants appear in all alternatives, this specification cannot include the LINEAR CLASSICAL LOGIT since in the latter case, only $M-1$ specific constants can be identified whereas in the LIN-IPT-L and BT-IPT-L cases, M specific constants are always identified provided that at least $M-1$ ϕ-parameters in LIN-IPT-L be different from 1, and at least $M-1$ $\tilde{\phi}$-parameters in BT-IPT-L be different from 0.

LIN-IPT-L

Like for the Standard Linear Model Type, the symmetry property **(d)** and the thin-equals-tails property **(e)** of the LINEAR CLASSICAL LOGIT no longer hold for the Generalized Linear Model Type:

(d.1) The probability curve of the LIN-IPT-L family, $P_i(V_i)_{LIL}$, is not symmetric with respect to its inflection point. Since the LINEAR CLASSICAL LOGIT is not nested in this model type, the slope of the curve cannot be expressed in terms of the slope of the probability curve of the LOGIT

TABLE 4 Special constraints on the extra parameters for the Generalized Linear Model Types of **LIN-IPT-L** and **BT-IPT-L**.

VARIABLE	SPECIAL CONTRAINTS ON THE EXTRA PARAMETERS	
LIN-IPT-L	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
If the CONSTANTS are present	• in $M-1$ alternatives exactly : No contraints	Not allowed
	• in all M alternatives: at least $M-1$ ϕ-parameters should be different from 1.	Same as left column
If the SOCIOECONOMIC and NETWORK variables are present	• in some alternatives : No constraints	Same as left column
	• in all M alternatives: at least $M-1$ ϕ-parameters should be different from 1.	Same as left column
BT-IPT-L	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
If the CONSTANTS are present	• in $M-1$ alternatives exactly : No contraints	Not allowed
	• in all M alternatives: at least $M-1$ $\tilde{\phi}$-parameters should be different from 0.	Same as left column
If the SOCIOECONOMIC and NETWORK variables are present	• in some alternatives : No constraints	Same as left column
	• in all M alternatives: at least $M-1$ $\tilde{\phi}$-parameters should be different from 0.	Same as left column

family $P_i(V_i)_L$:

$$\frac{\partial P_i(V_i)_{LIL}}{\partial V_i} = \frac{A_i^{\frac{1}{\phi_i}-1}[1-P_i(V_i)_{LIL}]\exp(V_i)}{\sum_m \left(A_m^{\frac{1}{\phi_m}} - \mu_m\right)}. \tag{89}$$

(d.2) The probability curve $P_i(X_{\cdot k})_{LIL}$ in function of an independent variable $X_{\cdot k}$ which appears both in the ith alternative and across some or all alternatives is also asymmetric with respect to its inflection point:

$$\frac{\partial P_i(X_{\cdot k})_{LIL}}{\partial X_{\cdot k}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{\cdot k}} + \sum_{j\neq i}\frac{\partial P_i(V_j)_{LIL}}{\partial V_j}\frac{\partial V_j}{\partial X_{\cdot k}} \tag{90}$$

where $\partial V_i/\partial X_{.k}$ is equal to β_k^i and $\partial V_j/\partial X_{.k}$ is equal to β_k^j if V_j is function of $X_{.k}$, or zero otherwise.

(e) The asymmetric-tails property of LIN-IPT-L (S-LIN) still holds for LIN-IPT-L (G-LIN).

BT-IPT-L

Similarly, the symmetry property **(d)** and the thin-equals-tails property **(e)** of the LINEAR CLASSICAL LOGIT no longer hold for the Generalized Linear Model Type:

(d.1) The probability curve of the BT-IPT-L family, $P_i(V_i)_{BIL}$, is not symmetric with respect to its inflection point. Since the LINEAR CLASSICAL LOGIT is not nested in this model type, the slope of the curve cannot be expressed in terms of the slope of the probabilitty curve of the LOGIT family $P_i(V_i)_L$:

$$\frac{\partial P_i(V_i)_{BIL}}{\partial V_i} = B_i^{\tilde{\phi}_i - 1} P_i(V_i)_{BIL}[1 - P_i(V_i)_{BIL}] \exp(V_i). \tag{91}$$

(d.2) The probability curve $P_i(X_{.k})_{BIL}$ in function of an independent variable $X_{.k}$ which appears both in the ith alternative and across some or all alternatives is also asymmetric with respect to its inflection point:

$$\frac{\partial P_i(X_{.k})_{BIL}}{\partial X_{.k}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{.k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_{BIL}}{\partial V_j}\frac{\partial V_j}{\partial X_{.k}} \tag{92}$$

where $\partial V_i/\partial X_{.k}$ is equal to β_k^i and $\partial V_j/\partial X_{.k}$ is equal to β_k^j if V_j is function of $X_{.k}$, or zero otherwise.

(e) The asymmetric-tails property of BT-IPT-L (S-LIN) still holds for BT-IPT-L (G-LIN).

4.3 Standard Box-Cox Model Types

The standard Box-Cox Model Types of the IPT-L families only differ from the Standard Linear Model Types by relaxing one additional assumption (**iv.a**) of the LINEAR CLASSICAL LOGIT to allow for the use of the Box-Cox transformations on the strictly independent variables X_{itk}'s included in each $V_{it} : V_{it} = \sum_k \beta_k X_{itk}^{(\lambda_k)}$.

The use of the Box-Cox transformation on an independent variable X_{ik} implies that:

(d.2) The response curve $P_i(X_{ik})_{LIL}$ for the (S-BC) model type remains asymmetric with respect to its inflection point as for the (S-LIN) model type. The slope of $P_i(X_{ik})_{LIL}$ can be written as:

$$\frac{\partial P_i(X_{ik})_{LIL}}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i}\beta_k X_{ik}^{\lambda_k - 1} \tag{93}$$

where $\partial P_i(V_i)_{LIL}/\partial V_i$ is already given in (78) and $\partial V_i/\partial X_{ik}$ is equal to $\beta_k X_{ik}^{\lambda_k - 1}$.
Since this slope can be positive or negative depending on the sign of β_k, the curve $P_i(X_{ik})_{LIL}$ can be nondecreasing or nonincreasing respectively. Moreover the slope depends also on $X_{ik}^{\lambda_k - 1}$ which is not present in the (S-LIN) model type, hence the effect on V_i due to a variation of X_{ik}, $\partial V_i/\partial X_{ik}$, is identical to the one already analyzed for the LOGIT (S-BC) model type.
The response curve $P_i(X_{ik})_{BIL}$ for the (S-BC) model type also remains asymmetric with respect to its inflection point as for the (S-LIN) model type. The slope of $P_i(X_{ik})_{BIL}$ has the following form:

$$\frac{\partial P_i(X_{ik})_{BIL}}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\beta_k X_{ik}^{\lambda_k - 1} \tag{94}$$

where $\partial P_i(V_i)_{BIL}/\partial V_i$ is already given in (83) and $\partial V_i/\partial X_{ik}$ is equal to $\beta_k X_{ik}^{\lambda_k - 1}$.
The general characteristics of the slope of $P_i(X_{ik})_{BIL}$ are similar to those of the slope of $P_i(X_{ik})_{LIL}$.

4.4 Generalized Box-Cox Model Types

The Generalized Box-Cox Model Types of the IPT-L families only differ from the Generalized Linear Model Types by relaxing one additional assumption (**iv.a**) of the LINEAR CLASSICAL LOGIT to allow for the use of the Box-Cox transformations on the strictly independent variables X_{itk}'s included in each $V_{it}\,:\,V_{it}\,=\,\sum_k \beta_k X_{itk}^{(\lambda_k)}$.

