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 S-1/S-5 algorithm known in its last version as: Liem, T., Gaudry, M. and R. Laferrière, "SHARE: The S-1 to S-5 programs for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT models with aggregate data", Publication CRT-899, Centre de recherche sur les transports, Université de Montréal, 70 p., 1993, 1997. The 1993 version incorporated Richard Laferrière's verbal suggestion on how to extend the previous estimation procedures of 1989 to handle data sets with alternatives that are unavailable for some observations. The 1997 version was enriched by work on parameter invariance supported by the "STEMM" project (Contract No.: ST-96-SC.301 of the Commission of the European Communities). This document is complemented by a user guide for the program: Tran, C-L., and Gaudry, M., "The S-1/S-5 Program user guide with databases, examples and Tablex outputs", Publication AJD-115, Agora Jules Dupuit, Université de Montréal, April 2008.

PUBLICATION
AJD-114

SHARE: The S-1 to S-5 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 aggregate 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-114, Agora Jules Dupuit
Université de Montréal
April 2008 — Revised October 2008

ABSTRACT

The S-1/S-5 programs are designed for the maximum likelihood estimation of five model families which all include the **Multinomial LOGIT** model as a special case, for a system of equations in which the dependent variables are expressed in terms of aggregate shares. The S-1 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, may imply a systematic underestimation (or overestimation) at the tails of the distribution, if consumers will choose certain alternatives despite their apparent unattractiveness. Various more general specifications have been established to overcome the limitations of the IIA assumption and the "thin equal tails" property. The S-2/S-5 programs have been developped following these directions. The **S-DOGIT** and **G-DOGIT** families (S-2/S-3 programs) relax the IIA restriction by allowing the relative shares 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 (S-4/S-5 programs) totally remove the IIA restriction so that the relative shares 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 parameters associated with the independent variables can be fixed or jointly estimated with the regression coefficients (except for the G-DOGIT family where only fixed Box-Cox parameters can be used) to introduce flexible functional forms other than the linear and logarithmic forms frequently used. Elasticities, marginal rates of substitution and forecasts obtained by simulation and maximum likelihood techniques are also computed.

Keywords: Box-Cox Share Models, Aggregate Data, Symmetric/Asymmetric Tails, Captivity, Similarity/Dissimilarity, Maximum Likelihood Estimation, Elasticities, Marginal Rates of Substitution, Maximum Likelihood Forecast.

RÉSUMÉ

Les programmes S-1/S-5 sont destinés à l'estimation par la méthode du maximum de vraisemblance de cinq familles de modèles qui incluent toutes le modèle **LOGIT Multinomial** comme cas particulier, pour un système d'équations dans lequel les variables dépendantes sont sous forme de parts agrégées. Le programme S-1 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 formulations plus générales ont été construites pour surmonter les limitations de l'hypothèse IANP et de la propriété des queues "minces et égales". Les programmes S-2/S-5 ont été développés selon ces directions. Les familles du **S-DOGIT** et du **G-DOGIT** (programmes S-2/S-3) assouplissent la restriction IANP en permettant aux parts 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 S-4/S-5) ne respectent plus du tout la restriction IANP, de telle sorte que les parts 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, différents paramètres de la transformation de Box-Cox appliquée aux variables indépendantes peuvent être fixés ou estimés conjointement avec les coefficients de la régression (sauf pour la famille du G-DOGIT où seulement des paramètres Box-Cox fixés peuvent être utilisés) pour introduire des formes fonctionnelles flexibles autres que les formes linéaire et logarithmique fréquemment utilisées. Des élasticités, des taux marginaux de substitution et les prévisions obtenues par des méthodes de simulation et du maximum de vraisemblance sont aussi calculés.

Mots-clés: Modèles de parts avec Box-Cox, Données Agrégées, Queues Symétriques/Assymétriques, Captivité, Similarité/Dissimilarité, Estimation du Maximum de Vraisemblance, Elasticités, Taux Marginaux de Substitution, Prévisions par Maximisation de la Vraisemblance.

Contents

List of Tables and Figures

INTRODUCTION

Statistical models involving a system of equations in which the sum of the dependent variables is constrained have been widely used in various fields of economic analysis. Since the dependent variables are expressed in terms of shares — budget shares in the expenditure system, profit shares in the theory of the firm or cost shares in the production theory — these models are specifically called SHARE EQUATION MODELS. Due the constraint that the shares should add up to unity, one well known characteristic of these models is that the equation system is singular and can only be estimated after dropping one equation (Barten 1969).

A convenient approach in the estimation of the share equations is to introduce an additive error term in each equation and to estimate the equation system by the Seemingly Unrelated Regression Equation (SURE) method of Zellner (1962). Econometric applications of this approach can be found in the linear expenditure system based on the Stone-Geary utility function (Stone 1954, Pollack and Wales 1969, Parks 1971), the budget share models derived from the translog indirect utility functions of Christensen, Jorgenson and Lau (1975), the profit share models derived from the indirect profit functions (Lau 1979, Weaver 1983) or the translog cost models (Berndt and Christensen 1973, 1974, Binswanger 1974, Ray 1982), notably in the field of energy economics (Hudson and Jorgenson 1974, Berndt and Wood 1975, Griffin and Gregory 1976), but the procedure has two serious shortcomings:

i. Since the additive error terms are assumed to be normally distributed, hence not restricted to any particular domain, the estimated shares may lie outside the interval [0,1] (Woodland 1979);
ii. In a time series context, if the share equations are estimated with autoregressive disturbances, the constraint on the dependent variables leads to rather strong restrictions on the parameters of the autoregressive process in order to preserve the invariance of the parameter estimates with respect to the equation dropped (Berndt and Savin 1975).

An alternative approach which does not suffer from these drawbacks is to assume a multiplicative error term in each share equation and to estimate the equation system using the LINEAR LOGIT specification, by the SURE or the maximum likelihood methods depending on the linearity or nonlinearity of the system. Although the term “LOGIT” is used in connection with the MULTINOMIAL LOGIT model which applies strictly to the discrete choice case where the dependent variables are of discrete type, this does not imply that the share equations result from an underlying discrete choice structure, but rather the shares which are continuous variables over the interval [0,1] can be interpreted as a set of independent, conditional probabilities of the logistic form. This approach has been adopted in a variety of studies, e.g. choice of transportation modes (Theil 1969), occupational choice of individuals (Schmidt and Strauss 1975), choice of automobile ownership (Parks 1980), pattern of household expenditures (Tyrrell and Mount 1982), dynamic input demand systems (Considine and Mount 1984) or cost share estimation with autoregressive disturbances (Chavas and Segerson 1986).

This computer package is based on the second approach and contains five different programs developed earlier for the estimation of aggregate mode shares in transportation demand analysis:

1. S-1 program for the **LOGIT** model family,
2. S-2 / S-3 programs for the **S**tandard and **G**eneralized **DOGIT** model families,
3. S-4 / S-5 programs for the **LIN**ear and **B**ox-**T**ukey **I**nverse **P**ower **T**ransformation **L**OGIT model families.

The five model families (**LOGIT**, **S-DOGIT**, **G-DOGIT**, **LIN-IPT-L**, **BT-IPT-L**) are now presented as different options in this computer package and they all include the LINEAR LOGIT model 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 was 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 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

• **CASE A : All the alternatives are available for each observation**

For a random sample of T aggregate observations, consider a system of M share equations in which the M theoretical independent probabilities P_{it}'s — interpreted as shares in this context — are specified by the following form:

$$P_{it} = \frac{\mathcal{U}_{it}(\Pi)}{\Sigma_m \mathcal{U}_{mt}(\Pi)} \text{ , } (i, m = 1, \ldots, M \text{ and } t = 1, \ldots, T) \tag{1}$$

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

For expositional purposes only, the utility function $\mathcal{U}_{it}(\Pi)$ can be represented by a general form that includes all the five share model families and also provides a theoretical framework formulated in Gaudry (1989) within which the relationships among the families can be examined:

$$\mathcal{U}_{it}(\Pi) = \sum_j w_{ij}\left(\tilde{\theta}_{ij}\right) f_j\left(g_j\left(\exp\left(V_{jt}\right); \tilde{\phi}_j, \tilde{\mu}_j\right); \phi_j, \mu_j\right) \geq 0 \text{ , } (i, j = 1, \ldots, M) \tag{2}$$

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_{jt};

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

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 (2) are respectively the direct and inverse Box-Tukey transformations on a real variable or function Z_j defined as follows:

$$g_j\left(Z_j; \tilde{\phi}_j, \tilde{\mu_j}\right) = Z_j^{\left(\tilde{\phi}_j, \tilde{\mu}_j\right)} = \begin{cases} \frac{1}{\tilde{\phi}_j}\left[\left(Z_j + \tilde{\mu_j}\right)^{\tilde{\phi}_j} - 1\right] & , if\ \tilde{\phi}_j \neq 0 \\ \ln\left(Z_j + \tilde{\mu}_j\right) & , if \tilde{\phi}_j = 0 \end{cases} \tag{3}$$

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

$$f_j(Z_j;\phi_j,\mu_j) = Z_j^{(\phi_j,\mu_j)^{-1}} = \begin{cases} (\phi_j Z_j + 1)^{\frac{1}{\phi_j}} - \mu_j & , if \phi_j \neq 0 \\ \exp(Z_j) - \mu_j & , if \phi_j = 0 \end{cases} \tag{4}$$

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

The direct Box-Tukey transformation (3) 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 $\exp(V_{jt})$ considered here. The inverse Box-Tukey transformation (4) 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 (3) without the location parameter is known as the Box-Cox transformation, which can be applied to an independent variable X_{jtk} specified in V_{jt} 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_{jtk} should be strictly positive so that $\left(X_{jtk}^{\lambda_k} - 1\right)/\lambda_k$ and $ln(X_{jtk})$ 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_{jt}. In the compact notation of $V_{jt} = \Sigma_k \beta_k X_{jtk}^{(\lambda_k)}$, it is clearly 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 $\mathcal{U}_{it}$ is not identified, different constraints on the "extra parameters", namely the $\theta, \phi, \mu, \tilde{\phi}$ and $\tilde{\mu}$ – parameters, should be imposed to yield an estimable form of $\mathcal{U}_{it}$ 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 $\phi_j, \mu_j, \tilde{\phi}_j$ 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 $\mathcal{U}_{it}$ is nonnegative over the whole range of $V_{it}(-\infty, +\infty)$, 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 $1 - \mu_i \geq 0$ and for **LIN-IPT-L**, and $\tilde{\mu}_i \geq 0$ for **BT-IPT-L**.

Assume that the error terms $\tilde{u}_{it}$'s are specified multiplicatively for the share equations (1), then the ith observed share S_{it} can be fitted against the theoretical probability P_{it} as follows:

$$S_{it} = \frac{\mathcal{U}_{it}(\Pi)\exp(\tilde{u}_{it})}{\Sigma_m[\mathcal{U}_{mt}(\Pi)\exp(\tilde{u}_{mt})]},\ (i = 1, \ldots, M) \tag{5}$$

where S_{it} is continuous over the open interval]0,1[and also satisfies the adding-up constraint $\Sigma_m S_{mt} = 1$, and the $\tilde{u}_{it}$'s are assumed to be serially independent and M-variate normal $N\left(0, \tilde{\Sigma}\right)$.

The system can then be linearized so that the residuals enter additively by using the logarithmic transformation on the share ratios:

$$\ln\frac{S_{it}}{S_{jt}} = \ln\frac{\mathcal{U}_{it}(\Pi)}{\mathcal{U}_{jt}(\Pi)} + \tilde{u}_{it} - \tilde{u}_{jt},\ (i, j = 1, \ldots, M) \tag{6}$$

This transformed system generates a set of $M(M-1)/2$ equations which are linearly dependent due to the adding-up constraint on the shares. Note that the covariance matrix $\tilde{\Sigma}$ of the residuals $\tilde{u}_{it}$'s is singular for the same reason. It can be shown (Barten 1969) that an estimable form of the system may include only $M-1$ linearly independent equations in which the share associated with the dropped equation is used as a reference share. Taking for example the Mth share as a reference share, the system now reduces to:

$$\ln \frac{S_{it}}{S_{Mt}} = \ln \frac{\mathcal{U}_{it}(\Pi)}{\mathcal{U}_{Mt}(\Pi)} + \tilde{u}_{it} - \tilde{u}_{Mt},\ (i = 1, \ldots, M-1) \tag{7}$$

or

$$\ln \frac{S_{it}}{S_{Mt}} = \ln \frac{P_{it}}{P_{Mt}} + u_{it} \tag{8}$$

where the new error terms u_{it}'s, expressed as the differences of the original residuals $\tilde{u}_{it} - \tilde{u}_{Mt}$, are also serially independent and $(M-1)$-variate normal $N(0, \Sigma)$. Note that due to the logarithmic transformation on the share ratios, each of the S_{mt}'s and P_{mt}'s $(m = 1, \ldots, M)$ is defined only for the open interval]0,1[.

TABLE 1 Relationships among the five Share 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 $\mathcal{U}_{it}(\Pi) \geq 0$
$\theta_{ii} = \theta_{jj} = 0$	$\phi_j = \mu_j = 1$	$\tilde{\phi}_j = \tilde{\mu}_j = 1$	**LOGIT** $\exp(V_{it})$
$\theta_{i1} = \ldots = \theta_{iM} = \theta_i \geq 0$	as above	as above	**S-DOGIT** $\exp(V_{it}) + \theta_i \sum_j \exp(V_{jt})$
$\theta_{ii} = 0$ and $\theta_{ij} \geq 0$	as above	as above	**G-DOGIT** $\exp(V_{it}) + \sum_{j \neq i} \theta_{ij} \exp(V_{jt})$
$\theta_{ii} = \theta_{ij} = 0$	$\phi_j \geq 0$ and $\mu_j \leq 1$	$\tilde{\phi}_j = \tilde{\mu}_j = 1$	**LIN-IPT-L** $[\phi_i \exp(V_{it}) + 1]^{\frac{1}{\phi_i}} - \mu_i$
$\theta_{ii} = \theta_{ij} = 0$	$\phi_j = \mu_j = 0$	$\tilde{\mu}_j \geq 0$	**BT-IPT-L** $\exp\left\{\frac{[\exp(V_{it}) + \tilde{\mu}_i]^{\tilde{\phi}_i} - 1}{\tilde{\phi} i}\right\}$

For ease of notation, the summations $\sum_{t=1}^{T}$ and $\sum_{k=1}^{K}$ will be abbreviated to $\sum_t$ and $\sum_k$ respectively, when no ambiguity arises in the text.