The use of the Box-Cox transformation on an independent variable $X_{\cdot k}$ which appears not only in the ith alternative but also across some or all alternatives implies that:

(d.2) The response curve $P_i(X_{\cdot k})_{LIL}$ for the (G-BC) model type remains asymmetric with respect to its inflection point as for the (G-LIN) model type. The slope of $P_i(X_{\cdot k})_{LIL}$ can be written as:

$$\frac{\partial P_i(X_{\cdot k})_{LIL}}{\partial X_{\cdot k}} = \frac{\partial P_i(V_i)_{LIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{\cdot k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_{LIL}}{\partial V_j}\frac{\partial V_j}{\partial X_{\cdot k}} \tag{95}$$

which is similar to the slope associated with the (G-LIN) model type, except for the effects on V_i and the V_j's due to a change of $X_{\cdot k}$ which now depend on the level of $X_{\cdot k}$: $\partial V_i/\partial X_{\cdot k} = \beta_k^i X_{\cdot k}^{\lambda_k^i - 1}$ and $\partial V_j/\partial X_{\cdot k} = \beta_k^j X_{\cdot k}^{\lambda_k^j - 1}$ if V_j is function of $X_{\cdot k}$, and zero otherwise.
The total effect on $P_i(X_{\cdot k})_{LIL}$ due to a variation of $X_{\cdot k}$ is hard to predict since it depends not only on the signs of the β_k-coefficients associated with $X_{\cdot k}$ across the alternatives, but also on the level of $X_{\cdot k}$ itself and on the signs of the λ_k-parameters across the alternatives.
Similarly, the response curve $P_i(X_{\cdot k})_{BIL}$ for the (G-BC) model type remains asymmetric with respect to its inflection point as for the (G-LIN) model type. The slope of $P_i(X_{\cdot k})_{BIL}$ has the following form:

$$\frac{\partial P_i(X_{\cdot k})_{BIL}}{\partial X_{\cdot k}} = \frac{\partial P_i(V_i)_{BIL}}{\partial V_i}\frac{\partial V_i}{\partial X_{\cdot k}} + \sum_{j \neq i} \frac{\partial P_i(V_j)_{BIL}}{\partial V_j}\frac{\partial V_j}{\partial X_{\cdot k}} \tag{96}$$

which is analogous to the slope associated with the (G-LIN) model type, except for $\partial V_i/\partial X_{\cdot k}$ and $\partial V_j/\partial X_{\cdot k}$ which are identical to those associated with the LIN-IPT-L (G-BC) model type.

Again, the total effect on $P_i(X_{\cdot k})_{BIL}$ due to a variation of $X_{\cdot k}$ is difficult to evaluate for the same reasons as for the LIN-IPT-L (G-BC) model type.

(f) For a socioeconomic or network variable, the mode-specific effect can be identified by specifying the same variable in all alternatives:

1. With a common λ-parameter across the alternatives, hence at least $M-1$ ϕ-parameters should be different from 1 for LIN-IPT-L (G-BC) and at least $M-1$ $\tilde{\phi}$–parameters should be different from 0 for BT-IPT-L (G-BC) so that the β-coefficients across the alternatives remain identified.
2. With distinct λ-parameters across the alternatives, hence the special constraints on the extra parameters can be removed since the β-coefficients of the variable across the alternatives are always identified.

(g) Like for the Generalized Linear Model Types, the mode-specific effect obtained by specifying on the alternative-specific constant for every alternative can be only be identified:

1. If at least $M - 1$ ϕ-parameters are different from 1 for LIN-IPT-L (G-BC).
2. If at least $M - 1$ $\tilde{\phi}$-parameters are different from 0 for BT-IPT-L (G-BC).

Table 5 summarizes the special constraints on the extra parameters for LIN-IPT-L (G-BC) and BT-IPT-L (G-BC) required for the identification of the mode-specific effects of the constant and the socioeconomic and network variables. Note that for each family, detailed analysis of the subcases I and II concerning the specification of a network variable is not given as it was done for LOGIT (G-BC) (see Table 3). As a general guideline, these subcases still hold for the two IPT-L families. But due to the nonlinearities introduced by the direct and inverse Box-Tukey transformation on every exponential of V_i, there would be other less constrained specifications which could be used in addition to the identified subcases of LOGIT (G-BC).

4.5 Specification of the Independent Variables

Except for the Generalized Linear and Box-Cox model types which are already reported in Tables 4 and 5 respectively due to the special contraints on the extra parameters for the identification of the β-coefficients associated with the constant terms and the socioeconomic and network variables, the specification of these variables for the Standard Linear and Box-Cox model types remains the same as for the Standard Linear and Box-Cox model types of the LOGIT family respectively (see Table 3).

TABLE 5 Special constraints on the extra parameters for the Generalized Box-Cox Model Types of LIN-IPT-L and BT-IPT-L.

VARIABLE	SPECIAL CONTRAINTS ON THE EXTRA PARAMETERS	
LIN-IPT-L	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
If the CONSTANTS are present	• in $M-1$ alternatives exactly : No contraints	Not allowed
	• in all M alternatives: at least $M-1$ ϕ-parameters should be different from 1.	Not allowed
If the SOCIOECONOMIC and NETWORK variables are present	• in some alternatives : No constraints	Same as left column
	• in all M alternatives: 1. With a common λ-parameter across the alternatives, at least $M-1$ ϕ-parameters should be different from 1.	Same as left column
	2. With distinct λ-parameters across the alternatives, no contraints.	Not allowed
BT-IPT-L	ALTERNATIVE-SPECIFIC CASE	GENERIC CASE
If the CONSTANTS are present	• in $M-1$ alternatives exactly : No contraints	Not allowed
	• in all M alternatives: at least $M-1$ $\tilde{\phi}$-parameters should be different from 0.	Not allowed
If the SOCIOECONOMIC and NETWORK variables are present	• in some alternatives : No constraints	Same as left column
	• in all M alternatives: 1. With a common λ-parameter across the alternatives, at least $M-1$ $\tilde{\phi}$-parameters should be different from 0.	Same as left column
	2. With distinct λ-parameters across the alternatives, no contraints.	Not allowed

5 ESTIMATION RESULTS

5.1 Elasticity and Marginal Rate of Substitution (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:

$$E^{\bar{P}_i}_{X_{jk}} = \sum_n P_{in} E^{P_{in}}_{X_{jnk}} / \sum_n P_{in} \tag{97}$$

where $E^{P_{in}}_{X_{jnk}}$ 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:

$$E^{P_{in}}_{X_{jnk}} = \frac{\partial P_{in}}{\partial X_{jnk}} \frac{X_{jnk}}{P_{in}}. \tag{98}$$

The weighted aggregate elasticity (97) can be viewed as evaluating (98) for each individual n, then aggregating by weighting each point elasticity by the individual's estimated choice probability.

In (98), the derivative of P_{in} with respect to X_{jnk} has the general form:

$$\frac{\partial P_{in}}{\partial X_{jnk}} = \frac{1}{\sum\limits_m \mathcal{U}_{mn}} \left[\frac{\partial \mathcal{U}_{in}}{\partial X_{jnk}} - P_{in} \sum_m \frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} \right] \quad . \tag{99}$$

An explicit form of $\partial \mathcal{U}_{mn} / \partial X_{jnk}$ can be written for each family:

LOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} = D^{mn}_{jnk} \tag{100}$$

S-DOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} = D^{mn}_{jnk} + \theta_m \sum_h D^{hn}_{jnk} \tag{101}$$

G-DOGIT

$$\frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} = D^{mn}_{jnk} + \sum_{h \neq m} \theta_{mh} D^{hn}_{jnk} \tag{102}$$

LIN-IPT-L

$$\frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} = A_{mn}^{\frac{1}{\phi_m}-1} D_{jnk}^{mn} \tag{103}$$

BT-IPT-L

$$\frac{\partial \mathcal{U}_{mn}}{\partial X_{jnk}} = \mathcal{U}_{mn} B_{mn}^{\tilde{\phi}_m-1} D_{jnk}^{mn} \tag{104}$$

where

$$D_{jnk}^{mn} = \frac{\partial \exp\left(V_{mn}\right)}{\partial X_{jnk}} = \exp\left(V_{mn}\right) \frac{\partial V_{mn}}{\partial X_{jnk}} \tag{105}$$

and

$$\frac{\partial V_{mn}}{\partial X_{jnk}} = \begin{cases} 0 & \text{if } X_{jnk} \text{ is not present in alternative } m \neq j, \\ \beta_r X_{mnr}^{\lambda_r - 1} & \text{if } X_{jnk} \text{ is alternative-specific and also appears} \\ & \text{in alternative } m \neq j, \text{ with a coefficient } \beta_r \neq \beta_k, \\ \beta_k X_{mnk}^{\lambda_k - 1} & \text{if } X_{jnk} \text{ is generic and also appears in alternative} \\ & m \neq j, \text{ with a same coefficient } \beta_k, \\ \beta_k X_{jnk}^{\lambda_k - 1} & \text{if } m = j \text{ and } X_{jnk} \text{ is either generic or alternative-specific.} \end{cases} \tag{106}$$

Note that the explicit form of $\partial V_{mn}/\partial X_{jnk}$ corresponds to the case where an independent variable may appear across the alternatives but only once in each of them. The program can also handle the more complex case where the variable is present across several alternatives in each of which it may be specified as two or more independent variables with distinct λ-parameters.

• **Components of the weighted aggregate elasticity: probability point and partial derivative**

Two components of the weighted aggregate elasticity (97) are also computed:

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

$$C_{X_{jk}}^{\bar{P}_i} = \sum_n P_{in} \left[\frac{\partial P_{in}}{\partial X_{jnk}} X_{jnk} \right] / \sum_n P_{in} \tag{107}$$

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

$$D_{X_{jk}}^{\bar{P}_i} = \sum_n P_{in} \left[\frac{\partial P_{in}}{\partial X_{jnk}} \right] / \sum_n P_{in}, \tag{108}$$

where $\partial P_{in}/\partial X_{jnk}$ is given in (99).

• **LOGIT: Sign of the weighted aggregate elasticity**

Among the five families, only the **LOGIT** family has a built-in property based on the **Independence from Irrelevant Alternatives** (**IIA**) axiom 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.

Case where the variable appears in one alternative only

In this case, the point elasticity (98) can be expressed as:

$$E_{X_{jnk}}^{P_{in}} = \beta_k [\delta_{ij} - P_{jn}] X_{jnk}^{\lambda_k} \tag{109}$$

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_{in}, they are uniform, i.e. equal across the alternatives $i \neq j$, which is one of the implications of the **IIA** property

The weighted aggregate elasticity (97) can be written as:

$$E_{X_{jk}}^{\bar{P}_i} = \beta_k \sum_n P_{in} [\delta_{ij} - P_{jn}] X_{jnk}^{\lambda_k} / \sum_n P_{in} \tag{110}$$

which depends on both P_{in} and P_{jn} 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 (110) 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 (110) 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 (110) 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 LOGIT Model Type.

The point elasticity (98) can now be expressed as:

$$E^{P_{in}}_{X_{jnk}} = \left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_m P_{mn}\frac{\partial V_{mn}}{\partial X_{jnk}}\right] X_{jnk} \tag{111}$$

where $\partial V_{mn}/\partial X_{jnk}$ is given in (106)

Note that in the explicit expression of this derivative, each V_{mn} is assumed to include only one X_{mnk} if any. In the case where the variable appears as two independent variables with distinct functional forms in the same alternative m, 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 (97) can now be written as:

$$E^{\bar{P}_i}_{X_{jk}} = \sum_n P_{in}\left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_m P_{mn}\frac{\partial V_{mn}}{\partial X_{jnk}}\right] X_{jnk} / \sum_n P_{in} \tag{112}$$

which depends on P_{in} as well as the other P_{mn}'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_{mn}'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.

b) Elasticity at the sample means of the independent variables

The point elasticities corresponding to the two cases (109) and (111) 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.

•Components of the elasticity at the sample means: probability point and partial derivative

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

1. The change of the probability P_{in} relative to a percent change of X_{jnk}:

$$C^{P_{in}}_{X_{jnk}}|_{\bar{X}} = \frac{\partial P_{in}}{\partial X_{jnk}} X_{jnk}|_{\bar{X}}, \tag{113}$$

2. The derivative of the probability P_{in} with respect to X_{jnk}:

$$D^{P_{in}}_{X_{jnk}}|_{\bar{X}} = \frac{\partial P_{in}}{\partial X_{jnk}}|_{\bar{X}}, \tag{114}$$

• LOGIT: Sign of the elasticity at the sample means

Case where the variable appears in one alternative only

The point elasticity (109) evaluated at the sample means can be written as:

$$E^{P_{in}}_{X_{jnk}}|_{\bar{X}} = \beta_k[\delta_{ij} - P_{jn}]X^{\lambda_k}_{jnk}|_{\bar{X}} \ . \tag{115}$$

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 (115):

— 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

The point elasticity (111) evaluated at the sample means can be written as:

$$E^{P_{in}}_{X_{jnk}}|_{\bar{X}} = \left[\frac{\partial V_{in}}{\partial X_{jnk}} - \sum_m P_{mn}\frac{\partial V_{mn}}{\partial X_{jnk}}\right] X_{jnk}|_{\bar{X}} \ . \tag{116}$$

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_{mn}'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.

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:

$$\tilde{E}^{P_{in}}_{X_{jk}} = E^{\bar{P}_i}_{X_{jk}} \frac{\Sigma_n P_{in}}{\Sigma_{n,X_{jnk>0}} P_{in}}, \tag{117}$$

where $E^{\bar{P}_i}_{X_{jk}}$ is given in (97) and $\Sigma_{n,X_{jnk>0}} P_{in}$ 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:

$$\tilde{C}^{\bar{P}_i}_{X_{jk}} = C^{\bar{P}_i}_{X_{jk}} \frac{\Sigma_n P_{in}}{\Sigma_{n,X_{jnk>0}} P_{in}}. \tag{118}$$

2.A. Elasticity at the sample means:

$$\tilde{E}^{P_{in}}_{X_{jnk}}|_{\bar{X}} = E^{P_{in}}_{X_{jnk}}|_{\bar{X}} \frac{\bar{X}^{+}_{jk}}{\bar{X}_{jk}}, \tag{119}$$

where $E^{P_{in}}_{X_{jnk}}|_{\bar{X}}$ is given in (116), $\bar{X}_{jk}$ 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:

$$\tilde{C}^{P_{in}}_{X_{jnk}}|_{\bar{X}} = C^{P_{in}}_{X_{jnk}}|_{\bar{X}} \frac{\bar{X}^{+}_{jk}}{\bar{X}_{jk}}. \tag{120}$$

d) Marginal rate of substitution or 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:

$$\mathrm{MRS}^{V_i}_{X_{ik},X_{ir}} = \frac{\partial V_i/\partial X_{ir}}{\partial V_i/\partial X_{ik}} \tag{121}$$

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:

$$\mathrm{VOT}^{V_i}_{C_i,T_i} = \frac{\partial V_i/\partial T_i}{\partial V_i/\partial C_i} \tag{122}$$

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.

Since the value of time is strictly defined with respect to V_i which differs only across the model types within a family, but not across the families for a same model type, an explicit form of VOT will be given in terms of the model types irrespective of the family considered. Apart from the three main model types, namely (S-LIN), (S-BC) and (G-BC), which can be specified for each of the five families, a fourth model type (G-LIN) is only allowed in the two IPT-L families.

- **(S–LIN) Model Type**

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

$$\mathrm{VOT}^{i}_{ii} = \frac{\beta^{i}_{Ti}}{\beta^{i}_{C_i}} \tag{123}$$

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

- **(S–BC) Model Type**

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

$$\text{VOT}\,^i_{ii} = \frac{\beta^i_{T_i} T_i^{\lambda^i_{T_i}-1}}{\beta^i_{C_i} C_i^{\lambda^i_{C_i}-1}} \tag{124}$$

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.

- **(G–BC) 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:

$$\text{VOT}\,^i = \left[\text{VOT}\,^i_{jh}\right] = \begin{bmatrix} \text{VOT}\,^i_{11} & \text{VOT}\,^i_{12} & \cdots & \text{VOT}\,^i_{1M} \\ \text{VOT}\,^i_{21} & \text{VOT}\,^i_{22} & \cdots & \text{VOT}\,^i_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}\,^i_{M1} & \text{VOT}\,^i_{M2} & \cdots & \text{VOT}\,^i_{MM} \end{bmatrix} \tag{125}$$

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

$$\text{VOT}\,^i_{jh} = \frac{\partial V_i/\partial T_h}{\partial V_i/\partial C_j} = \frac{\beta^i_{T_h} T_h^{\lambda^i_{T_h}-1}}{\beta^i_{C_j} C_j^{\lambda^i_{C_h}-1}}. \tag{126}$$

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:

$$\text{VOT} = \left[\text{VOT}\,^j_{jh}\right] = \begin{bmatrix} \text{VOT}\,^1_{11} & \text{VOT}\,^1_{12} & \cdots & \text{VOT}\,^1_{1M} \\ \text{VOT}\,^2_{21} & \text{VOT}\,^2_{22} & \cdots & \text{VOT}\,^2_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}\,^M_{M1} & \text{VOT}\,^M_{M2} & \cdots & \text{VOT}\,^M_{MM} \end{bmatrix} \tag{127}$$

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 (123) and (124) 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.