The system (8) has a nonlinear multivariate regression structure since no endogenous variable appears in the set of explanatory variables specified in the probability functions P_{it} and P_{Mt}. The associated log-likelihood function can therefore be written as:

$$\mathcal{L}(\Pi, \Sigma) = -\frac{T}{2}(M-1)\ln(2\pi) - \frac{T}{2}\ln(\det \Sigma) - \frac{1}{2}\sum_t \left(U_t' \Sigma^{-1} U_t\right) \tag{9}$$

where U_t is an $[(M-1) \times 1]$ vector of residuals, $u_{it} = \ln(S_{it}/S_{Mt}) - \ln(P_{it}/P_{Mt})$, and Σ is the associated covariance matrix of order $M-1$.

Three forms of Σ can be specified in the estimation of Π and Σ:

1. Full matrix Σ: all the different elements σ_{ij} of Σ are estimated. Since the matrix Σ is symmetric, $\sigma_{ij} = \sigma_{ji}\ (i \neq j)$, there are only $(M-1)M/2$ different elements to be estimated.
2. Diagonal matrix Σ: all the $M-1$ variances σ_{ii} are distinctly estimated, but all the covariances $\sigma_{ij}(i \neq j)$ are set equal to zero.
3. Scalar matrix Σ: all the variances are equal to a common scalar, say σ_u^2, which is estimated, and all the covariances are equal to zero.

Note that the model (8) is invariant with respect to the choice of the reference share only if the covariance matrix Σ is full, for the five families. Wills (1982) has shown this property for the LOGIT family with no Box-Cox transformation on the independent X_{jtk}'s. But the invariance still holds in the Box-Cox case since the property depends strictly on the fullness of the covariance matrix, but not on the functional form of the independent variables.

Setting the partial derivative of (9) with respect to Σ equal to zero, we obtain an estimate of Σ:

$$\hat{\Sigma} = \begin{cases} \left[\hat{\sigma}_{ij} = \frac{1}{T}\sum_t u_{it}u_{jt}, \forall i,j\right] & (\text{Full } \Sigma) \\ \left[\hat{\sigma}_{ii} = \frac{1}{T}\sum_t u_{it}^2, \sigma_{ij} = 0 (i \neq j)\right] & (\text{Diagonal } \Sigma) \\ \left[\hat{\sigma}_{ii} = \hat{\sigma}_u^2 = \frac{1}{T(M-1)}\sum_t \sum_{m=1}^{M-1} u_{mt}^2, \sigma_{ij} = 0 (i \neq j)\right] & (\text{Scalar } \Sigma) \end{cases} \tag{10}$$

where u_{it} is evaluated at the tried values of the parameters Π during the iterative procedure which maximizes the log-likelihood function; $\hat{\Sigma}$ will become only a maximum likelihood estimate when the maximum likelihood estimate $\hat{\Pi}$ is obtained at the end of the iterations.

Replacing Σ by $\hat{\Sigma}$ in (9), we get the log-likelihood function concentrated on Σ:

$$L(\Pi) = -\frac{T}{2}(M-1)[1 + \ln(2\pi)] - \frac{T}{2}\ln\left(\det \hat{\Sigma}\right). \tag{11}$$

As $\hat{\Sigma}$ is expressed as an implicit function of Π, maximizing $L(\Pi)$ with respect to Π only is equivalent to maximizing $\mathcal{L}(\Pi, \Sigma)$ for obtaining the joint maximum likelihood estimates of Π and Σ. This procedure is computationally more efficient since it reduces significantly the number of parameters explicitly involved, specially when Σ is a full matrix.

• **CASE B: Not all the alternatives are available for an observation**[1]

For this special case, since the two first forms of Σ are not identified, only the last form (Scalar Σ) can be estimated in the present version of the program. The system of equations (8) should be rewritten as:

$$\ln \frac{S_{it}}{S_{Rt}} = \ln \frac{P_{it}}{P_{Rt}} + u_{it}, \quad i \neq R \tag{12}$$

[1] This case was suggested to us by Richard Laferrière of University of Montreal in 1988.

where the residual u_{it} represents $\tilde{u}_{it} - \tilde{u}_{Rt}$, the alternative i is the ith alternative included in the set of M_t $(2 \leq M_t \leq M)$ available alternatives, say C_t, for an observation t, and the alternative R used as a reference should be available for every observation t in the estimation sample to ensure that the difference $\tilde{u}_{it} - \tilde{u}_{Rt}$ is always defined whenever an ith alternative is available for some observations t's.

Even if there are two or more alternatives which are always available in the estimation sample, the model is not invariant with respect to the reference alternative chosen among these alternatives. This is also true in the CASE A with a scalar matrix Σ, that can be considered as a limit case where all the alternatives are always available in the sample.

The residuals u_{it}'s are assumed to be serially independent and normally distributed with zero mean and covariance matrix $\Sigma_t = \sigma_u^2 I_{M_t-1}$, where σ_u^2 is the common variance of the u_{it}'s across the $M_t - 1$ equations and I_{M_t-1} is an identity matrix of order $M_t - 1$.

The log-likelihood function can now be written as:

$$\mathcal{L}(\Pi, \sigma_u^2) = -\frac{1}{2}\sum_t (M_t-1)\ln(2\pi) - \frac{1}{2}\sum_t \ln(\det \Sigma_t) - \frac{1}{2}\sum_t \mathrm{U_t}'\Sigma_t^{-1}\mathrm{U_t} \tag{13}$$

where U_t is a column vector of $(M_t - 1)$ elements u_{it}'s.

The function can be explicitly written in terms of the common variance σ_u^2:

$$\mathcal{L}(\Pi, \sigma_u^2) = -\frac{1}{2}\sum_t (M_t-1)\ln(2\pi) - \frac{1}{2}\sum_t (M_t - 1)\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_t \sum_{i\neq R} u_{it}^2 \tag{14}$$

Setting the partial derivative of (14) with respect to σ_u^2 equal to zero, an estimate of σ_u^2 is obtained:

$$\hat{\sigma}_u^2 = \frac{1}{\sum_t (M_t - 1)} \sum_t \sum_{i\neq R} u_{it}^2 \tag{15}$$

where the u_{it}'s are functions of Π only: $u_{it} = \ln(S_{it}/S_{Rt}) - \ln(P_{it}/P_{Rt})$.

Replacing σ_u^2 by $\hat{\sigma}_u^2$ in (14), the log-likelihood function concentrated on σ_u^2 can be written as:

$$L(\Pi) = -\frac{1}{2}\sum_t (M_t - 1)[1 + \ln(2\pi)] - \frac{1}{2}\sum_t (M_t - 1)\ln\hat{\sigma}_u^2. \tag{16}$$

As in CASE A, maximizing this function with respect to Π will yield the maximum likelihood estimate of Π and using this estimate to compute the u_{it}'s in (15) will give the maximum likelihood estimate of σ_u^2.

1.2 Computational Aspects

The Davidon-Fletcher-Powell (DFP) algorithm (Fletcher and Powell 1963) is used to maximize the concentrated log-likelihood function L in (11) with respect to the $(P \times 1)$ vector of parameters Π redefined as $\Pi = [\beta', \psi', \lambda']'$, where β is a $(K \times 1)$ vector of β-coefficients, ψ is an $(M \times 1)$ vector of θ_i's for **S-DOGIT**, an $[M(M-1) \times 1]$ vector of θ_{ij}'s $(i \neq j)$ for **G-DOGIT**, or a $(2M \times 1)$ vector of (ϕ_i, μ_i)'s or $\left(\tilde{\phi}_i, \tilde{\mu}_i\right)$'s for **LIN-IPT-L** or **BT-IPT-L** respectively, and λ is a $(K \times 1)$ vector of Box-Cox parameters which can be estimated in all families, except for **G-DOGIT** which includes only fixed Box-Cox parameters due the great number of estimated θ_{ij}'s.

- **CASE A**

The gradient of L which is used in DFP can be written as:

$$\frac{\partial L}{\partial \Pi} = \sum_t \frac{\partial L_t}{\partial \Pi} = -\sum_t \left(\frac{\partial U_t}{\partial \Pi}\right)' \hat{\Sigma}^{-1} U_t \tag{17}$$

where the typical elements of the $[(M-1) \times P]$ matrix $\partial U_t / \partial \Pi$ are given below for each family.

At the maximum point of the log-likelihood function L, the asymptotic covar iance matrix of the estimated parameters $\hat{\Pi}$ is evaluated by the method of Berndt, Hall, Hall and Hausman (1974) applied to the special case of the nonlinear multivariate regressions:

$$\text{Var}\left(\hat{\Pi}\right) = \left[\text{E}\left(\sum_t \frac{\partial L_t}{\partial \Pi}\frac{\partial L_t}{\partial \Pi'}\right)\right]^{-1}_{\Pi=\hat{\Pi}} = \left[\sum_t \left(\frac{\partial U_t}{\partial \Pi}\right)' \hat{\Sigma}^{-1} \left(\frac{\partial U_t}{\partial \Pi}\right)\right]^{-1}_{\Pi=\hat{\Pi}}. \tag{18}$$

- **CASE B**

The gradient of L has the following form:

$$\frac{\partial L}{\partial \Pi} = \sum_t \frac{\partial L_t}{\partial \Pi} = -\frac{1}{\hat{\sigma}_u^2} \sum_t \sum_{i \neq R} \left(\frac{\partial u_{it}}{\partial \Pi}\right) u_{it} \tag{19}$$

where the typical elements of the column vector $\partial u_{it} / \partial \Pi$ are given below for each family.

At the maximum point of the log-likelihood function L, the asymptotic covariance matrix of the estimated parameters $\hat{\Pi}$ is similarly computed as in CASE A:

$$\text{Var}\left(\hat{\Pi}\right) = \left[\text{E}\left(\sum_t \frac{\partial L_t}{\partial \Pi}\frac{\partial L_t}{\partial \Pi'}\right)\right]^{-1}_{\Pi=\hat{\Pi}} = \left[\frac{1}{\hat{\sigma}_u^2} \sum_t \sum_{i \neq R} \frac{\partial u_{it}}{\partial \Pi}\frac{\partial u_{it}}{\partial \Pi'}\right]^{-1}_{\Pi=\hat{\Pi}}. \tag{20}$$

From now on, the subscript t referring to the observations is omitted from u_{it}, P_{it}, V_{it} and X_{itk}.

LOGIT

$$\frac{\partial u_i}{\partial \beta_k} = X_{Mk}^{(\lambda_k)} - X_{ik}^{(\lambda_k)} \tag{21}$$

$$\frac{\partial u_i}{\partial \lambda_k} = \beta_k [\Delta_{Mk} - \Delta_{ik}] \tag{22}$$

S-DOGIT

$$\frac{\partial u_i}{\partial \beta_k} = \frac{X_{Mk}^{(\lambda_k)} + \theta_M \sum_j X_{jk}^{(\lambda_k)} \exp\left(V_j - V_M\right)}{1 + \theta_M \sum_j \exp\left(V_j - V_M\right)} - \frac{X_{ik}^{(\lambda_k)} + \theta_i \sum_j X_{jk}^{(\lambda_k)} \exp\left(V_j - Vi\right)}{1 + \theta_i \sum_j \exp\left(V_j - V_i\right)} \tag{23}$$

$$\frac{\partial u_i}{\partial \theta_m} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ -\dfrac{\sum_j \exp\left(V_j\right)}{\left[\exp\left(V_i\right) + \theta_i \sum_j \exp\left(V_j\right)\right]} & \text{, if } m = i \\ \dfrac{\sum_j \exp\left(V_j\right)}{\left[\exp\left(V_M\right) + \theta_M \sum_j \exp\left(V_j\right)\right]} & \text{, if } m = M \end{cases} \tag{24}$$

$$\frac{\partial u_i}{\partial \lambda_k} = \beta_k \left[\frac{\Delta_{Mk} + \theta_M \sum_j \Delta_{jk} \exp\left(V_j - V_M\right)}{1 + \theta_M \sum_j \exp\left(V_j - V_M\right)} - \frac{\Delta_{ik} + \theta_i \sum_j \Delta_{jk} \exp\left(V_j - V_i\right)}{1 + \theta_i \sum_j \exp\left(V_j - V_i\right)} \right] \tag{25}$$

G-DOGIT

$$\frac{\partial u_i}{\partial \beta_k} = \frac{X_{Mk}^{(\lambda_k)} + \sum_{j \neq M} \theta_{Mj} X_{jk}^{(\lambda_k)} \exp\left(V_j - V_M\right)}{1 + \sum_{j \neq M} \theta_{Mj} \exp\left(V_j - V_M\right)} - \frac{X_{ik}^{(\lambda_k)} + \sum_{j \neq i} \theta_{ij} X_{jk}^{(\lambda_k)} \exp\left(V_j - V_i\right)}{1 + \sum_{j \neq i} \theta_{ij} \exp\left(V_j - V_i\right)} \tag{26}$$

$$\frac{\partial u_i}{\partial \theta_{mr}} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ -\dfrac{\exp\left(V_r\right)}{\left[\exp\left(V_i\right) + \sum_{j \neq i} \theta_{ij} \exp\left(V_j\right)\right]} & \text{, if } m = i \text{ and } m \neq r \\ \dfrac{\exp\left(V_r\right)}{\left[\exp\left(V_M\right) + \sum_{j \neq M} \theta_{Mj} \exp\left(V_j\right)\right]} & \text{, if } m = M \text{ and } m \neq r \end{cases} \tag{27}$$

LIN-IPT-L

$$\frac{\partial u_i}{\partial \beta_k} = \frac{A_M^{\frac{1}{\phi_M}-1} X_{Mk}^{(\lambda_k)} \exp\left(V_M\right)}{A_M^{\frac{1}{\phi_M}} - \mu_M} - \frac{A_i^{\frac{1}{\phi_i}-1} X_{ik}^{(\lambda_k)} \exp\left(V_i\right)}{A_i^{\frac{1}{\phi_i}} - \mu_i} \tag{28}$$

$$\frac{\partial u_i}{\partial \phi_m} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ -\dfrac{A_i^{\frac{1}{\phi_i}-1}\left[\phi_i \exp\left(V_i\right) - A_i \ln A_i\right]}{\phi_i^2 \left(A_i^{\frac{1}{\phi_i}} - \mu_i\right)} & \text{, if } m = i \\ \dfrac{A_M^{\frac{1}{\phi_M}-1}\left[\phi_M \exp\left(V_M\right) - A_M \ln A_M\right]}{\phi_M^2 \left(A_M^{\frac{1}{\phi_M}} - \mu_M\right)} & \text{, if } m = M \end{cases} \tag{29}$$

Note that $\lim_{\phi_i \to 0} \partial u_i / \partial \phi_i = \frac{1}{2} \exp\left[\exp\left(V_i\right) + 2V_i\right] / \{\exp\left[\exp\left(V_i\right)\right] - \mu_i\}$ for $m = i$ and $\lim_{\phi_M \to 0} \partial u_i / \partial \phi_M = -\frac{1}{2} \exp\left[\exp\left(V_M\right) + 2V_M\right] / \{\exp\left[\exp\left(V_M\right)\right] - \mu_M\}$ for $m = M$.

$$\frac{\partial u_i}{\partial \mu_M} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ \left(A_i^{\frac{1}{\phi_i}} - \mu_i\right)^{-1} & \text{, if } m = i \\ -\left(A_M^{\frac{1}{\phi_M}} - \mu_M\right)^{-1} & \text{, if } m = M \end{cases} \tag{30}$$

$$\frac{\partial u_i}{\partial \lambda_k} = \beta_k \left[\frac{A_M^{\frac{1}{\phi_M}-1} \Delta_{Mk} \exp(V_M)}{A_M^{\frac{1}{\phi_M}} - \mu_M} - \frac{A_i^{\frac{1}{\phi_i}-1} \Delta_{ik} \exp(V_i)}{A_i^{\frac{1}{\phi_i}} - \mu_i} \right] \tag{31}$$

where $A_j = \phi_j \exp(V_j) + 1 > 0 \ (j = 1, \ldots, M)$ and the subscript t is omitted from A_{jt}. Note that $\lim_{\phi_j \to 0} A_j^{1/\phi_j} = \lim_{\phi_j \to 0} A_j^{1/\phi_j - 1} = \exp[\exp(V_j)]$.

BT-IPT-L

$$\frac{\partial u_i}{\partial \beta_k} = B_M^{\tilde{\phi}_M - 1} X_{Mk}^{(\lambda_k)} \exp(V_M) - B_i^{\tilde{\phi}_i - 1} X_{ik}^{(\lambda_k)} \exp(V_i) \tag{32}$$

$$\frac{\partial u_i}{\partial \tilde{\phi}_m} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ -\frac{1}{\tilde{\phi}_i} \left[B_i^{\tilde{\phi}_i} \ln B_i - B_i^{(\tilde{\phi}_i)} \right] & \text{, if } m = i \\ \frac{1}{\tilde{\phi}_M} \left[B_M^{\tilde{\phi}_M} \ln B_M - B_M^{(\tilde{\phi} M)} \right] & \text{, if } m = M \end{cases} \tag{33}$$

Note that $\lim_{\tilde{\phi} i \to 0} \partial u_i / \partial \tilde{\phi}_i = -\frac{1}{2} \ln^2 B_i$ for $m = i$ and $\lim_{\tilde{\phi}_M \to 0} \partial u_i / \partial \tilde{\phi}_M = \frac{1}{2} \ln^2 B_M$ for $m = M$.

$$\frac{\partial u_j}{\partial \tilde{\mu}_m} = \begin{cases} 0 & \text{, if } m \neq i \text{ and } M \\ -B_i^{\tilde{\phi}_i - 1} & \text{, if } m = i \\ B_M^{\tilde{\phi}_M - 1} & \text{, if } m = M \end{cases} \tag{34}$$

$$\frac{\partial u_i}{\partial \lambda_k} = \beta_k \left[B_M^{\tilde{\phi}_M - 1} \Delta_{Mk} \exp(V_M) - B_i^{\tilde{\phi}_i - 1} \Delta_{ik} \exp(V_i) \right] \tag{35}$$

where $B_j = \exp(V_j) + \tilde{\mu}_j > 0 \ (j = 1, \ldots, M)$ and $B_j^{(\tilde{\phi}_j)} = \left(B_j^{\tilde{\phi}_j} - 1 \right) / \tilde{\phi}_j$.Note that the subscript t is omitted from B_{jt}.

In (22), (25), (2) and (35), the derivative of $X_{jk}^{(\lambda_k)}$ with respect to $\lambda_k, \Delta_{jk} \ (j = 1, \ldots, M)$, is computed as follows:

$$\Delta_{jk} = \frac{\partial X_{jk}^{(\lambda_k)}}{\partial \lambda_k} = \begin{cases} \frac{1}{\lambda_k} \left[X_{jk}^{\lambda_k} \ln X_{jk} - X_{jk}^{(\lambda_k)} \right] & \text{, if } \lambda_k \neq 0 \\ \frac{1}{2} \ln^2 X_{jk} & \text{, if } \lambda_k = 0. \end{cases} \tag{36}$$

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 which vary only across the observations but not across the modes, e.g. aggregate indices of population or income associated with each origin-destination pair in the context of an intercity passenger modal market share model;
3. The network variables which represent the characteristics of the transportation modes and which vary both across the observations and the modes, e.g. aggregate indices of modal service levels (fare, frequency, travel time) associated with each O-D pair.

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 for **CASE A** where all the alternatives are available for every observation, but only the (S-LIN) and (S-BC) model types can be estimated for **CASE B** due to the fact that a network characteristic of an ith alternative can be unavailable when it also appears across the other alternatives for some observations, 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 $\mathcal{U}_{it}(\Pi)$ to a single term $\exp(V_{it})$ where the representative utility component V_{it} only contains a linear specification of the independent variables X_{itk}'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 $\mathcal{U}_{it}(\Pi)$ only depends on $f_i(\cdot)$, $g_i(\cdot)$ and $\exp(V_{it})$ 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 $\exp(V_{it})$.
iv. Every V_{it} includes:

   a. The independent variables X_{itk}'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, $V_{mt} - V_{it} = \sum_k \beta_k (X_{mtk} - X_{itk})$, 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_{mt}'s including V_{it} leaves P_{it} unchanged, although the level of every V_{mt} will shift by the same amount: $P^*_{it} = \exp(V_{it}+c)/\sum_m \exp(V_{mt}+c) = 1/\sum_m \exp(V_{mt}+c-V_{it}-c) = 1/\sum_m \exp(V_{mt}-V_{it}) = P_{it}$. Multiplying all V_{mt}'s including V_{it} by a positive factor d affects P_{it}: $P^*_{it} = \exp(dV_{it})/\sum_m \exp(dV_{mt}) = [\exp(V_{it})]^d/\sum_m [\exp(V_{mt})]^d = 1/\sum_m [\exp(V_{mt}-V_{it})]^d = 1/\sum_m [\exp(V_{mt}-V_{it})]^d \neq P_{it.}$ This means that the function $\sum_m [\exp(V_{mt})]^d$ 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_{itk}'s included in each V_{it} : $V_{it} = \sum_k \beta_k X_{itk}^{(\lambda_k)}$.

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:

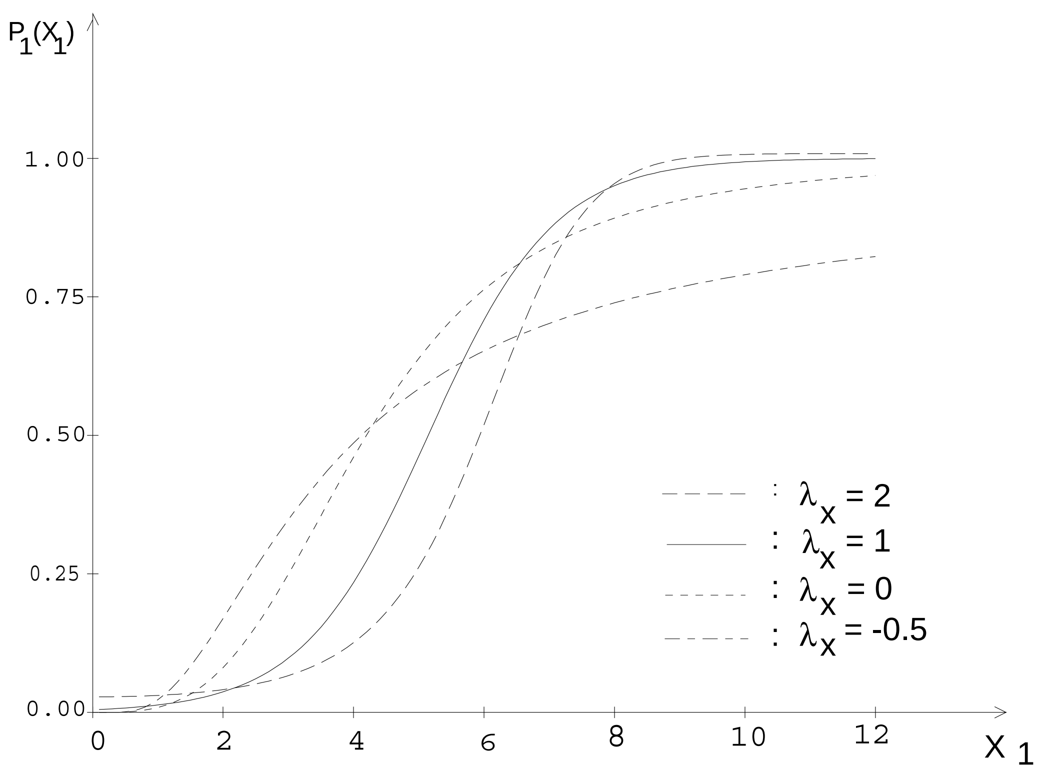
$$\frac{\partial P_i(X_{ik})}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\frac{\partial V_i}{\partial X_{ik}} = \frac{\partial P_i(V_i)}{\partial V_i}\beta_k X_{ik}^{\lambda_k - 1} \tag{38}$$

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.

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



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_{it} 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_{jt}'s $(j = 1, \ldots, 3)$ for an observation t can be written in matrix form as:

$$\begin{bmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{bmatrix} = \begin{bmatrix} X_{1t1} & 0 & 0 & X_{1t4} & 0 & 0 \\ 0 & X_{2t2} & 0 & 0 & X_{2t5} & 0 \\ 0 & 0 & X_{3t3} & 0 & 0 & X_{3t6} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{bmatrix} \tag{40}$$

where X_{1t1}, X_{2t2} and X_{3t3} are the alternative-specific constant terms $(\beta_1 \neq \beta_2 \neq \beta_3)$, and X_{1t4}, X_{2t5} and X_{3t6} 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_{1t1}, X_{2t2} and X_{3t3} are all equal to one for each observation t, the mode share P_{1t} can be expressed without explicitly writing them as follows:

$$P_{1t} = \frac{\exp(V_{1t})}{\sum_{j=1}^{3} \exp(V_{jt})} = \frac{\exp(\beta_1 + \beta_4 X_{1t4})}{\exp(\beta_1 + \beta_4 X_{1t4}) + \exp(\beta_2 + \beta_5 X_{2t5}) + \exp(\beta_3 + \beta_6 X_{3t6})}. \tag{41}$$

This expression can be represented in three equivalent forms:

$$P_{1t} = \begin{cases} \dfrac{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t4})}{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t4}) + \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2t5}) + \exp(\beta_6 X_{3t6})} \\ \dfrac{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1t4})}{\exp(\beta_1 - \beta_2)\exp(\beta_4 X_{1t4}) + \exp(\beta_5 X_{2t5}) + \exp(\beta_3 - \beta_2)\exp(\beta_6 X_{3t6})} \\ \dfrac{\exp(\beta_4 X_{1t4})}{\exp(\beta_4 X_{1t4}) + \exp(\beta_2 - \beta_1)\exp(\beta_5 X_{2t5}) + \exp(\beta_3 - \beta_1)\exp(\beta_6 X_{3t6})} \end{cases} \tag{42}$$

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 $\tilde{u}_{1t} - \tilde{u}_{3t}$ and $\tilde{u}_{2t} - \tilde{u}_{3t}$ 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_{1t4}, X_{2t5} and X_{3t6} represent the same socioeconomic variable $(X_{1t4} = X_{2t5} = X_{3t6})$ which is specified as alternative-specific $(\beta_4 \neq \beta_5 \neq \beta_6)$, P_{1t} can be rewritten in three equivalent forms:

$$P_{1t} = \begin{cases} \dfrac{\exp(\beta_1^*)\exp[(\beta_4-\beta_6)X_{1t4}]}{\exp(\beta_1^*)\exp[(\beta_4-\beta_6)X_{1t4}] + \exp(\beta_2^*)\exp[(\beta_5-\beta_6)X_{1t4}] + 1} \\ \dfrac{\exp(\beta_1^*)\exp[(\beta_4-\beta_5)X_{1t4}]}{\exp(\beta_1^*)\exp[(\beta_4-\beta_5)X_{1t4}] + \exp(\beta_2^*) + \exp[(\beta_6-\beta_5)X_{1t4}]} \\ \dfrac{\exp(\beta_1^*)}{\exp(\beta_1^*) + \exp(\beta_2^*)\exp[(\beta_5-\beta_4)X_{1t4}] + \exp[(\beta_6-\beta_4)X_{1t4}]} \end{cases} \tag{43}$$

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_{1t4}, X_{2t5} and X_{3t6} represent the same socioeconomic variable $(X_{1t4} = X_{2t5} = X_{3t6})$ but specified as generic $(\beta_4 = \beta_5 = \beta_6)$, P_{1t} can be rewritten as:

$$P_{1t} = \frac{\exp(\beta_1^*)\exp[\beta_4 X_{1t4}]}{\exp(\beta_1^*)\exp[\beta_4 X_{1t4}] + \exp(\beta_2^*)\exp[\beta_4 X_{1t4}] + \exp[\beta_4 X_{1t4}]} \tag{44}$$

which reduces to:

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

It can be seen that the socioeconomic variable has no influence at all on the mode share P_{1t} 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_{1t} can be expressed as:

$$P_{1t} = \frac{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1t4}^{(\lambda_4)}\right]}{\exp(\beta_1^*)\exp\left[\beta_4^* X_{1t4}^{(\lambda_4)}\right] + \exp(\beta_2^*)\exp\left[\beta_5^* X_{1t4}^{(\lambda_5)}\right] + 1} \tag{46}$$

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_{1t4}, X_{2t5} and X_{3t6} which represent the same socioeconomic variable $(X_{1t4} = X_{2t5} = X_{3t6})$, the mode share P_{1t} can be rewritten as:

$$P_{1t} = \frac{\exp(\beta_1^*)\exp\left[\beta_4 X_{1t4}^{(\lambda_4)}\right]}{\exp(\beta_1^*)\exp\left[\beta_4 X_{1t4}^{(\lambda_4)}\right] + \exp(\beta_2^*)\exp\left[\beta_5 X_{1t4}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{1t4}^{(\lambda_6)}\right]}. \tag{47}$$

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

Generic case

In (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_{1t4}, X_{2t5} and X_{3t6} represent the travel times associated with the three modes respectively $(X_{1t4} \neq X_{2t5} \neq X_{3t6})$ 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_{1t4}, X_{2t5} and X_{3t6}, if they are strictly positive, are subject to the Box-Cox transformations.