• **(G–LIN) Model Type**

Since this model type is nested in the (G-BC) model type for each of the two IPT-L families, all the formulas of the (G-BC) model type still hold for the (G-LIN) model type with every λ-parameter fixed at the value of 1.

• **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 (127) 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 (G-BC) model type which includes as special cases the (S-LIN), (S-BC) and (G-LIN) model types.

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

$$V_j = V_j(C, T, etc.) \tag{128}$$

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:

$$P_j = P_j(V) = \exp(V_j) / \sum_{m=1}^{M} \exp(V_m) \tag{129}$$

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

Analogous to the matrix VOT in (127), the matrix $\widetilde{\mathrm{VOT}}$ computed in terms of the choice probabilities P_j's can be defined as $\widetilde{\mathrm{VOT}} = \left[\widetilde{\mathrm{VOT}}_{jh}^{j}\right]$ where a typical element is:

$$\widetilde{\mathrm{VOT}}_{jh}^{j} = \frac{\partial P_j / \partial T_h}{\partial P_j / \partial C_j}. \tag{130}$$

Using the chain rule for functions of functions, the derivative of P_j with respect to T_h in the numerator of (130) can be written as:

$$\begin{aligned} \frac{\partial P_j}{\partial T_h} &= \frac{\partial P_j}{\partial V_1}\frac{\partial V_1}{\partial T_h} + \frac{\partial P_j}{\partial V_2}\frac{\partial V_2}{\partial T_h} + \ldots + \frac{\partial P_j}{\partial V_M}\frac{\partial V_M}{\partial T_h} = \sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial T_h} \\ &= \frac{\partial V_j}{\partial T_h}\sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial T_h}\frac{\partial T_h}{\partial V_j} = \frac{\partial V_j}{\partial T_h}\sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial V_j}. \end{aligned} \tag{131}$$

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:

$$\frac{\partial P_j}{\partial T_h} = \frac{\partial V_j}{\partial T_h}\frac{\partial P_j}{\partial V_j}. \tag{132}$$

Similarly, the derivative of P_j with respect to C_j in the denominator of (130) can be shown to be:

$$\frac{\partial P_j}{\partial C_j} = \frac{\partial V_j}{\partial C_j}\frac{\partial P_j}{\partial V_j}. \tag{133}$$

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 $\widetilde{\text{VOT}}^j_{jh}$ is identical to VOT^j_{jh} in (127).

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:

$$\widetilde{\text{VOT}}^j_{jj} = \frac{\partial P_j/\partial T_j}{\partial P_j/\partial C_j} = \frac{\partial P_j/\partial V_j}{\partial P_j/\partial V_j}\,\frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \text{VOT}^j_{jj}. \tag{134}$$

- **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:

$$\widetilde{\text{VOT}}^{\bar{P}_j}_{jh} = \sum_n P_{jn}\widetilde{\text{VOT}}^j_{jh.n} / \sum_n P_{jn} \tag{135}$$

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

$$\widetilde{\text{VOT}}^j_{jh.n} = \frac{\partial P_{jn}/\partial T_{hn}}{\partial P_{jn}/\partial C_{hn}} = \frac{\beta^j_{T_h} T_{hn}^{\lambda^j_{T_h}-1}}{\beta^j_{C_h} C_{jn}^{\lambda^j_{C_j}-1}} \tag{136}$$

Note that for the (S-LIN) 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:

$$\widetilde{\text{VOT}}^j_{jh}|_{\bar{X}} = \frac{\partial P_j/\partial T_h}{\partial P_j/\partial C_j}|_{\bar{X}} = \frac{\beta^j_{T_h} \bar{T}_h^{\lambda^j_{T_h}-1}}{\beta^j_{C_j} \bar{C}_j^{\lambda^j_{C_j}-1}}. \tag{137}$$

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 (S-BC) and (G-BC) 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 (S-LIN) and (G-LIN) model types $(\lambda^j_{C_j} = \lambda^j_{T_j} = 1)$, 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 $\widetilde{\text{VOT}}^{\bar{P}_j}_{jj}$.

TABLE 6 Prediction success table.

ALTERNATIVES	PREDICTED ALTERNATIVES $(j = 1, \ldots, M)$	OBSERVED COUNT	OBSERVED SHARE
ACTUAL ALTERNATIVES $(i = 1, \ldots, M)$	$N_{11} \ \ldots \ N_{1M}$ $N_{21} \ \ldots \ N_{2M}$ $\vdots \quad \vdots$ $\vdots \quad \vdots$ $N_{M1} \ \ldots \ N_{MM}$	$\Sigma_j N_{1j}$ $\Sigma_j N_{2j}$ $\vdots$ $\vdots$ $\Sigma_j N_{Mj}$	$\Sigma_j N_{1j}/N$ $\Sigma_j N_{2j}/N$ $\vdots$ $\vdots$ $\Sigma_j N_{Mj}/N$
PREDICTION 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.	$\frac{\Sigma_i N_{i1} - \Sigma_j N_{1j}}{N} \ \ldots \ \frac{\Sigma_i N_{iM} - \Sigma_j N_{Mj}}{N}$		

5.2 Prediction Success Table

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

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:

$$N_{ij} = \sum_n d_{in} P_{jn}. \tag{138}$$

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 $j = 1, \ldots, M$:

$$\sigma = \sum_j \left(\sum_i \frac{N_{ij}}{N} \right) \left(\frac{N_{jj}}{\Sigma_i N_{ij}} - \sum_i \frac{N_{ij}}{N} \right) = \sum_j \left[\frac{N_{jj}}{N} - \left(\sum_i \frac{N_{ij}}{N} \right)^2 \right]. \tag{139}$$

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:

$$\max\,(\sigma) = \sum_j \frac{N_{jj}}{N} - \sum_j \left(\sum_i \frac{N_{ij}}{N} \right)^2 = 1 - \sum_j \left(\sum_i \frac{N_{ij}}{N} \right)^2 \tag{140}$$

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.

5.3 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:

$$X_i'X_i = \sum_{k=1}^{K_i} \gamma_k p_k p_k' \tag{141}$$

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 $X_i'X_ib = X_i'Y_i$;

TABLE 7 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}	...	$\Pi_{K_iK_i}$

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:

$$\text{Var}\,(b_k) = \sigma^2 \sum_{m=1}^{K_i} \left(p_{km}^2/\gamma_m\right). \tag{142}$$

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

$$\Pi_{rk} = \frac{p_{kr}^2/\gamma_r}{\Sigma_{m=1}^{K_i}\left(p_{km}^2/\gamma_m\right)} \tag{143}$$

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.

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

i. High values of the condition indexes signal the existence of near dependencies. Although 30 has been suggested, values depend on units and differences between two situations, and robustness of results may be of more interest than absolute levels of η (Erkel-Rousse 1995). In addition, 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

β-coefficients

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 the covariance matrix Var$\left(\widehat{\Pi}\right)$ in (29) for the LOGIT family or in (35) for the other families. 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 below. 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 matrices:

Logit

$$\text{Var}\left(\widehat{\beta} \mid \lambda\right) = \left[-\frac{\partial^2 L}{\partial \beta \, \partial \beta'}\right]^{-1}_{\Pi=\widehat{\Pi}} \tag{144}$$

Other families

$$\text{Var}\left(\widehat{\Pi}^* \mid \lambda\right) = \left[\sum_n \frac{\partial L_n}{\partial \Pi^*} \frac{\partial L_n}{\partial \Pi^{*\prime}}\right]^{-1}_{\Pi=\widehat{\Pi}} \tag{145}$$

where Π^* is a subvector of Π containing only β and ψ.

These statistics are *invariant* to changes of measurement units in X. Note that when there is no estimated Box-Cox transformations, the conditional and unconditional t-statistics coincide.