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_{1t4}, X_{2t5} and X_{3t6} representing the travel times associated with the three modes respectively ($X_{1t4} \neq X_{2t5} \neq X_{3t6}$), the mode share P_{1t} can be expressed as:

$$P_{1t} = \frac{\exp\left(\beta_1^*\right) \exp\left[\beta_4 X_{1t4}^{(\lambda_4)}\right]}{\exp\left(\beta_1^*\right) \exp\left[\beta_4 X_{1t4}^{(\lambda_4)}\right] + \exp\left(\beta_2^*\right) \exp\left[\beta_5 X_{2t5}^{(\lambda_5)}\right] + \exp\left[\beta_6 X_{3t6}^{(\lambda_6)}\right]}. \tag{48}$$

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.

Generic case

In (48), if the variables X_{1t4}, X_{2t5} and X_{3t6} 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 $(j \neq i)$.

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 $M-1$ V_{jt}'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_{jt}'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_{jt}' s involving in the mode share P_{1t} can be represented as follows:

$$\begin{bmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1t4}^{(\lambda_4)} & 0 & 0 & 0 & X_{2t5}^{(\lambda_8)} & X_{3t6}^{(\lambda_9)} \\ 0 & 1 & 0 & X_{2t5}^{(\lambda_5)} & 0 & X_{1t4}^{(\lambda_7)} & 0 & 0 \\ 0 & 0 & 0 & 0 & X_{3t6}^{(\lambda_6)} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_9 \end{bmatrix} \tag{49}$$

where the characteristic of an ith mode appears in its own mode and one of the other modes, i.e. X_{1t4} is specified in modes 1 and 2, X_{2t5} in modes 2 and 1, and X_{3t6} 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_{1t4} : $\beta_4 \neq \beta_7$ and $\lambda_4 \neq \lambda_7$,
- for X_{2t5} : $\beta_5 \neq \beta_8$ and $\lambda_5 \neq \lambda_8$,
- for X_{3t6} : $\beta_6 \neq \beta_9$ 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_{1t4} : $\beta_4 = \beta_7$ and $\lambda_4 = \lambda_7$,
- for X_{2t5} : $\beta_5 = \beta_8$ and $\lambda_5 = \lambda_8$,
- for X_{3t6} : $\beta_6 = \beta_9$ and $\lambda_6 = \lambda_9$,
- with $\lambda_4 \neq \lambda_5 \neq \lambda_6$.
- <u>Subcase II</u>

This subcase is in fact a generalization of the subcase I since each representative utility component V_{jt} 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:

$$\begin{bmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & X_{1t4}^{(\lambda_4)} & 0 & 0 & 0 & 0 & X_{2t5}^{(\lambda_9)} & 0 & X_{3t6}^{(\lambda_{11})} & 0 \\ 0 & 1 & 0 & X_{2t5}^{(\lambda_5)} & 0 & X_{1t4}^{(\lambda_7)} & 0 & 0 & 0 & 0 & X_{3t6}^{(\lambda_{12})} \\ 0 & 0 & 0 & 0 & X_{3t6}^{(\lambda_6)} & 0 & X_{1t4}^{(\lambda_8)} & 0 & X_{2t5}^{(\lambda_{10})} & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \beta_4 \\ \vdots \\ \beta_{12} \end{bmatrix} \tag{50}$$

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_{1t4} : $\beta_4 \neq \beta_7 \neq \beta_8$ and $\lambda_4 \neq \lambda_7 \neq \lambda_8$,
- for X_{2t5} : $\beta_5 \neq \beta_9 \neq \beta_{10}$ and $\lambda_5 \neq \lambda_9 \neq \lambda_{10}$,
- for X_{3t6} : $\beta_6 \neq \beta_{11} \neq \beta_{12}$ 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_{1t} : $\lambda_4 = \lambda_9 = \lambda_{11}$,
- for V_{2t} : $\lambda_5 = \lambda_7 = \lambda_{12}$,
- for V_{3t} : $\lambda_6 = \lambda_8 = \lambda_{10}$,
- 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_{1t4} with one common λ-parameter $(\lambda_4 = \lambda_7 = \lambda_8)$, cannot be specified across all the alternatives:

- for X_{1t4} : $\beta_4 = \beta_7 = \beta_8$ and $\lambda_4 = \lambda_7 = \lambda_8$,
- for X_{2t5} : $\beta_5 = \beta_9 = \beta_{10}$ and $\lambda_5 = \lambda_9 = \lambda_{10}$,
- for X_{3t6} : $\beta_6 = \beta_{11} = \beta_{12}$ 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_{1t4}, X_{2t5} and X_{3t6} 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_{it}, but not in some or all of the other V_{jt}'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_{it} but also in the other V_{jt}'s:	same as the alternative-specific case:
	• Subcase I (in some V_{jt}'s): the alternative-specific and generic subcases are allowed.	• Subcase I : only the alternative-specific is allowed.
	• Subcase II (in all V_{jt}'s): only the alternative-specific subcase 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:

$$P_i(V_i) = \frac{\exp(V_i) + \theta_i \sum_j \exp(V_j)}{\sum_m \left[\exp(V_m) + \theta_m \sum_h \exp(V_h)\right]}, (\theta_i \geq 0), \tag{51}$$

(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:

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

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:

$$\begin{aligned} P_i(V_i)_{SD} &= \frac{\exp(V_i) + \theta_i \sum_j \exp(V_j)}{\sum_m \exp(V_m) + \sum_m \theta_m \sum_h \exp(V_h)} \\ &= \frac{P_i(V_i)_L + \theta_i}{1 + \sum_m \theta_m} = \frac{1}{1 + \sum_m \theta_m} P_i(V_i)_L + \frac{\theta_i}{1 + \sum_m \theta_m} \end{aligned} \tag{53}$$

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:

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

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:

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

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 $P_i(V_i)_{GD} < 1/\sum\limits_{m\neq i} \theta_{mi}$. 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: $\lim\limits_{V_i\to\infty} P_i(V_i)_{GD} = 1/\left(1 + \sum\limits_{m\neq i} \theta_{mi}\right)$, 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 $P_i(V_i)_L$:

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

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:

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

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:

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

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:

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

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 $\prod\limits_k [\exp(X_{jk} - X_{ik})]^{\beta_k}$ 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 $\ln[1 + \theta_{ij} \exp(\beta_{j0} - \beta_{i0})]$. 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.

(e) 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:

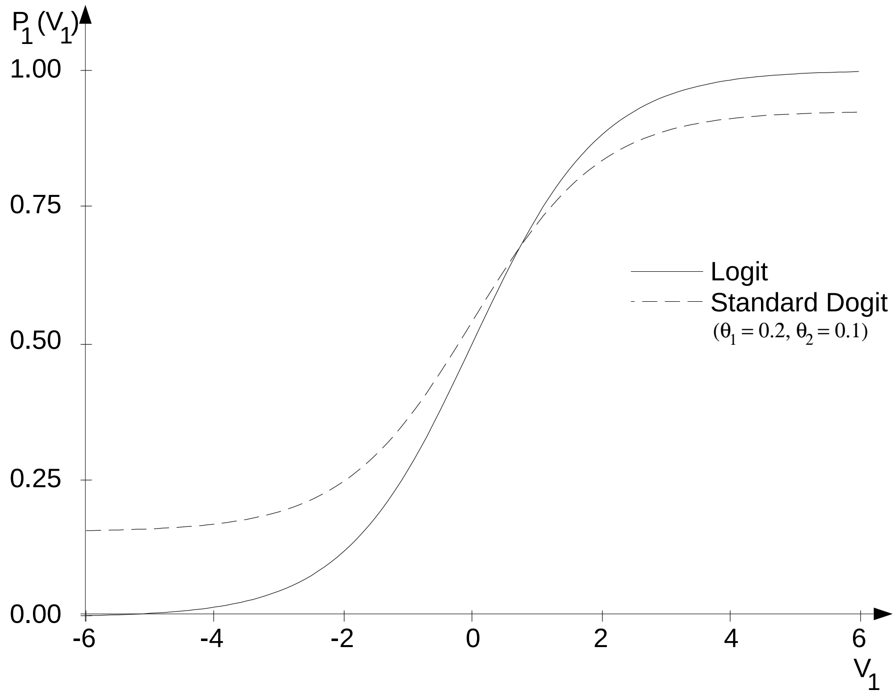
$$1-\lim_{V_i\to\infty} P_i(V_i)_{SD} = 1-\left[\frac{1}{1+\sum\limits_m \theta_m} \lim_{V_i\to\infty} P_i(V_i)_L + \frac{\theta_i}{1+\sum\limits_m \theta_m}\right]. \tag{65}$$

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 $P_i(V_i)_L$:

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

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:

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

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.

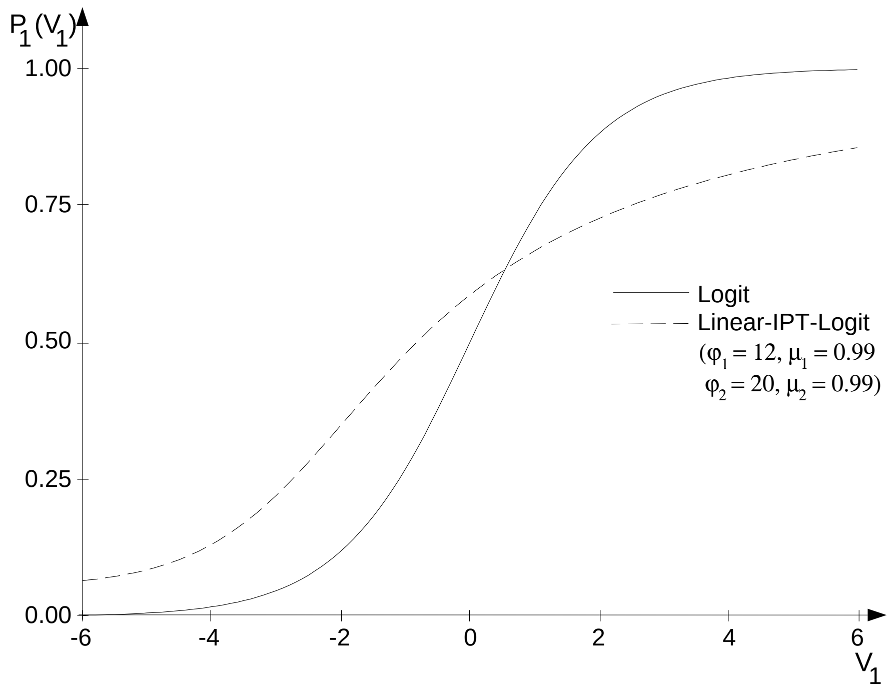
(d.2) The probability curve $P_i(X_{ik})_{LIL}$ in function of an independent variable X_{ik} is also asymmetric with respect to its inflection point since its slope is just in constant proportion to the slope of the asymmetric probability curve $P_i(V_i)_{LIL}$:

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

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:

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

  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:

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

• **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 $P_i(V_i)_L$:

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

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

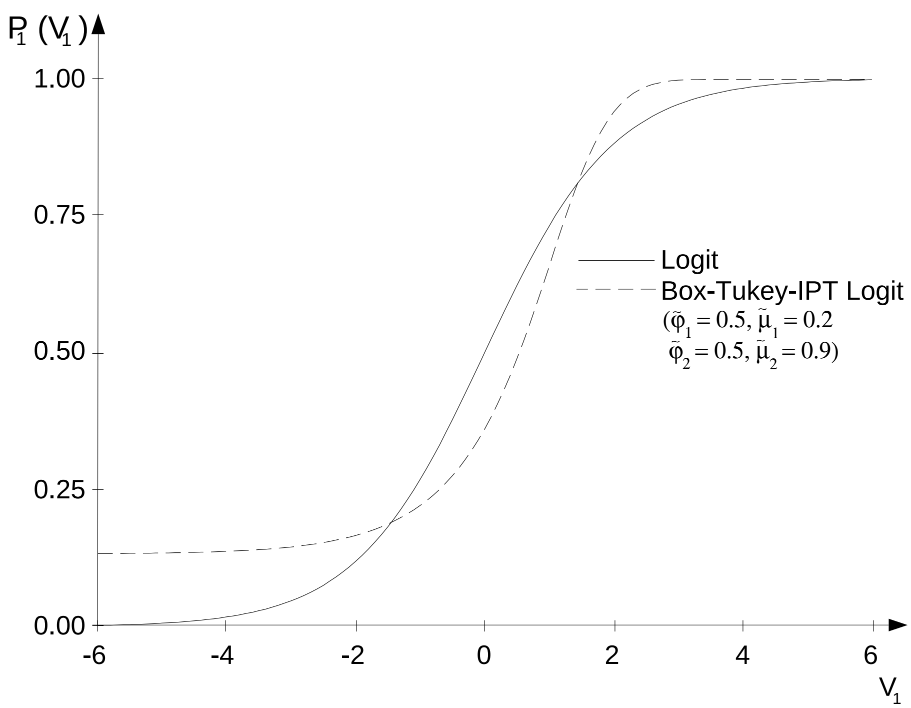
(d.2) The probability curve $P_i(X_{ik})_{BIL}$ in function of X_{ik} is also asymmetric with respect to its inflection point for the same reason. Its slope can be written as:

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

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:

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

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:

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

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

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

  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:

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

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.

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.