Extra parameters

For the estimated ψ-parameters, only **unconditional** t-statistics are computed from the standard errors given by the square roots of the diagonal elements of the covariance matrix $\text{Var}\left(\widehat{\Pi}\right)$ in (35), since the parameters and their t-statistics are *invariant* to changes of measurement units in X. Note that these t-statistics are implicitly conditional on the fixed values of the extra parameters and Box-Cox parameters if there are any.

The null hypotheses associated with the t-statistics are as follows:

S-DOGIT: $H_0 : \theta_i = 0,$

G-DOGIT: $H_0 : \theta_{ij} = 0,$

LIN-IPT-L: $H_0 : \phi_i = 0$ and 1, and $H_0 : \mu_i = 0$ and 1,

BT-IPT-L: $H_0 : \tilde{\phi}_i = 0$ and 1, and $H_0 : \tilde{\mu}_i = 0$ and 1.

Box-Cox parameters

For the estimated λ-parameters, only **unconditional** t-statistics are computed from the standard errors given by the square roots of the diagonal elements of the covariance matrix $\text{Var}\left(\widehat{\Pi}\right)$ in (29) for the LOGIT family or in (35) for the other families, since the parameters 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.

Two null hypotheses are considered:

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

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

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, ψ and λ-parameters in the model.

iv. Sum of squared residuals: $\Sigma_n \Sigma_i \left[d_{in} - \widehat{P}_{in} \right]^2$, where $\widehat{P}_{in}$ 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 $\widehat{P}_{in} = \max\limits_{j \in C_n} \left[\widehat{P}_{jn} \right]$.

5.4 Conditions under which the results are invariant

Generally speaking, there can be different types of model invariance depending on the structure of the model considered. One type of invariance always encountered in econometrics is related to the problem of the measurement units of the variables. For example, in transportation economics, data on the independent variables such as income, distance, travel time and travel cost are often recorded in thousands of dollars, kilometers, minutes and dollars respectively. If the modeler uses these variables in their original units, sometimes they can create numerical problems such as in the inversion of a matrix involving these variables. To avoid these problems, the modeler should resort to the change of scales so that the variables in their new units are expressed in an appropriate order of magnitude. But he does not want that his model changes numerically due to the change of the measurement units, since the estimation results are not meaningful if his model changes numerically.

In the case of a change in the measurement units of the independent variables, the invariance of a model means that the change does not affect:

1. The β-coefficients not related to the change.
2. The Box-Cox λ-parameters.
3. The extra parameters: θ_i's for S-DOGIT, θ_{ij}'s for G-DOGIT, ϕ_i's and μ_i's for LIN-IPT-L and $\tilde{\phi}_i$'s and $\tilde{\mu}_i$'s for BT-IPT-L.
4. The choice probabilities P_{in}'s.
5. The log-likelihood function L.

6. The elasticities.

In other words, only the β-coefficients directly involved in the change adjust themselves so that all the remaining parameters, as well as the choice probabilities, the log-likelihood function and the elasticities are not affected by the change.

For all the five families (LOGIT, S-DOGIT, G-DOGIT, LIN-IPT-L and BT-IPT-L), two types of model invariance can be considered:

1. Invariance with respect to the choice of an alternative where an alternative specific constant is omitted, i.e. when only $M-1$ specific constants remain to be estimated. In this case, model invariance means that only the $M-1$ specific constants will change but not the other β-coefficients, the Box-Cox parameters, the extra parameters, the choice probabilities, the log-likelihood function and the elasticities.
2. Invariance with respect to a change in the measurement units of an independent variable X_{ink} which can be:

   A. A linear variable $(-\infty < X_{ink} < +\infty)$ without a Box-Cox transformation.
   B. A positive variable $(X_{ink} > 0)$ with a Box-Cox transformation.
   C. A quasi-dummy $(X_{ink} \geq 0)$ with a Box-Cox transformation, with simultaneous use of an associated dummy variable.

Model invariance means that for case A, only the β-coefficient associated with X_{ink}, say β_{ik}, will change. For case B, the coefficient β_{ik} will change with a shift in the alternative-specific constant β_{i0}, created by the BCT on X_{ink} expressed in the new units. For case C, the coefficient β_{ik} will change with a shift in the β-coefficient of the associated dummy created by the BCT on X_{ink} expressed in the new units.

Within each category of variables, the specification of the constants should jointly be considered since, as we shall see later, the invariance will depend also on how the constants are specified.

Specification of the constants

In a model with M alternatives, the constants can be specified as:

i. M specific constants.
ii. $M-1$ specific constants.
iii. M equal constants.

Note that in the first three families (LOGIT, S-DOGIT and G-DOGIT), a model cannot be estimated for the cases (i) and (iii), since the M constants are not identified. But for the last two families (LIN-IPT-L and BT-IPT-L), the constants are identified in all the three cases so that a model can always be estimated.

Consider a model with three alternatives in which the representative utility components V_{jn}'s have the following forms:

$$\begin{bmatrix} V_{1n} \\ V_{2n} \\ V_{3n} \end{bmatrix} = \begin{bmatrix} C_{1n} & 0 & 0 & X_{1n} & 0 & 0 \\ 0 & C_{2n} & 0 & 0 & X_{2n} & 0 \\ 0 & 0 & C_{3n} & 0 & 0 & X_{3n} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{bmatrix} \tag{146}$$

where β_1, β_2 and β_3 are the alternative specific coefficients associated with the constant terms C_{1n}, C_{2n} and C_{3n} respectively, and β_4, β_5 and β_6 are those associated with the independent variables X_{1n}, X_{2n} and X_{3n} respectively.

LOGIT

The choice probability of alternative 1 can be written as:

$$P_{1n} = \frac{\exp(V_{1n})}{\sum_{m=1}^{3} \exp(V_{mn})} = \frac{\exp(\beta_1 C_{1n} + \beta_4 X_{1n})}{\exp(\beta_1 C_{1n} + \beta_4 X_{1n}) + \exp(\beta_2 C_{2n} + \beta_5 X_{2n}) + \exp(\beta_3 C_{3n} + \beta_6 X_{3n})} \quad . \tag{147}$$

Dropping the C_{mn}'s in the V_{mn}'s since they are all equal to 1 for every m and n, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1n} = \frac{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n})}{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n}) + \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2n}) + \exp(\beta_6 X_{3n})} \quad . \tag{148}$$

i. ***M* specific constants**: all we can estimate are the two differences $(\beta_1 - \beta_3)$ and $(\beta_2 - \beta_3)$, but not β_1, β_2 and β_3 separately.

ii. ***M-1* specific constants**: if we drop the third constant β_3 for example, i. e. setting it to zero, then β_1 and β_2 can be estimated.

iii. ***M* equal constants**: if an equality constraint is imposed on all three β's $(\beta_1 = \beta_2 = \beta_3)$, the common β disappears since the choice probability reduces to:

$$P_{1n} = \frac{\exp(\beta_4 X_{1n})}{\exp(\beta_4 X_{1n}) + \exp(\beta_5 X_{2n}) + \exp(\beta_6 X_{3n})} \quad . \tag{149}$$

S-DOGIT

The choice probability of alternative 1 can be written as:

$$P_{1n} = \frac{\exp(V_{1n}) + \theta_1 \sum_j \exp(V_{jn})}{\left(1 + \sum_m \theta_m\right) \sum_j \exp(V_{jn})} \quad . \tag{150}$$

Dropping the C_{jn}'s in the V_{jn}'s since they are all equal to 1 for every j and n, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1n} = \frac{\exp(V_{1n} - \beta_3) + \theta_1 \sum_j \exp(V_{jn} - \beta_3)}{\left(1 + \sum_m \theta_m\right) \sum_j \exp(V_{jn} - \beta_3)} \tag{151}$$

where each $\exp(V_{jn} - \beta_3)$ has the following form:

$$\begin{aligned} \exp(V_{1n} - \beta_3) &= \exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1n}) \\ \exp(V_{2n} - \beta_3) &= \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2n}) \\ \text{and } \exp(V_{3n} - \beta_3) &= \exp(\beta_6 X_{3n}) \quad . \end{aligned} \tag{152}$$

i. ***M* specific constants**: all we can estimate are the two differences $(\beta_1 - \beta_3)$ and $(\beta_2 - \beta_3)$, but not β_1, β_2 and β_3 separately.

ii. ***M-1* specific constants**: if we drop the third constant β_3 for example, i. e. setting it to zero, then β_1 and β_2 can be estimated.