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

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 probabilitty curve of the

LOGIT 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 $\partial V_i/\partial X_{\cdot k}$ is equal to β_k^i and $\partial V_j/\partial X_{\cdot k}$ is equal to β_k^j if V_j is function of $X_{\cdot k}$, 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 $P_i(X_{\cdot k})_{BIL}$ 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:

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

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

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.	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: 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.	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: 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

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).

5 MAIN RESULTS

5.1 Elasticity and Marginal Rate of Substitution (Value of Time)

Elasticity

The point elasticity of choosing alternative i with respect to the kth independent variable which describes the utility of the hth alternative for observation t is defined as:

$$E_{htk}^{it} = \frac{\partial P_{it}}{\partial X_{htk}} \frac{X_{htk}}{P_{it}} \tag{97}$$

The program computes the point elasticity E_{htk}^{it} and its components, namely the derivative $\partial P_{it}/\partial X_{htk}$ and the variation of P_{it} relative to a percent variation of X_{htk}, $(\partial P_{it}/\partial X_{htk})X_{htk}$, at the sample means of all the independent variables in their original units. Note that E_{htk}^{it} and $(\partial P_{it}/\partial X_{htk})X_{htk}$ are invariant to changes of measurement units in X_{htk}, but not $\partial P_{it}/\partial X_{htk}$.

The derivative $\partial P_{it}/\partial X_{htk}$ is given by the following formulas in which the subscript t is omitted from P_{it}, X_{htk}, V_{mt}, D_{htk}^{mt} and A_{mt}:

LOGIT

$$\frac{\partial P_i}{\partial X_{hk}} = \frac{D_{hk}^i - P_i \sum\limits_m D_{hk}^m}{\sum\limits_m \exp\left(V_m\right)} \tag{98}$$

S-DOGIT

$$\frac{\partial P_i}{\partial X_{hk}} = \frac{\left(D_{hk}^i + \theta_i \sum\limits_j D_{hk}^j\right) - P_i \sum\limits_m \left(D_{hk}^m + \theta_m \sum\limits_j D_{hk}^j\right)}{\sum\limits_m \left[\exp\left(V_m\right) + \theta_m \sum\limits_j \exp\left(V_j\right)\right]} \tag{99}$$

G-DOGIT

$$\frac{\partial P_i}{\partial X_{hk}} = \frac{\left(D_{hk}^i + \sum\limits_{j \neq i} \theta_{ij} D_{hk}^j\right) - P_i \sum\limits_m \left(D_{hk}^m + \sum\limits_{j \neq m} \theta_{mj} D_{hk}^j\right)}{\sum\limits_m \left[\exp\left(V_m\right) + \sum\limits_{j \neq m} \theta_{mj} \exp\left(V_j\right)\right]} \tag{100}$$

LIN-IPT-L

$$\frac{\partial P_i}{\partial X_{hk}} = \frac{\left[A_i^{\frac{1}{\phi_i}-1} D_{hk}^i\right] - P_i \sum\limits_m \left[A_m^{\frac{1}{\phi_m}-1} D_{hk}^m\right]}{\sum\limits_m \left(A_m^{\frac{1}{\phi_m}} - \mu_m\right)} \tag{101}$$

BT-IPT-L

$$\frac{\partial P_i}{\partial X_{hk}} = \frac{\left[B_i^{\tilde{\phi}_i - 1} D_{hk}^i \exp\left(B_i^{(\tilde{\phi}_i)}\right)\right] - P_i \sum\limits_m \left[B_m^{\tilde{\phi}_m - 1} D_{hk}^m \exp\left(B_m^{(\tilde{\phi}_m)}\right)\right]}{\sum\limits_m \exp\left(B_m^{(\tilde{\phi}_m)}\right)} \tag{102}$$

where

$$D_{hk}^m = \exp\left(V_m\right) \frac{\partial V_m}{\partial X_{hk}} \tag{103}$$

and

$$\frac{\partial V_m}{\partial X_{hk}} = \begin{cases} 0 & \text{, if } X_{hk} \text{ is not present in alternative } m \neq h, \\ \beta_r X_{mr}^{\lambda_r - 1} & \text{, if } X_{hk} \text{ is alternative-specific and also appears} \\ & \text{in alternative } m \neq h, \text{ with a coefficient } \beta_r \neq \beta_k, \\ \beta_k X_{mk}^{\lambda_k - 1} & \text{, if } X_{hk} \text{ is generic and also appears in alternative} \\ & m \neq h, \text{ with a same coefficient } \beta_k, \\ \beta_k X_{hk}^{\lambda_k - 1} & \text{, if } m = h \text{ and } X_{hk} \text{ is either generic or alternative-specific.} \end{cases} \tag{104}$$

Note that the explicit form of $\partial V_m / \partial X_{hk}$ 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.

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 given observation t, the ratio of the estimated shares for any pair of alternatives remains the same irrespective of the total number M of choices considered. If X_{hk} appears uniquely in the hth alternative, the cross elasticities are then uniform across the alternatives $i \neq h$ since they do not depend on the estimated shares P_i's:

$$E_{hk}^{i} = \beta_k(\delta_{ih} - P_h) X_{hk}^{\lambda_k} \tag{105}$$

where $\delta_{ih} = 1$, if $i = h$ (direct elasticity) and $\delta_{ih} = 0$, if $i \neq h$ (cross elasticity).

In all other cases, the cross elasticities will not be equal:

i. If a socioeconomic variable appears in more than one alternative;
ii. If a network characteristic of a given mode which usually appears in its own alternative only as in the Standard Linear or Box-Cox Logit, may be present in other alternatives as allowed in the Generalized Box-Cox Logit.

Corrected elasticities and percentage points for a quasi-dummy or a real dummy

The concept of elasticity strictly implies that an independent variable X_{hk} 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 elasticity at the sample means of the independent variable 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.

Category 3: Real dummy which has only two values: 0 and a positive constant, for example a binary (0–1) variable such as a dummy for an O-D pair within the same province 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 elasticity at the sample means is 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 elasticity is 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_{kh} is corrected for the portion of the positive values of the variable as follows:

$$\tilde{E}^{it}_{htk}|_{\bar{X}} = E^{it}_{htk}|_{\bar{X}} \frac{\bar{X}^{+}_{hk}}{\bar{X}_{hk}}, \tag{106}$$

where $E^{it}_{htk}|_{\bar{X}}$ is given in (97) evaluated at the sample means, $\bar{X}_{hk}$ is the sample mean computed from the whole sample for a socioeconomic variable and from the subset of observations available in alternative h for a network variable, and $\bar{X}^{+}_{hk}$ 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 h for a network variable.

Note that the change of P_{it} (the change in percentage points) relative to a percent variation of X_{htk} at the sample means is similarly corrected:

$$\tilde{C}^{it}_{htk} \Big|_{\bar{X}} = \frac{\partial P_{it}}{\partial X_{htk}} X_{htk} \Big|_{\bar{X}} \frac{\bar{X}^{+}_{hk}}{\bar{X}_{hk}}. \tag{107}$$

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:

$$\text{MRS}\,^{V_i}_{X_{ik},X_{ir}} = \frac{\partial V_i/\partial X_{ir}}{\partial V_i/\partial X_{ik}} \tag{108}$$

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 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:

$$\text{VOT}\,^{Vi}_{C_i,T_i} = \frac{\partial V_i/\partial T_i}{\partial V_i/\partial C_i} \tag{109}$$

For notational compactness, 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 (see Section 1.3).

- **(S–LIN) Model Type**

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

$$\text{VOT}\,_{ii}^{\,i} = \frac{\beta_{T_i}^i}{\beta_{C_i}^i} \tag{110}$$

where $\beta_{T_i}^i$ and $\beta_{C_i}^i$ 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 C_i and T_i : with a Box-Cox transformation on each variable:

$$\text{VOT}\,_{ii}^{\,i} = \frac{\beta_{T_i}^i T_i^{\lambda_{T_i}^i - 1}}{\beta_{C_i}^i C_i^{\lambda_{C_i}^i - 1}} \tag{111}$$

where $\lambda_{T_i}^i$ and $\lambda_{C_i}^i$ 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 C_i and T_j 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:

$$\text{VOT}^{\,i} = \left[\text{VOT}\,_{jh}^{\,i}\right] = \begin{bmatrix} \text{VOT}\,_{11}^{\,i} & \text{VOT}\,_{12}^{\,i} & \dots & \text{VOT}\,_{1M}^{\,i} \\ \text{VOT}\,_{21}^{\,i} & \text{VOT}\,_{22}^{\,i} & \dots & \text{VOT}\,_{2M}^{\,i} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}\,_{M1}^{\,i} & \text{VOT}\,_{M2}^{\,i} & \dots & \text{VOT}\,_{MM}^{\,i} \end{bmatrix} \tag{112}$$

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

$$\text{VOT}\,_{jh}^{\,i} = \frac{\partial V_i / \partial T_h}{\partial V_i / \partial C_j} = \frac{\beta_{T_h}^i T_h^{\lambda_{T_h}^i - 1}}{\beta_{C_j}^i C^{\lambda_{C_j}^i - 1}} \tag{113}$$

These VOT$\,_{jh}^{\,i}$'s are unique since each VOT$\,_{jh}^{\,i}$ is equal to VOT$\,_{jh}^{\,m}$, $\forall m \neq i$. In other words, they are invariant with respect to the utility component chosen. 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:

$$\text{VOT} = \left[\text{VOT}\,_{jh}^{\,j}\right] = \begin{bmatrix} \text{VOT}\,_{11}^{\,1} & \text{VOT}\,_{12}^{\,1} & \dots & \text{VOT}\,_{1M}^{\,1} \\ \text{VOT}\,_{21}^{\,2} & \text{VOT}\,_{22}^{\,2} & \dots & \text{VOT}\,_{2M}^{\,2} \\ \vdots & \vdots & \ddots & \vdots \\ \text{VOT}\,_{M1}^{\,M} & \text{VOT}\,_{M2}^{\,M} & \dots & \text{VOT}\,_{MM}^{\,M} \end{bmatrix} \tag{114}$$

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

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 (S–LIN) and (S–BC) model types, where the direct VOT's are defined and already given in (110) and (111) 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 involved 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** $\mathrm{VOT}^{\,j}_{jh}$ **in terms of the estimated share** P_j

In the program, for computational convenience, the $\mathrm{VOT}^{\,j}_{jh}$'s in (114) 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.

For CASE A defined in Section 1.1, 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. Similar results can be obtained for CASE B defined in Section 1.1.

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

$$V_j = V_j(C, T, etc.) \tag{115}$$

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

By definition, for every family, the estimated share P_j is a function of all the V_m's:

$$P_j = P_j(V) \tag{116}$$

where $V = [V_1, \ldots, V_M]$.

Analogous to the matrix VOT in (114), the matrix $\widetilde{VOT}$ computed in terms of the estimated shares P_j's can be defined as $\widetilde{VOT} = \left[\widetilde{VOT}^{j}_{jh}\right]$ where a typical element is:

$$\widetilde{VOT}^{j}_{jh} = \frac{\partial P_j/\partial T_h}{\partial P_j/\partial C_j} \tag{117}$$

Using the chain rule for functions of functions, the derivative of P_j with respect to T_h in the numerator of (117) can be written as:

$$\begin{aligned}\frac{\partial P_j}{\partial T_h} &= \frac{\partial P_j}{\partial V_1}\frac{\partial V_1}{\partial T_h} + \frac{\partial P_j}{\partial V_2}\frac{\partial V_2}{\partial T_h} + \ldots + \frac{\partial P_j}{\partial V_M}\frac{\partial V_M}{\partial T_h} = \sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial T_h} \\ &= \frac{\partial V_j}{\partial T_h}\sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial T_h}\frac{\partial T_h}{\partial V_j} = \frac{\partial V_j}{\partial T_h}\sum_{i=1}^{M}\frac{\partial P_j}{\partial V_i}\frac{\partial V_i}{\partial V_j}\end{aligned} \tag{118}$$

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:

$$\frac{\partial P_j}{\partial T_h} = \frac{\partial V_j}{\partial T_h}\frac{\partial P_j}{\partial V_j} \tag{119}$$

Similarly, the derivative of P_j with respect to C_j in the denominator of (117) can be shown to be:

$$\frac{\partial P_j}{\partial C_j} = \frac{\partial V_j}{\partial C_j}\frac{\partial P_j}{\partial V_j} \tag{120}$$

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 $\widetilde{VOT}^j_{jh}$ is identical to VOT$^j_{jh}$ in (114).

For the (S–LIN) and (S–BC) model types where each V_j is exclusively a function of C_j and T_j, only the diagonal elements $(j = h)$ of $\widetilde{VOT}$ are defined:

$$\widetilde{VOT}^j_{jj} = \frac{\partial P_j/\partial T_j}{\partial P_j/\partial C_j} = \frac{\partial P_j/\partial V_j}{\partial P_j/\partial V_j}\frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \frac{\partial V_j/\partial T_j}{\partial V_j/\partial C_j} = \text{VOT}^j_{jj} \tag{121}$$

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

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

$$\left.\widetilde{VOT}^j_{jh}\right|_{\bar{X}} = \left.\frac{\partial P_j/\partial T_h}{\partial P_j/\partial C_j}\right|_{\bar{X}} = \frac{\beta^j_{T_h}\bar{T}_h^{\lambda^j_{T_h}-1}}{\beta^j_{C_j}\bar{C}_j^{\lambda^j_{C_j}-1}} \tag{122}$$

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 estimated shares if X_{jk} also appears across the alternatives, the value of time at the sample means $\left.\widetilde{VOT}^j_{jh}\right|_{\bar{X}}$ 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 $\left(\lambda^j_{C_j} = \lambda^j_{T_j} = 1\right)$.

5.2 Other Useful Statistics

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.

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:

$$X_i' X_i = \sum_{k=1}^{K_i} \gamma_k q_k q_k' \tag{123}$$

where q_k is the kth eigenvector $(K_i \times 1)$ 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 $X_i : \kappa(X_i) = (\gamma_{max}/\gamma_{min})^{1/2}$, which measures the sensivity of b to changes in $X_i' X_i$ or $X_i' Y_i$ in linear systems like the normal equations $X_i' X_i b = X_i' Y_i$;

ii. Condition indexes: $\eta_k = (\gamma_{max}/\gamma_k)^{1/2}, (k = 1, \ldots, K_i)$.