iii. ***M* equal constants**: if an equality constraint is imposed on all three β's $(\beta_1 = \beta_2 = \beta_3)$, the common β disappears since the choice probability reduces to:

$$P_{1n} = \frac{\exp(\beta_4 X_{1n}) + \theta_1[\exp(\beta_4 X_{1n}) + \exp(\beta_5 X_{2n}) + \exp(\beta_6 X_{3n})]}{\left(1 + \sum_m \theta_m\right)[\exp(\beta_4 X_{1n}) + \exp(\beta_5 X_{2n}) + \exp(\beta_6 X_{3n})]} \quad . \tag{153}$$

G-DOGIT

The choice probability of alternative 1 can be written as:

$$P_{1n} = \frac{\exp(V_{1n}) + \sum\limits_{j\neq 1} \theta_{1j} \exp(V_{jn})}{\sum\limits_{m} \exp(V_{mn}) + \sum\limits_{m} \sum\limits_{j\neq m} \theta_{mj} \exp(V_{jn})} \quad . \tag{154}$$

Dropping the C_{jn}'s in the V_{jn}'s since they are all equal to 1 for every j and n, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1n} = \frac{\exp(V_{1n} - \beta_3) + \sum\limits_{j\neq 1} \theta_{1j} \exp(V_{jn} - \beta_3)}{\sum\limits_{m} \exp(V_{mn} - \beta_3) + \sum\limits_{m} \sum\limits_{j\neq m} \theta_{mj} \exp(V_{jn} - \beta_3)} \tag{155}$$

where each $\exp(V_{jn} - \beta_3)$ or $\exp(V_{mn} - \beta_3)$ has the following form:

$$\begin{aligned} \exp(V_{1n} - \beta_3) &= \exp(\beta_1 - \beta_3) \exp(\beta_4 X_{1n}) \\ \exp(V_{2n} - \beta_3) &= \exp(\beta_2 - \beta_3) \exp(\beta_5 X_{2n}) \\ \text{and } \exp(V_{3n} - \beta_3) &= \exp(\beta_6 X_{3n}) \quad . \end{aligned} \tag{156}$$

i. ***M* specific constants**: all we can estimate are the two differences $(\beta_1 - \beta_3)$ and $(\beta_2 - \beta_3)$, but not β_1, β_2 and β_3 separately.

ii. ***M-1* specific constants**: if we drop the third constant β_3 for example, i. e. setting it to zero, then β_1 and β_2 can be estimated.

iii. ***M* equal constants**: if an equality constraint is imposed on all three β's $(\beta_1 = \beta_2 = \beta_3)$, the common β disappears since the choice probability reduces to:

$$P_{1n} = \frac{\exp(\beta_4 X_{1n}) + \theta_{12} \exp(\beta_5 X_{2n}) + \theta_{13} \exp(\beta_6 X_{3n})}{(1 + \theta_{21} + \theta_{31}) \exp(\beta_4 X_{1n}) + (1 + \theta_{12} + \theta_{32}) \exp(\beta_5 X_{2n}) + (1 + \theta_{13} + \theta_{23}) \exp(\beta_6 X_{3n})} \quad . \tag{157}$$

LIN-IPT-L

The choice probability of alternative 1 can be written as:

$$P_{1n} = \frac{[\phi_1 \exp(V_{1n}) + 1]^{\frac{1}{\phi_1}} - \mu_1}{\sum\limits_{m} \left\{ [\phi_m \exp(V_{mn}) + 1]^{\frac{1}{\phi_m}} - \mu_m \right\}} \quad . \tag{158}$$

Dropping the C_{mn}'s in the V_{mn}'s since they are all equal to 1 for every m and n, we get:

$$P_{1n} = \frac{[\phi_1 \exp(\beta_1 + \beta_4 X_{1n}) + 1]^{\frac{1}{\phi_1}} - \mu_1}{\sum\limits_{m} \left\{ [\phi_m \exp(\beta_m + \beta_{m+3} X_{mn}) + 1]^{\frac{1}{\phi_m}} - \mu_m \right\}} \quad . \tag{159}$$

i. ***M* specific constants**: due to the nonlinear structure of the expression $[\phi_m \exp(\beta_m + \beta_{m+3} X_{mn}) + 1]^{\frac{1}{\phi_m}}$, $m = 1, \ldots, 3$, all three constants β_1, β_2 and β_3 can be estimated separately.

ii. ***M-1* specific constants**: setting one of the constants to zero, the two remaining ones can always be estimated.

iii. ***M* equal constants**: setting all three constants equal, the common β can be estimated.

BT-IPT-L

The choice probability of alternative 1 can be written as:

$$P_{1n} = \frac{\exp\left\{\frac{[\exp(V_{1n})+\tilde{\mu}_1]^{\tilde{\phi}_1}-1}{\tilde{\phi}_1}\right\}}{\sum_m \exp\left\{\frac{[\exp(V_{mn})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}} \quad . \tag{160}$$

Dropping the C_{mn}'s in the V_{mn}'s since they are all equal to 1 for every m and n, we get:

$$P_{1n} = \frac{\exp\left\{\frac{[\exp(\beta_1+\beta_4 X_{1n})+\tilde{\mu}_1]^{\tilde{\phi}_1}-1}{\tilde{\phi}_1}\right\}}{\sum_m \exp\left\{\frac{[\exp(\beta_m+\beta_{m+3} X_{mn})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}} \quad . \tag{161}$$

i. ***M* specific constants**: due to the nonlinear structure of the expression $\exp\left\{\frac{[\exp(\beta_m+\beta_{m+3} X_{mn})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}$, $m = 1,\ldots,3$, all three constants β_1, β_2 and β_3 can be estimated separately.

ii. ***M-1* specific constants**: setting one of the constants to zero, the two remaining ones can always be estimated.

iii. ***M* equal constants**: setting all three constants equal, the common β can be estimated.

In the last two families, LIN-IPT-L and BT-IPT-L, the cases of $M-1$ specific constants appear to be analogous to the LOGIT case, but in effect constitute distinct forms which reduce to the LOGIT form only if all the ϕ and μ-parameters are equal to 1 for LIN-IPT-L, and all the $\tilde{\phi}$ and $\tilde{\mu}$-parameters are equal to 0 for BT-IPT-L.

Invariance properties

A *partial* type of invariance arises when a particular specification option is adopted. All the invariance properties according to the type of option chosen and the family considered are given in Table 8, where the abbreviations are defined as follows:

- **(Model not identified)**: the model is not identified since in the cases (i) and (iii), the constants are not identified for LOGIT, S-DOGIT and G-DOGIT;
- **UI** (Unconditional Invariance): the invariance holds with or without the equality constraints on the parameters ϕ's and μ's for LIN-IPT-L, and on the parameters $\tilde{\phi}$'s and $\tilde{\mu}$'s for BT-IPT-L;
- **CI** (Conditional Invariance): the invariance holds only if all the ϕ's are equal and all the μ's are equal for LIN-IPT-L, and all the $\tilde{\phi}$'s are equal and all the $\tilde{\mu}$'s are equal for BT-IPT-L.

TABLE 8 Partial Invariance Types in Probability Models

TYPE	INVARIANCE	LOGIT S-DOGIT G-DOGIT	LIN-IPT-L	BT-IPT-L
P1	with respect to the choice of the alternative where a specific constant is omitted (case ii)	YES	NO	NO
P2	with respect to the change of the measurement units of an independent variable X_{ink}:			
P2.A	**Linear Variable** $(-\infty < X_{ink} < +\infty)$ i) M specific constants ii) $M-1$ specific constants iii) M equal constants	 (Model not identified) YES (Model not identified)	 UI UI NO	 UI UI NO
P2.B	**Positive Variable** $(X_{ink} > 0)$ with Box-Cox i) M specific constants ii) $M-1$ specific constants iii) M equal constants	 (Model not identified) YES (Model not identified)	 CI NO NO	 CI NO NO
P2.C	**Quasi-Dummy** $(X_{ink} \geq 0)$ with Box-Cox i) M specific constants ii) $M-1$ specific constants iii) M equal constants	 (Model not identified) YES (Model not identified)	 CI NO NO	 CI NO NO

Clearly, for Types P2.B.ii or iii and Types P2.C.ii or iii, the use of Box-Cox transformations on the independent variables of the representative utility function is very restrictive. The reason for this is that the cases of $M-1$ specific constants differ profoundly from the LOGIT case. The Box-Cox transformations in effect shift the constants in ways that are hampered by the omission of one specific constant.