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:

$$\text{Var}(b_k) = \sigma^2 \sum_{m=1}^{K_i} \left(q_{km}^2 / \gamma_m \right). \tag{124}$$

The proportion of Var(b_k) associated with any γ_r is then computed:

$$\Pi_{rk} = \left(q_{kr}^2 / \gamma_r \right) / \sum_{m=1}^{K_i} \left(q_{km}^2 / \gamma_m \right). \tag{125}$$

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

TABLE 6 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)	$\dots$	Var(b_{K_i})
$\gamma_i(\gamma_{min})$	$\eta_1 = \kappa(X_i)$	Π_{11}	Π_{12}	$\dots$	Π_{1K_i}
γ_2	η_2	Π_{21}	Π_{22}	$\dots$	Π_{2K_i}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
$\gamma_{K_i}(\gamma_{max})$	$\eta_{K_i} = 1$	Π_{K_i1}	Π_{K_i2}	$\dots$	$\Pi_{K_iK_i}$

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

i. High values of the condition indexes 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.

Correlation Matrix of Residuals

For CASE A, according to (10), the estimated covariance matrix $\hat{\Sigma}$ of the residuals u_{it}'s across the equations defined in (8) is computed at the maximum likelihood estimates of the parameters Π and can be given in terms of the correlations r_{ij}'s between the residuals u_i and u_j of the ith and jth equations by using the following relation:

$$r_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \tag{126}$$

where $\hat{\sigma}_{ii}$ and $\hat{\sigma}_{jj}$ are the estimated variances of u_i and u_j respectively, and $\hat{\sigma}_{ij}$ is the estimated covariance between u_i and u_j. Note that for the last two cases in (10), diagonal and scalar Σ, the correlation matrix of the residuals reduces to an identity matrix of order $M-1$ since the covariances, hence the correlations, of the residuals across the equations are constrained equal to zero.

For CASE B, since the order of the covariance matrix of the residuals Σ_t varies with each observation t and the matrix is scalar, i.e. all the variances of the residuals across the equations are equal to a common scalar σ_u^2 and all covariances are zero, it is not useful to compute the correlation matrix of the residuals which reduces to an identity matrix of order $M_t - 1$ for each observation t.

t-statistics

i) β-coefficients

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

• **Unconditional** t-statistics computed from the standard errors given by the square roots of the diagonal elements of the covariance matrices (18) and (20) for CASES A and B respectively. 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 for this case. Note that when some extra parameters or Box-Cox parameters are fixed, the t-statistics of the estimated β-coefficients are implicitly conditional on the fixed values of the former.

• T-statistics **conditional** on the estimated λ-values and the fixed λ-values if any, computed from the standard errors given by the square roots of the diagonal elements of the covariance matrices:

$$\operatorname{Var}\left(\hat{\Pi}^* | \lambda\right) = \begin{cases} \left[\sum_t \left(\frac{\partial U_t}{\partial \Pi^*}\right)' \hat{\Sigma}^{-1} \left(\frac{\partial U_t}{\partial \Pi^*}\right)\right]^{-1}_{\Pi=\hat{\Pi}} & \text{for CASE A} \\ \left[\frac{1}{\hat{\sigma}_u^2} \sum_t \sum_{i \neq R} \frac{\partial u_{it}}{\partial \Pi^*} \frac{\partial u_{it}}{\partial \Pi^{*\prime}}\right]^{-1}_{\Pi=\hat{\Pi}} & \text{for CASE B} \end{cases} \tag{127}$$

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 second row is not given since it is identical to the first row.

ii) Extra parameters and Box-Cox parameters

For the estimated parameters, only **unconditional** t-statistics are computed from the standard errors given by the square roots of the diagonal elements of the covariance matrices (18) and (20) for CASES A and B respectively, since they 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 hypothesis associated with the estimated extra parameters is as follows:

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

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

LIN-IPT-L: $H_0 \,:\, \phi_i$ or $\mu_i = \begin{cases} 0 & (\text{first row}) \\ 1 & (\text{second row}) \end{cases}$,

BT-IPT-L: $H_0 \,:\, \tilde{\phi}_i$ or $\tilde{\mu}_i = \begin{cases} 0 & (\text{first row}) \\ 1 & (\text{second row}) \end{cases}$.

The null hypothesis for the estimated Box-Cox parameters in **LOGIT**, **S-DOGIT**, **LIN-IPT-L** and **BT-IPT-L** is $H_0 \,:\, \lambda_k = \begin{cases} 0 & (\text{first row}) \\ 1 & (\text{second row}) \end{cases}$.

Other Statistics

i) Mean share :

• observed: $\bar{S}_i = \sum_t S_{it}/T, (i = 1, \ldots, M)$.

• estimated: $\bar{\hat{P}}_i = \sum_t \hat{P}_{it}/T, (i = 1, \ldots, M)$, where $\hat{P}_{it}$ is the value of P_{it} estimated at $\Pi = \hat{\Pi}$.

Note that for the **case A** where every observation t in the estimation sample has always M available alternatives, **the sum of observed mean shares $\bar{S}_i$'s over M alternatives is always equal to one** and the same is true for the sum of estimated mean shares $\bar{\hat{P}}_i$'s. But for the **case B** where some observations t's in the estimation sample have less than M available alternatives, **this adding up property no longer holds** since each observed mean share $\bar{S}_i$ and each estimated mean share $\bar{\hat{P}}_i$ can be computed only from the subset of T_i observations where the alternative i is available: $\bar{S}_i = \sum_t S_{it}/T_i$ and $\bar{\hat{P}}_i = \sum_t \hat{P}_{it}/T_i$. To preserve this property, one can be tempted to obtain global measures for the observed and estimated shares similar to those computed in the Prediction Success Table due to McFadden (1979) for the discrete choice models with disaggregate data (Gaudry and Liem, 1987). But this approach when applied to the share models with aggregate data also presents some difficulties:

1. The dependent variable S_{it} which is continuous over the interval [0,1] should be transformed into a discrete choice variable d_{it} which is a dummy variable having only two values 0 or 1, in order to compute the expected number of observations which "chose" alternative i but have been predicted to "choose" alternative j :

$$T_{ij} = \sum_t d_{it} \hat{P}_{jt} \ , \ (i,j = 1,\ldots,M) \tag{128}$$

where $d_{it} = 1$ if $S_{it} = \max_m (S_{mt}), \ m \in C_t$, and $d_{it} = 0$ otherwise.

2. Recalling that in a **Prediction Success Table**, the different T_{ij}'s are presented in a matrix where the rows correspond to the actual (or observed) alternatives and the columns to the predicted (or estimated) alternatives, **the sum of T_{ij}'s in each row i represents the *observed count***, i.e. the total number of observations which actually "select" alternative i, and similarly **the sum of T_{ij}'s in each column j denotes the *predicted count***, i.e. the total number of observations predicted to "choose" alternative j. Then from these *observed* and *predicted counts*, global measures for the *observed* and *predicted shares* can be computed as:

$$Observed\,Share_i = \frac{1}{T}\,Observed\,Count_i = \frac{1}{T}\sum_j T_{ij} \tag{129}$$

$$Predicted\,Share_j = \frac{1}{T}\,PredictedCount_j = \frac{1}{T}\sum_i T_{ij}. \tag{130}$$

Clearly, **the sum of all observed shares over i adds up to unity** and similarly for the sum of all predicted shares over j. But due to the transformation of the share variable S_{it} into a dummy variable d_{ij}, it happens frequently that for every observation t in a sample, there are certain alternatives whose observed shares S_{it}'s are always smaller than those associated with the other alternatives so that the associated dummy variables d_{it}'s are always zero, and hence **all the T_{ij}'s in the rows corresponding to the alternatives with smaller shares are zero**. Consequently, the global measures for the observed count and observed share computed from the T_{ij}'s for these rows are also null. In this sense, the approach cannot give any information about how small these smaller shares are.

ii) <u>Rho-squared</u>: an "intuitive" goodness-of-fit statistic computed for:

• each alternative: $R_i^2 = 1 - \frac{\sum_t (S_{it} - \hat{P}_{it})^2}{\sum_t (S_{it} - \bar{S}_i)^2}, (i = 1,\ldots,M).$

• the whole model: $R^2 = 1 - \frac{\sum_i \sum_t (S_{it} - \hat{P}_{it})^2}{\sum_i \sum_t (S_{it} - \bar{S}_i)^2}.$

Note that these statistics can be negative since $\hat{P}_{it}$ is used to compute the residuals $S_{it} - \hat{P}_{it}$ instead of the expected value of S_{it} which is more relevant for a nonlinear model, but difficult to obtain due to the numerical approximations of the multiple integrals involving the $(M-1)$-variate normal density function for each observation t.

5.3 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_{it}'s.
5. The log-likelihood function L.
6. The elasticities.

In other words, 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), three 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_{itk} which can be:

   A. A linear variable $(-\infty < X_{itk} < +\infty)$ without a Box-Cox transformation.
   B. A positive variable $(X_{itk} > 0)$ with a Box-Cox transformation.
   C. A quasi-dummy $(X_{itk} \geq 0)$ 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_{itk}, 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_{itk} 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_{itk} 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.

3. Invariance with respect to a reference alternative R, for two data cases:

   A. Full choice set: all the alternatives are available for every observation in the sample.
   B. Non full choice set: some or all the alternatives are available for an observation in the sample.

   For each case above, three forms of the covariance matrix of the residuals across the equations, Σ, can be considered:

   a. Full Σ.
   b. Diagonal Σ.
   c. Scalar Σ.

   Note that in case B (non full choice set), the model is not identified if Σ is full or diagonal, since one cannot compute the different variances and covariances ($\sigma_{ij} \neq 0$ if Σ is full) of the residuals using the same number of observations due to the different availabilities of the alternatives across the observations.

We will consider in detail only the case A (full choice set) with a full matrix Σ jointly with each of the three specifications of the constants:

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

since we already know that the invariance holds for a LOGIT model with $M-1$ specific constants and without Box-Cox transformations (Wills, 1982) only if Σ is full. The same is true for S-DOGIT and G-DOGIT without Box-Cox. Moreover, the invariance still holds with Box-Cox for the three families because the property depends strictly on the fullness of the covariance matrix Σ, but not on the functional form of the independent variables. All the other cases (3.A.b, 3.A.c, 3.B.c) are not invariant, since Σ is diagonal or scalar, for any of the five families.

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_{jt}'s have the following forms:

$$\begin{bmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{bmatrix} = \begin{bmatrix} C_{1t} & 0 & 0 & X_{1t} & 0 & 0 \\ 0 & C_{2t} & 0 & 0 & X_{2t} & 0 \\ 0 & 0 & C_{3t} & 0 & 0 & X_{3t} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_6 \end{bmatrix} \tag{131}$$

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

LOGIT

The choice probability of alternative 1 can be written as:

$$P_{1t} = \frac{\exp(V_{1t})}{\sum_{m=1}^{3} \exp(V_{mt})} = \frac{\exp(\beta_1 C_{1t} + \beta_4 X_{1t})}{\exp(\beta_1 C_{1t} + \beta_4 X_{1t}) + \exp(\beta_2 C_{2t} + \beta_5 X_{2t}) + \exp(\beta_3 C_{3t} + \beta_6 X_{3t})} \quad . \tag{132}$$

Dropping the C_{mt}'s in the V_{mt}'s since they are all equal to 1 for every m and t, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1t} = \frac{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t})}{\exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t}) + \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2t}) + \exp(\beta_6 X_{3t})} \quad . \tag{133}$$

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:

$$P_{1t} = \frac{\exp(\beta_4 X_{1t})}{\exp(\beta_4 X_{1t}) + \exp(\beta_5 X_{2t}) + \exp(\beta_6 X_{3t})} \quad . \tag{134}$$

S-DOGIT

The choice probability of alternative 1 can be written as:

$$P_{1t} = \frac{\exp(V_{1t}) + \theta_1 \sum_j \exp(V_{jt})}{\left(1 + \sum_m \theta_m\right) \sum_j \exp(V_{jt})} \quad . \tag{135}$$

Dropping the C_{jt}'s in the V_{jt}'s since they are all equal to 1 for every j and t, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1t} = \frac{\exp(V_{1t} - \beta_3) + \theta_1 \sum_j \exp(V_{jt} - \beta_3)}{\left(1 + \sum_m \theta_m\right) \sum_j \exp(V_{jt} - \beta_3)} \tag{136}$$

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

$$\begin{aligned} \exp(V_{1t} - \beta_3) &= \exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t}) \\ \exp(V_{2t} - \beta_3) &= \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2t}) \\ \text{and } \exp(V_{3t} - \beta_3) &= \exp(\beta_6 X_{3t}) \quad . \end{aligned} \tag{137}$$

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:

$$P_{1t} = \frac{\exp(\beta_4 X_{1t}) + \theta_1[\exp(\beta_4 X_{1t}) + \exp(\beta_5 X_{2t}) + \exp(\beta_6 X_{3t})]}{\left(1 + \sum_m \theta_m\right)[\exp(\beta_4 X_{1t}) + \exp(\beta_5 X_{2t}) + \exp(\beta_6 X_{3t})]} \quad . \tag{138}$$

G-DOGIT

The choice probability of alternative 1 can be written as:

$$P_{1t} = \frac{\exp(V_{1t}) + \sum_{j \neq 1} \theta_{1j} \exp(V_{jt})}{\sum_m \exp(V_{mt}) + \sum_m \sum_{j \neq m} \theta_{mj} \exp(V_{jt})} \quad . \tag{139}$$

Dropping the C_{jt}'s in the V_{jt}'s since they are all equal to 1 for every j and t, and dividing the numerator and the denominator by $\exp(\beta_3)$ for example, we get:

$$P_{1t} = \frac{\exp(V_{1t} - \beta_3) + \sum_{j \neq 1} \theta_{1j} \exp(V_{jt} - \beta_3)}{\sum_m \exp(V_{mt} - \beta_3) + \sum_m \sum_{j \neq m} \theta_{mj} \exp(V_{jt} - \beta_3)} \tag{140}$$

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

$$\begin{aligned} \exp(V_{1t} - \beta_3) &= \exp(\beta_1 - \beta_3)\exp(\beta_4 X_{1t}) \\ \exp(V_{2t} - \beta_3) &= \exp(\beta_2 - \beta_3)\exp(\beta_5 X_{2t}) \\ \text{and } \exp(V_{3t} - \beta_3) &= \exp(\beta_6 X_{3t}) \quad . \end{aligned} \tag{141}$$

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:

$$P_{1t} = \frac{\exp(\beta_4 X_{1t}) + \theta_{12}\exp(\beta_5 X_{2t}) + \theta_{13}\exp(\beta_6 X_{3t})}{(1+\theta_{21}+\theta_{31})\exp(\beta_4 X_{1t}) + (1+\theta_{12}+\theta_{32})\exp(\beta_5 X_{2t}) + (1+\theta_{13}+\theta_{23})\exp(\beta_6 X_{3t})} \quad . \tag{142}$$

LIN-IPT-L

The choice probability of alternative 1 can be written as:

$$P_{1t} = \frac{[\phi_1 \exp(V_{1t}) + 1]^{\frac{1}{\phi_1}} - \mu_1}{\sum_m \left\{ [\phi_m \exp(V_{mt}) + 1]^{\frac{1}{\phi_m}} - \mu_m \right\}} \quad . \tag{143}$$

Dropping the C_{mt}'s in the V_{mt}'s since they are all equal to 1 for every m and t, we get:

$$P_{1t} = \frac{[\phi_1 \exp(\beta_1 + \beta_4 X_{1t}) + 1]^{\frac{1}{\phi_1}} - \mu_1}{\sum_m \left\{ [\phi_m \exp(\beta_m + \beta_{m+3} X_{mt}) + 1]^{\frac{1}{\phi_m}} - \mu_m \right\}} \quad . \tag{144}$$

i. ***M* specific constants**: due to the nonlinear structure of the expression $[\phi_m \exp(\beta_m + \beta_{m+3} X_{mt}) + 1]^{\frac{1}{\phi_m}}$, $m = 1, \ldots, 3$, 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:

$$P_{1t} = \frac{\exp\left\{\frac{[\exp(V_{1t})+\tilde{\mu}_1]^{\tilde{\phi}_1}-1}{\tilde{\phi}_1}\right\}}{\sum_m \exp\left\{\frac{[\exp(V_{mt})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}} \quad . \tag{145}$$

Dropping the C_{mt}'s in the V_{mt}'s since they are all equal to 1 for every m and t, we get:

$$P_{1t} = \frac{\exp\left\{\frac{[\exp(\beta_1+\beta_4 X_{1t})+\tilde{\mu}_1]^{\tilde{\phi}_1}-1}{\tilde{\phi}_1}\right\}}{\sum_m \exp\left\{\frac{[\exp(\beta_m+\beta_{m+3} X_{mt})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}} \quad . \tag{146}$$

i. ***M* specific constants**: due to the nonlinear structure of the expression $\exp\left\{\frac{[\exp(\beta_m+\beta_{m+3} X_{mt})+\tilde{\mu}_m]^{\tilde{\phi}_m}-1}{\tilde{\phi}_m}\right\}$, $m = 1,\ldots,3$, 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 7, where the abbreviations are defined as follows:

- **(Model not identified$^{(1)}$)**: the model is not identified since in the cases (i) and (iii), the constants are not identified for LOGIT, S-DOGIT and G-DOGIT.
- **(Model not identified$^{(2)}$)**: the model is not identified since the covariance matrix Σ (full or diagonal) cannot computed due to the different availabilities of the alternatives across the observations.
- **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.
- **NO*** : applies to the case (ii) only.
- **NO**** : applies to the three cases (i), (ii) and (iii).

TABLE 7 Partial Invariance Types in Share 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_{itk}:			
P2.A	**Linear Variable** $(-\infty < X_{itk} < +\infty)$ i) M specific constants ii) M-1 specific constants iii) M equal constants	 (Model not identified[(1)]) YES (Model not identified[(1)])	 UI UI NO	 UI UI NO
P2.B	**Positive Variable** $(X_{itk} > 0)$ with Box-Cox i) M specific constants ii) M-1 specific constants iii) M equal constants	 (Model not identified[(1)]) YES (Model not identified[(1)])	 CI NO NO	 CI NO NO
P2.C	**Quasi-Dummy** $(X_{itk} \geq 0)$ with Box-Cox i) M specific constants ii) M-1 specific constants iii) M equal constants	 (Model not identified[(1)]) YES (Model not identified[(1)])	 CI NO NO	 CI NO NO
P3	with respect to the reference alternative R			
P3.A	**Full choice set** a) Full covariance matrix Σ i) M specific constants ii) M-1 specific constants iii) M equal constants b) Diagonal Σ c) Scalar Σ	 (Model not identified[(1)]) YES (Model not identified[(1)]) NO* NO*	 CI CI CI NO** NO**	 CI CI CI NO** NO**
P3.B	**Non full choice set** a) Full covariance matrix Σ b) Diagonal Σ c) Scalar Σ	 (Model not identified[(2)]) (Model not identified[(2)]) NO*	 (Model not identified[(2)]) (Model not identified[(2)]) NO**	 (Model not identified[(2)]) (Model not identified[(2)]) 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 indicated in Table 8 where we note:

1. Type P1 can be combined with Type P2.A, Type P2.B or Type P2.C, and Type P3.A.a 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, Type P2.C.ii and Type P3.A.a.ii to give the most general specification where all the three groups of variables (linear, positive and quasi-dummy) are present for the full choice set case with a full covariance matrix Σ. 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, Type P2.C.ii and Type P3.A.a.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, Type P2.C.i and Type P3.A.a.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, Type P2.C.iii and Type P3.A.a.iii implies that the invariance will not hold globally.

TABLE 8 Global Invariance Types in Share Models with Full Choice Set and Full Σ only

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)+(P3.A.a.i)	(Model not identified[(1)])	CI	CI
GII	ii) ***M-1* specific constants** (P1)+(P2.A.ii)+(P2.B.ii)+(P2.C.ii)+(P3.A.a.ii)	YES	NO	NO
GIII	iii) ***M* equal constants** (P2.A.iii)+(P2.B.iii)+(P2.C.iii)+(P3.A.a.iii)	(Model not identified[(1)])	NO	NO

Note that global invariance will be destroyed for the five families:

1. With full choice set and diagonal or scalar Σ.
2. With non full choice set and scalar Σ.

6 FORECASTING

6.1 Simulation Forecast

Once the parameter vector Π has been estimated for a variant associated with a model type of a family, say $\hat{\Pi}$ the estimated parameter vector, the simulated shares $\hat{P}_{is}$'s $(i = 1, \ldots, M)$ can be computed from a given set of values of all the independent variables $X_s = [X_{1s}, X_{2s}, \ldots, X_{Ms}]$ where a typical element X_{ms} is a set of independent variables included in the alternative m for an observation s to be simulated:

$$\hat{P}_{is} = \frac{\mathcal{U}_{is}\left(\hat{\Pi}\right)}{\sum_m \mathcal{U}_{ms}\left(\hat{\Pi}\right)} \tag{147}$$

where $\mathcal{U}_{is}\left(\hat{\Pi}\right)$ is the utility function evaluated at $\Pi = \hat{\Pi}$ and already defined in Table 1 for each family.

For example, the simulated shares for **LOGIT** (S-BC) are computed at $\hat{\Pi} = \left[\hat{\beta}', \hat{\lambda}'\right]$ as follows:

$$\hat{P}_{is} = \frac{\exp\left(\hat{V}_{is}\right)}{\sum_m \exp\left(\hat{V}_{ms}\right)} = \frac{\exp\left[\sum_k \hat{\beta}_k X_{isk}^{(\hat{\lambda}_k)}\right]}{\sum_m \exp\left[\sum_k \hat{\beta}_k X_{msk}^{(\hat{\lambda}_k)}\right]}. \tag{148}$$

6.2 Maximum Likelihood Forecast

For CASE A, the simulated shares $\hat{P}_{is}$'s can be shown to be identical to the maximum likelihood forecasts $\hat{S}_{is}$'s which are obtained by maximizing the log-likelihood function associated with the observation s, $\mathcal{L}_s\left(S_{is}|\hat{\Pi}, \hat{\Sigma}\right)$, with respect to the unknown S_{is}'s $(i \neq M)$, given the estimated parameters $\hat{\Pi}$ and the estimated matrix of covariance of residuals $\hat{\Sigma}$:

$$\mathcal{L}_s\left(S_{is}|\hat{\Pi},\hat{\Sigma}\right) = -\frac{1}{2}(M-1)\ln(2\pi) - \frac{1}{2}\ln\left(\det\hat{\Sigma}\right) - \frac{1}{2}\hat{U}_s{}'\hat{\Sigma}^{-1}\hat{U}_s \tag{149}$$

where a typical element of $\hat{U}_s$ is $\hat{u}_{is} = \ln\left(\frac{S_{is}}{1-\sum_{m\neq M} S_{ms}}\right) - \ln\left(\frac{\hat{P}_{is}}{1-\sum_{m\neq M}\hat{P}_{ms}}\right)$, where S_{Ms} and $\hat{P}_{Ms}$ are replaced by $1 - \sum_{m\neq M} S_{ms}$ and $1 - \sum_{m\neq M}\hat{P}_{ms}$ respectively, due to the adding-up share constraint $\sum_{m=1}^{M} S_{ms} = 1$.

Setting the first derivatives of $\mathcal{L}_s$ with respect to the S_{is}'s equal to zero yields a system of $M-1$ equations with the unknown S_{is}'s :

$$\begin{aligned}\frac{\partial\mathcal{L}_s}{\partial S_{is}} &= -\left(\frac{\partial\hat{U}_s}{\partial S_{is}}\right)'\hat{\Sigma}^{-1}\hat{U}_s = -\sum_{j=1}^{M-1}\sum_{h=1}^{M-1}\frac{\partial\hat{u}_{js}}{\partial S_{is}}\hat{\sigma}^{jh}\hat{u}_{hs} \\ &= -\sum_{j=1}^{M-1}\sum_{h=1}^{M-1}\frac{\partial\hat{u}_{js}}{\partial S_{is}}\hat{\sigma}^{jh}\left[\ln\left(\frac{S_{hs}}{1-\sum_{m\neq M}S_{ms}}\right) - \ln\left(\frac{\hat{P}_{hs}}{1-\sum_{m\neq M}\hat{P}_{ms}}\right)\right] = 0\end{aligned} \tag{150}$$

where

$$\frac{\partial \hat{u}_{js}}{\partial S_{is}} = \begin{cases} \dfrac{1}{S_{is}} + \dfrac{1}{1 - \sum\limits_{m \neq M} S_{ms}} & , j = i \\ \dfrac{1}{1 - \sum\limits_{m \neq M} S_{ms}} & , j \neq i \end{cases} \tag{151}$$

and $\hat{\sigma}^{jh}$ is a typical element of the matrix $\hat{\Sigma}^{-1}$.

Clearly, this system has a solution if every S_{hs} is equal to the corresponding $\hat{P}_{hs}$, and the solution, $\hat{S}_{hs} = \hat{P}_{hs}$, is unique.

A similar result can be obtained for CASE B. The log-likelihood function associated with the observation s, involving the unknown S_{is}'s $(i \neq R_s)$ and given the estimated parameters $\widehat{\Pi}$ and the estimated common variance $\hat{\sigma}_u^2$ can be written as:

$$\mathcal{L}_s\left(S_{is}|\widehat{\Pi}, \hat{\sigma}_{\mathrm{u}}^2\right) = -\frac{1}{2}(M_s - 1)\ln(2\pi) - \frac{1}{2}(M_s - 1)\ln\hat{\sigma}_u^2 - \frac{1}{2\hat{\sigma}_u^2}\sum_{j \neq R_s} \hat{u}_{js}^2 \tag{152}$$

where R_s is the reference alternative in the set C_s of M_s available alternatives and $\hat{u}_{js} = \ln\left(\frac{S_{js}}{1 - \sum\limits_{m \neq R_s} S_{ms}}\right) - \ln\left(\frac{\widehat{P}_{js}}{1 - \sum\limits_{m \neq R_s} \widehat{P}_{ms}}\right)$.

Setting the first derivatives of $\mathcal{L}_s$ with respect to the S_{is}'s equal to zero yields a system of $M_s - 1$ equations with the unknown S_{is}'s:

$$\begin{aligned} \frac{\partial L_s}{\partial S_{is}} &= -\frac{1}{\hat{\sigma}_u^2}\sum_{j \neq R_s}\frac{\partial \hat{u}_{js}}{\partial S_{is}}\hat{u}_{js} \\ &= -\frac{1}{\hat{\sigma}_u^2}\sum_{j \neq R_s}\frac{\partial \hat{u}_{js}}{\partial S_{is}}\left[\ln\left(\frac{S_{js}}{1 - \sum\limits_{m \neq R_s} S_{ms}}\right) - \ln\left(\frac{\widehat{P}_{js}}{1 - \sum\limits_{m \neq R_s} \widehat{P}_{ms}}\right)\right] = 0 \end{aligned} \tag{153}$$

where

$$\frac{\partial \hat{u}_{js}}{\partial S_{is}} = \begin{cases} \frac{1}{S_{is}} + \frac{1}{1 - \sum\limits_{m \neq R_s} S_{ms}} & , j = i \\ \frac{1}{1 - \sum\limits_{m \neq R_s} S_{ms}} & , j \neq i \end{cases} . \tag{154}$$

Clearly, this system has a unique solution which corresponds to $\widehat{S}_{js} = \widehat{P}_{js}$ for every $j \in C_s$.

6.3 Standard Error of Forecast

Since for a general nonlinear equation system there is no analytical formula that can be used to compute directly the standard errors of forecast, and hence the corresponding confidence intervals for the S_{is}'s , approximate confidence intervals for the P_{is}'s can be derived using a first order Taylor series expansion of $\hat{P}_{is}$ about the true parameter vector Π:

$$\hat{P}_{is} = P_{is} + \left(\hat{\Pi} - \Pi\right)' \frac{\partial P_{is}}{\partial \Pi} \tag{155}$$

Since $\hat{\Pi}$ is asymptotically normal with mean Π and covariance matrix $\text{Var}\left(\hat{\Pi}\right)$ as given in (18) and (20) for CASES A and B respectively, $\hat{P}_{is}$ is also asymptotically normal with mean P_{is} and variance:

$$\text{Var } \hat{P}_{is} = \frac{\partial P_{is}}{\partial \Pi'} \text{Var}\left(\hat{\Pi}\right) \frac{\partial P_{is}}{\partial \Pi} \tag{156}$$

The estimated value of Var $\hat{P}_{is}$, say Var $\widehat{\hat{P}_{is}}$, is obtained by substituting $\hat{\Pi}$ for Π in the derivative $\partial P_{is}/\partial \Pi$ which has the following form:

$$\frac{\partial P_{is}}{\partial \Pi} = \frac{1}{\sum\limits_{m} U_{ms}(\Pi)} \left[\frac{\partial U_{is}}{\partial \Pi} - P_{is} \sum_{m} \frac{\partial U_{ms}(\Pi)}{\partial \Pi} \right] \tag{157}$$

where P_{is} is equal to $U_{is}(\Pi)/\sum\limits_{m} U_{ms}(\Pi)$ and each derivative $\partial U_{is}/\partial \Pi$ is given in the table below for each family.