By contrast, in the cases of M specific constants, it is surprising that changes in units generally imply results that are not invariant: invariance only occurs if equality restrictions are used on the outer function. In general there is no buffer to absorb the shifts: the effect goes on the ϕ's and μ's in LIN-IPT-L and the $\tilde{\phi}$'s and $\tilde{\mu}$'s in BT-IPT-L, and indeed everywhere, on all model parameters.

Moreover, *global* invariance properties arise when combinations of specification options are considered jointly, as in Table 9 where we note:

1. Type P1 can be combined with Type P2.A, Type P2.B or Type P2.C with regard to the case (ii) only, since for Type P1, only $M-1$ specific constants are specified.
2. For the first three families (LOGIT, S-DOGIT, G-DOGIT), Type P1 can be combined with Type P2.A.ii, Type P2.B.ii and Type P2.C.ii to give the most general specification where all the three groups of variables (linear, positive and quasi-dummy) are present. The combined result implies that the invariance will hold globally.

3. For LIN-IPT-L and BT-IPT-L, Type P1 combined with Type P2.A.ii, Type P2.B.ii and Type P2.C.ii implies that the invariance will not be preserved globally.
4. For LIN-IPT-L, the combination of Type P2.A.i, Type P2.B.i and Type P2.C.i implies that the invariance will not hold globally. In contrast, for BT-IPT-L, at least the conditional invariance (CI) will be preserved globally.
5. For LIN-IPT-L and BT-IPT-L, the combination of Type P2.A.iii, Type P2.B.iii and Type P2.C.iii implies that the invariance will not hold globally.

TABLE 9 Global Invariance Types in Probability Models

TYPE	**COMBINATION OF INVARIANCE TYPES**	**LOGIT S-DOGIT G-DOGIT**	**LIN-IPT-L**	**BT-IPT-L**
GI	i) M **specific constants** (P2.A.i)+(P2.B.i)+(P2.C.i)	(Model not identified)	CI	CI
GII	ii) $M-1$ **specific constants** (P1)+(P2.A.ii)+(P2.B.ii)+(P2.C.ii)	YES	NO	NO
GIII	iii) M **equal constants** (P2.A.iii)+(P2.B.iii)+(P2.C.iii)	(Model not identified)	NO	NO

6 SPECIAL OPTIONS

6.1 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 (11) 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 of the log-likelihood function, the weighted aggregate elasticity 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 $H(i)$.

6.2 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:

$$U_{md} = \widetilde{V}_m + \widetilde{V}_d + \widetilde{V}_{md} + \epsilon_{md} \tag{162}$$

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):

$$P_{md} = \frac{\exp\left(\widetilde{V}_m + \widetilde{V}_d + \widetilde{V}_{md}\right)}{\sum\limits_{(m',d') \in C_n} \exp\left(\widetilde{V}_{m'} + \widetilde{V}_{d'} + \widetilde{V}_{m'd'}\right)} \tag{163}$$

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).

6.3 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 5.1. 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 $\widehat{P}_{in}$ can be evaluated for each individual n in the second sample as:

$$\widehat{P}_{in} = \mathcal{U}_{in}\left(\widehat{\Pi}\right) / \sum_{m} \mathcal{U}_{mn}\left(\widehat{\Pi}\right) \tag{164}$$

where the utilility function $\mathcal{U}_{mn}\left(\widehat{\Pi}\right)$ associated with each model family is computed with the parameter vector $\widehat{\Pi}$ 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, for the five model families, the confidence intervals for the choice probabilities are computed by the linear approximation method in which each predicted choice probability $\widehat{P}_{in}$ in the second sample is linearized by a first order Taylor-series expansion about $\widehat{\Pi}$:

$$\widehat{P}_{in} = P_{in} + \left(\widehat{\Pi} - \Pi\right)' \frac{\partial P_{in}}{\partial \Pi} \tag{165}$$

where $\widehat{\Pi}$ is a column vector containing the β, ψ and λ-values estimated from the first sample, and P_{in} and Π are respectively the true choice probability value and the column vector of the true β, ψ and λ-values.

The column vector of derivatives $\partial P_{in}/\partial \Pi$ in (165) can be written as:

$$\frac{\partial P_{in}}{\partial \Pi} = \frac{\partial \left(\mathcal{U}_{in} / \sum_m \mathcal{U}_{mn} \right)}{\partial \Pi} = \frac{1}{\sum_m \mathcal{U}_{mn}} \left(\frac{\partial \mathcal{U}_{in}}{\partial \Pi} - P_{in} \sum_m \frac{\partial \mathcal{U}_{mn}}{\partial \Pi} \right) \tag{166}$$

where the typical elements of the column vector of derivatives $\partial \mathcal{U}_{mn}/\partial \Pi$ are given in (13-27) for the five model families.

Since $\widehat{\Pi}$ is asymptotically normal with mean Π and covariance matrix Var($\widehat{\Pi}$) given in (29) for the LOGIT family and in (35) for the other families, each predicted choice probability $\widehat{P}_{in}$ in its linearized form (165) is asymptotically normal with mean P_{in} and variance:

$$\text{Var}\left(\widehat{P}_{in}\right) = \frac{\partial P_{in}}{\partial \Pi'} \text{ Var } \left(\widehat{\Pi}\right) \frac{\partial P_{in}}{\partial \Pi}. \tag{167}$$

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

$$\widehat{P}_{in} - Z_{\epsilon/2}\sqrt{\text{Var}\left(\widehat{P}_{in}\right)} \leq P_{in} \leq \widehat{P}_{in} + Z_{\epsilon/2}\sqrt{\text{Var}\left(\widehat{P}_{in}\right)} \tag{168}$$

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_{in} and hence the value of $Z_{0.025}$ is equal to 1.96.

7 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. **Batsell, R.R.** (1981), “A Multiattribute Extension of the Luce Model which simultaneously scales Utility and Substitutability”, in J. Huber (ed.): *The Effect of Item Similarity on Choice Probabilities: A collection of working papers from a conference at Quail Roost*, Duke University, 4–23.
5. **Belsley, D., E. Kuh and R.E. Welsh** (1980), *Regression Diagnostics*, John Wiley, New York.
6. **Ben-Akiva, M.** (1977), “Choice Models with Simple Choice Set Generating Processes”, Working paper, Department of Civil Engineering, MIT, Cambridge, Mass.
7. **Ben-Akiva, M. and S.R. Lerman** (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass
8. **Berkson, J.** (1944), “Applications of the Logistic Function to Bioassay", *Journal of the American Statistical Association*, 39, 357–365.
9. **Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman** (1974), “Estimation and Inference in Nonlinear Structural Models”, *Annals of Economic and Social Measurement*, 3, 653–665.
10. **Bock, R.D. and L.V. Jones** (1968), *The Measurement and Prediction of Judgement and Choice*, Holden-Day, San Francisco.
11. **Bordley, R.F.** (1989), “Relaxing the Loyalty Condition in the Colombo/Morrison Model”, *Marketing Science*, 8(1), 100-103.
12. **Bordley, R.F.** (1990), “The Dogit Model is Applicable even Without Perfectly Captive Buyers”, *Transportation Research*, 24B, 4, 315–323.
13. **Borgers, A.W.J. and H.J.P. Timmermans** (1984), “A Context-sensitive Model of Spatial Choice Behaviour”, Paper prepared for the meeting of the IGU working group on Mathematical Models and Systems Analysis, Besançon.
14. **Box, G.P. and D.R. Cox** (1964), “An Analysis of Transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211–243.
15. **Box, G.P. and P.W. Tidwell** (1962), “Transformations of the Independent Variables", *Technometrics*, 4, 531–550.
16. **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.
17. **Cascetta, E., A. Nuzzolo, F. Russo and A. Vitetta** (1996), “A New Route Choice Logit Model overcoming IIA problems: Specification and Some Calibration Results for Interurban Networks", in J. B. Lesort (ed.), *Transportation and Traffic Theory: Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 1996, 15 pages.

18. **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.

19. **Clarke, R.R.** (1957), "Constant-Ratio Rule for Confusion Matrices in Speech Communication", *Journal of Acoustical Society of America*, 30, 715–720.

20. **Colombo, R. and D. Morrison** (1989), "A Brand-switching Model with Implications for Marketing Strategies", *Marketing Science*, 8(1), 89–99.

21. **Coombs, C.H.** (1958), "On the Use of Inconsistency of Preferences in Psychological Measurements", *Journal of Experimental Psychology*, 55, 1–7.

22. **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.

23. **Cox, D.R.** (1970), *Analysis of Binary Data*, Methuen, London.

24. **Cragg, J.G. and T.E. Lisco** (1968), *PROLO-User Manual*, Department of Economics, University of Chicago (mimeo).

25. **Crittle, F.J. and L.W. Johnson** (1980), *Basic Logit (BLOGIT) – Technical Manual*, ATM no.9, Australian Road Research Board

26. **De Donnea, F.X.** (1971), *The Determinants of Transport Mode Choice in Dutch Cities*, Rotterdam University Press, Rotterdam, The Netherlands.

27. **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.

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

29. **Erkel-Rousse, H.** (1995), "Détection de la multicolinéarité dans un modèle linéaire ordinaire: Quelques éléments pour un usage averti des indicateurs de Belsley, Kuh et Welsch", *Revue de Statistique Appliquée*, Vol. XLIII (4), 19–42.

30. **Finney, D.J.** (1971), *Probit Analysis*, Third Edition , Cambridge University Press, Cambridge, England.

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

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

33. **Gaudry, M.J.I.** (1981), "The Inverse Power Transformation Logit and Dogit Mode Choice Models", *Transportation Research*, Part B, 15B, 97–103.

34. **Gaudry, M.J.I.** (1989), "Asymmetric Shape and Variable Tail Thickness in Multinomial Probabilistic Response: Three Model Type Families in a Quasi-Direct Format Application to Intercity Travel Demand with Aggregate Canadian Data", Publication #893, Centre de Recherche sur les Transports, Université de Montréal.

35. **Gaudry, M.J.I. and M.G. Dagenais** (1978), "The Generalized Dogit Mode Choice Model", Publication 107, Centre de Recherche sur les Transports, Université de Montréal.

36. **Gaudry, M.J.I. and M.G. Dagenais** (1979), "The Dogit Model", *Transportation Research*, Part B, 13B, 105–111.

37. **Gaudry, M.J.I. and M.J. Wills** (1979), “Testing the Dogit Model with Aggregate Time-Series and Cross-Sectional Travel Data”, *Transportation Research*, Part B, 13B, 155–166.
38. **Gronau, R.** (1970), *The Value of Time in Passenger Transportation: The Demand for Air Travel*, National Bureau of Economic Research, New York.
39. **Hensher, D.A.** (1979), *MLOGPROT User’s Manual*, School of Economic and Financial Studies, Macquarie University, Sidney.
40. **Hensher, D.A. and L.W. Johnson** (1981), *Applied Discrete-Choice Modelling*, Croom Helm, London.
41. **Horowitz, J.L.** (1979), “Confidence Intervals for Choice Probabilities of the Multinomial Logit Model”, *Transportation Research Record*, 728, 23–29.
42. **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.
43. **Horowitz, J.L.** (1983), “Statistical Comparison of Non-Nested Probabilistic Choice Models", *Transportation Science*, 17, 319–350.
44. **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.
45. **Koppelman, F.** (1975), *Travel Prediction with Models of Individualistic Choice Behavior*, Ph.D. dissertation, Department of Civil Engineering, MIT, Cambridge, Mass.
46. **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.
47. **Krishnan, K.** (1977), “Incorporating Thresholds of Indifference in Probabilistic Choice Models”, *Management Science*, 23, 1224–33.
48. **Lave, C.A.** (1969), “A Behahioral Approach to Modal Split Forecasting", *Transportation Research*, 3, 463–480.
49. **Lerman, S.R.** (1976), “Location, Housing, Auto Ownership and Mode to Work: A Joint-Choice Model", *Transportation Research Record*, no. 610, 6–11.
50. **Lerman, S.R.** (1982), *Micro Data Analyzer: the MDA User’s Manual*, Release 1.0, Cambridge Systematics, Inc.
51. **Lioukas, S.** (1984), “Thresholds and Transitivity in Stochastic Consumer Choice: A Multinomial Logit Analysis”, *Management Science*, 30, 110–22.
52. **Lisco, T.E.** (1967), *The Value of Commuters’ Travel Time: A Study in Urban Transportation*, Ph.D. Dissertation, Department of Economics, University of Chicago.
53. **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.
54. **Luce, R.D.** (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.
55. **Manski, C.** (1977), “The Structure of Random Utility Models”, *Theory and Decision*, 8, 229–254.
56. **Manski, C.F.** (1981), “Structural Models of Discrete Data", *Sociological Methodology*, 58–109.
57. **Manski, C.F. and S.R. Lerman** (1977), “The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45, 1977–1988.
58. **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.

59. **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.
60. **McFadden, D.** (1974), “Conditional Logit Analysis of Qualitative Choice Behavior", in P.C. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 105–142.
61. **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.
62. **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.
63. **McFadden, D.** (1982), “Econometric Analysis of Qualitative Response Models", Working Paper, Department of Economics, MIT, Cambridge, Mass.
64. **McFadden, D., J. Berkman, D. Browstone and G.M. Duncan** (1979), *QUAIL 3.0 User’s Manual*, Department of Economics, University of California, Berkeley.
65. **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.
66. **McGillivray, R.G.** (1972), “Binary Choice of Urban Transport Mode in the San Francisco Region", *Econometrica*, 40, 827–848.
67. **Meyer, R.J. and T.C. Eagle** (1981), “A Parsimonious Multinomial Choice Model Recognizing Alternative Interdependence and Context-dependent Utility Functions”, Working paper 26-80-81 (Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA).
68. **Meyer, R.J. and T.C. Eagle** (1982), “Context-induced Parameter Instability in a Disaggregate-Stochastic Model Of Store Choice”, *Journal of Marketing Research*, 19, 62–71.
69. **Montmarquette, C. and S. Mahseredjian** (1985), “Functional Forms and Educational Production Functions”, *Economics Letters*, 19, 291–294.
70. **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.
71. **Quandt, R.** (1970), *The Demand for Travel*, D.C. Heath, Lexington, Mass.
72. **Quarmby, D.A.** (1967), “Choice of Travel Mode for the Journey to Work: Some Findings", *Journal of Transportation and Economic Policy*, 1, 273–314.
73. **Restle, F.** (1961), *Psychology of Judgement and Choice*, New York: John Wiley and Sons.
74. **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.
75. **Stone, R.** (1954), “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand”, *Economic Journal*, 64, 511–527.
76. **Stopher, P.R.** (1969), “A Probability Model of Travel Mode Choice for the Work Journey", *Highway Research Record*, 283, 57–65.
77. **Swait, J. and M. Ben-Akiva** (1987), “Empirical Test of a Constrained Choice Discrete Model:

Mode Choice in Saõ Paulo, Brazil”, *Transportation Research*, Part B, 21B, 103–115.

78. **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.
79. **Talvitie, A.D.** (1972), “Comparison of Probabilistic Modal Choice Models: Estimation Methods and System Inputs", *Highway Research Record*, 392, 111–120.
80. **Thurstone, L.** (1927), “A Law of Comparative Judgement", *Psychological Review*, 34, 273–286.
81. **Tukey, J.W.** (1957), “On the Comparative Anatomy of Transformations", *Annals of Mathematical Statistics*, 28, 602–632.
82. **Tversky, A.** (1972), “Elimination by Aspects: A Theory of Choice", *Psychological Review*, 79, 281–299.
83. **Warner, S.L.** (1962), *Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice*, Northwestern University Press, Evanston, Ill.
84. **Watson, P.L.** (1974), *The Value of Time: Behavioral Models of Modal Choice*, D.C. Heath, Lexington, Mass.
85. **Westin, R. and D. Gillen** (1978), “Parking Location and Transit Demand", *Journal of Econometrics*, 8, 75–101.
86. **Wigner, M.F.** (1973), “Disaggregate Modal Choice Models of Downtown Trips in the Chicago Region", *Highway Research Record*, 446, 49–65.
87. **Zarembka, P.C.** (1974), “Transformation of Variables in Econometrics", in P.C. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 81–104.