An asymptotic $100(1-\epsilon)$ percent confidence interval for each P_{is} can be computed using the estimated standard error of forecast $\hat{\sigma}\left(\hat{P}_{is}\right)$ which is equal to the square root of Var $\widehat{\hat{P}_{is}}$:

$$\hat{P}_{is} - Z_{\epsilon/2}\hat{\sigma}\left(\hat{P}_{is}\right) \leq P_{is} \leq \hat{P}_{is} + Z_{\varepsilon/2}\hat{\sigma}\left(\hat{P}_{is}\right) \tag{158}$$

where $Z_{\epsilon/2}$ is the $(1-\varepsilon/2)$ percentile of the standard normal distribution.

In the program, the value of ε is chosen to be 0.05, hence $Z_{\epsilon/2}$ is equal to 1.96, to yield a 95 percent confidence interval for P_{is}.

TABLE 9 Derivatives of $U_{is}(\Pi)$ with respect to Π for the five families.

Utility $U_{is}(\Pi)$	Π	$\partial U_{is}(\Pi)/\partial\Pi$
LOGIT		
$\exp(V_{is})$	β_k	$\exp(V_{is})\dfrac{\partial V_{is}}{\partial \beta_k}$, where $\dfrac{\partial V_{is}}{\partial \beta_k} = X_{isk}^{(\lambda_k)}$
	λ_k	$\exp(V_{is})\dfrac{\partial V_{is}}{\partial \lambda_k}$, where $\dfrac{\partial V_{is}}{\partial \lambda_k} = \sum_{k'} \beta_{k'} \dfrac{\partial X_{isk'}^{(\lambda_{k'})}}{\partial \lambda_k}$
S-DOGIT		
$\exp(V_{is}) + \theta_i \sum_j \exp(V_{is})$	β_k	$\exp(V_{is})\dfrac{\partial V_{is}}{\partial \beta_k} + \theta_i \sum_j \exp(V_{js})\dfrac{\partial V_{js}}{\partial \beta_k}$
	θ_i	$\sum_j \exp(V_{js})$
	λ_k	$\exp(V_{is})\dfrac{\partial V_{is}}{\partial \lambda_k} + \theta_i \sum_j \exp(V_{js})\dfrac{\partial V_{js}}{\partial \lambda_k}$
G-DOGIT		
$\exp(V_{is}) + \theta_{ij} \sum_{j \neq i} \exp(V_{js})$	β_k	$\exp(V_{is})\dfrac{\partial V_{is}}{\partial \beta_k} + \sum_{j \neq i} \theta_{ij} \exp(V_{js})\dfrac{\partial V_{js}}{\partial \beta_k}$
	θ_{ij}	$\exp(V_{js})$

TABLE 9 (Continued) Derivatives of $U_{is}(\Pi)$ with respect to Π for the five families.

LIN-IPT-L		
$[\phi_i \exp(V_{is}) + 1]^{\frac{1}{\phi_i}} - \mu_i$	β_k	$A_{is}^{\frac{1}{\phi_i}-1} \exp(V_{is}) \frac{\partial V_{is}}{\partial \beta_k}$, where $A_{is} = \phi_i \exp(V_{is}) + 1$
	$\phi_i > 0$	$\frac{1}{\phi_i^2} A_{is}^{\frac{1}{\phi_i}-1} [\phi_i \exp(V_{is}) - A_{is} \ln(A_{is})]$
	$\phi_i = 0$	$-\frac{1}{2} \exp[\exp(V_{is}) + 2V_{is}]$
	μ_i	-1
	λ_k	$A_{is}^{\frac{1}{\phi_i}-1} \exp(V_{is}) \frac{\partial V_{is}}{\partial \lambda_k}$
BT-IPT-L		
$\exp\left\{\frac{[\exp(V_{is}) + \tilde{\mu}_i]^{\tilde{\phi}_i} - 1}{\tilde{\phi}_i}\right\}$	β_k	$\exp\left\{B_{is}^{(\tilde{\phi}_i)}\right\} B_{is}^{\tilde{\phi}_i - 1} \exp(V_{is}) \frac{\partial V_{is}}{\partial \beta_k}$, where $B_{is} = \exp(V_{is}) + \tilde{\mu}_i$
	$\tilde{\phi}_i \neq 0$	$\exp\left\{B_{is}^{(\tilde{\phi}_i)}\right\} \frac{1}{\tilde{\phi}_i} \left[B_{is}^{\tilde{\phi}_i} \ln(B_{is}) - B_{is}^{(\tilde{\phi}_i)}\right]$
	$\phi_i = 0$	$B_{is} \left(\frac{1}{2} \ln^2 B_{is}\right)$
	$\tilde{\mu}_i$	$\exp\left\{B_{is}^{(\tilde{\phi}_i)}\right\} B_{is}^{\tilde{\phi}_i - 1}$
	λ_k	$\exp\left\{B_{is}^{(\tilde{\phi}_i)}\right\} B_{is}^{\tilde{\phi}_i - 1} \exp(V_{is}) \frac{\partial V_{is}}{\partial \lambda_k}$

7 BIBLIOGRAPHY FOR THE SHARE CLASS

1. **Anscombe, F.J. and J.W. Tukey** (1954), “The Criticism of Transformations”, Paper presented before the American Statistical Association and Biometric Society, Montreal.
2. **Barten, A.P.** (1969), “Maximum Likelihood Estimation of a Complete System of Demand Equations”, *European Economic Review*, 1, 7–73.
3. **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.
4. **Belsley, D., E. Kuh and R.E. Welsh** (1980), *Regression Diagnostics*, John Wiley, New York.
5. **Ben-Akiva, M.** (1977), “Choice Models with Simple Choice Set Generating Processes”, Working paper, Department of Civil Engineering, MIT, Cambridge, Mass.
6. **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.
7. **Berndt, E.R. and L.R. Christensen** (1973), “The Translog Function and the Substitution of Equipment, Structures and Labor in U.S. Manufacturing 1929–1968”, *Journal of Econometrics*, 1, 81–114.
8. **Berndt, E.R. and L.R. Christensen** (1974), “Testing for the Existence of a Consistent Aggregated Index of Labor Inputs”, *American Economic Review*, 64, 391–404.
9. **Berndt, E.R. and N.E. Savin** (1975), “Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances”, *Econometrica*, 43, 937–957.
10. **Berndt, E.R. and D.O. Wood** (1975), “Technology, Prices and the Derived Demand for Energy”, *Review of Economics and Statistics*, 57, 259–268.
11. **Binswanger, H.P.** (1974), “The Measurement of Technical Change Biases with Many Factors of Production”, *American Economic Review*, 64, 964–976.
12. **Bordley, R.F.** (1989), “Relaxing the Loyalty Condition in the Colombo/Morrison Model”, *Marketing Science*, 8(1), 100-103.
13. **Bordley, R.F.** (1990), “The Dogit Model is Applicable even Without Perfectly Captive Buyers”, *Transportation Research*, 24B, 4, 315–323.
14. **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.
15. **Box, G.P. and D.R. Cox** (1964), “An Analysis of Transformations”, *Journal of the Royal Statistical Society*, Series B, 26, 211–243.
16. **Box, G.P. and P.W. Tidwell** (1962), “Transformations of the Independent Variables”, *Technometrics*, 4, 531–550.
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. **Chavas, J.P. and K. Segerson** (1986), “Singularity and Autoregressive Disturbances Linear Logit Models”, *Journal of Business and Economic Statistics*, 4(2), 161–169.
19. **Christensen, L.R., D.W. Jorgenson and L.J. Lau** (1975), “Transcendental Logarithmic Utility Functions”, *American Economic Review*, 65, 367–383.
20. **Clarke, R.R.** (1957), “Constant-Ratio Rule for Confusion Matrices in Speech Communication”, *Journal of Acoustical Society of America*, 30, 715–720.

21. **Colombo, R. and D. Morrison** (1989), "A Brand-switching Model with Implications for Marketing Strategies", *Marketing Science*, 8(1), 89–99.
22. **Considine, T.J. and T.D. Mount** (1984), "The Use of Linear Logit Models for Dynamic Input Demand Systems", *Review of Economics and Statistics*, 66, 434–443.
23. **Coombs, C.H.** (1958), "On the Use of Inconsistency of Preferences in Psychological Measurements", *Journal of Experimental Psychology*, 55, 1–7.
24. **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.
25. **Fletcher, R. and M.J.D. Powell** (1963), "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, 6, 163–168.
26. **Gaudry, M.J.I.** (1978), "A Generalized Logit mode choice model", Publication #98, Centre de Recherche sur les Transports, Université de Montréal.
27. **Gaudry, M.J.I.** (1981), "The Inverse Power Transformation Logit and Dogit Mode Choice Models", *Transportation Research*, Part B, 15B, 97–103.
28. **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.
29. **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.
30. **Gaudry, M.J.I. and M.G. Dagenais** (1979), "The Dogit Model", *Transportation Research*, Part B, 13B, 105–111.
31. **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.
32. **Griffin, J.M. and P.R. Gregory** (1976), "An Intercountry Translog Model of Energy Substitution Responses", *American Economic Review*, 66, 845–857.
33. **Hudson, E.A. and D.W. Jorgenson** (1974), "U.S. Energy Policy and Economic Growth", *Bell Journal of Economics*, 5, 461–514.
34. **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.
35. **Krishnan, K.** (1977), "Incorporating Thresholds of Indifference in Probabilistic Choice Models", *Management Science*, 23, 1224–33.
36. **Lau, L.J.** (1979), "Applications of Profit Functions", in M. Fuss and D. McFadden (eds.),*Production Economics: A Dual Approach to Theory and Applications*, North-Holland, Amsterdam, 133–216.
37. **Liem, T.C. and M.J.I. Gaudry** (1987), "PROBABILITY: The P-2 Program for the Standard and Generalized BOX-COX LOGIT Models with Disaggregate Data", Publication 527, Centre de Recherche sur les Transports, Université de Montréal.
38. **Lioukas, S.** (1984), "Thresholds and Transitivity in Stochastic Consumer Choice: A Multinomial Logit Analysis", *Management Science*, 30, 110–22.
39. **Luce, R.D.** (1959), *Individual Choice Behavior: A Theoretical Analysis*, John Wiley, New York.
40. **Manski, C.** (1977), "The Structure of Random Utility Models", *Theory and Decision*, 8, 229–254.
41. **Marschak, J.** (1960), "Binary-Choice Constraints and Random Utility Indicators", in K. Arrow, S. Karlin and P. Suppes (eds.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford University Press, Stanford, Calif.
42. **McFadden, D.** (1974a), "Conditional Logit Analysis of Qualitative Choice Behavior", in P.C.

Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 105–142.

43. **McFadden, D.** (1974b), “The Measurement of Urban Travel Demand”, *Journal of Public Economics*, 3, 303–328.
44. **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.
45. **McFadden, D.** (1978), “Modelling the Choice of Residential Location”, in A. Karlquist et al. (eds.), *Spatial Interaction Theory and Residential Location*, North Holland, Amsterdam, 75–96.
46. **McLynn, J.M.** (1973), “A Technical Note on a Class of Fully Competitive Modal Choice Models”, DIM Corporation, Bethesda, Md.
47. **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).
48. **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.
49. **Montmarquette, C. and S. Mahseredjian** (1985), “Functional Forms and Educational Production Functions”, *Economics Letters*, 19, 291–294.
50. **Parks, R.W.** (1971), “Maximum Likelihood Estimation of the Linear Expenditure System”, *Journal of the American Statistical Association*, 66, 900–903.
51. **Parks, R.W.** (1980), “On the Estimation of Multinomial Logit Models from Relative Frequency Data”, *Journal of Economics*, 13, 293–303.
52. **Pollack, R.A. and T.J. Wales** (1969), “Estimation of the Linear Expenditure System”, *Econometrica*, 37, 611–627.
53. **Ray, S.C.** (1982), “A Translog Cost Function Analysis of U.S. Agriculture”, *American Journal of Agricultural Economics*, 64, 490–498.
54. **Restle, F.** (1961), *Psychology of Judgement and Choice*, New York: John Wiley and Sons.
55. **Schmidt, P. and R.P. Strauss** (1975), “The Predictions of Occupation Using Multiple Logit Models”, *International Economic Review*, 16, 471–486.
56. **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.
57. **Stone, R.** (1954), “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand”, *Economic Journal*, 64, 511–527.
58. **Swait, J. and M. Ben-Akiva** (1985), “Constraints to Individual Travel Behavior in a Brazilian City”, *Transportation Research Record*, to be published.
59. **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.
60. **Theil, H.** (1969), “A Multinomial Extension of the Linear Logit Model”, *International Economic Review*, 10, 251–259.
61. **Tukey, J.W.** (1957), “On the Comparative Anatomy of Transformations”, *Annals of Mathematical Statistics*, 28, 602–632.
62. **Tyrrell, T. and T. Mount** (1982), “A Nonlinear Expenditure System Using a Linear Logit Specification”, *American Journal of Agricultural Economics*, 64, 539–546.
63. **Weaver, R.D.** (1983), “Multiple Input, Multiple Output Production Choices and Technology in the U.S. Wheat Region”, *American Journal of Agricultural Economics*, 65, 45–56.
64. **Wills, M.J.** (1982), “On the Equivalence of Common Rearrangements for Sum Constrained Travel Estimation Problems”,*Transportation Research*, Part B, 16B, 339–359.

65. **Woodland, A.D.** (1979), “Stochastic Specification and the Estimation of Share Equations”,*Journal of Econometrics*, 10, 361–383.
66. **Zarembka, P.C.** (1974), “Transformation of Variables in Econometrics”, in P.C. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 81–104.
67. **Zellner, A.** (1962), “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias”, *Journal of the American Statistical Association*, 57, 348–368.