EVALUATION OF MIXED LOGIT AS A PRACTICAL MODELLING ALTERNATIVE

Marcela A. Munizaga
Universidad de Chile, Casilla 228-3, Santiago, Chile
Ricardo Alvarez-Daziano
CIS Asociados Consultores en Transporte, Austria 2042, Santiago, Chile

1. INTRODUCTION

In the context of discrete choice modelling, the approach used the most is based on random utility theory (McFadden, 1974). According to this theory, each individual n has a utility function U_{in} associated to each of the alternatives i, choosing the one which maximises his (her) utility. This individual function can be divided into a systematic component V_{in}, which considers the effect of the explanatory variables (measurable or observable by the modeller attributes), and a random component e_{in} that takes into account all the effects not included in the systematic component of the utility function; for example, the incapacity of the modeller to observe all the variables that have an influence in the decision, measurement errors, differences between individuals, incorrect perceptions of attributes and the randomness inherent to human nature. Depending on the assumptions made for the distribution of the random error term, different models can be derived (Ortúzar and Willumsen, 1994).

The commonly used models are Multinomial Logit (McFadden, 1974), which is derived assuming that the error terms e_{in} are iid Gumbel, and the Nested Logit (Williams, 1977), that is derived as an extension of MNL, where it is considered the existence of an additional error component, which represents correlation in a group of alternatives. In synthesis, these models have very simple structures of covariance (of the error term), which is a simplifying assumption that not always is sustainable, but it allows to obtain models easy to understand and use.

The Probit model (Daganzo, 1979), on the other side, is derived assuming multivariate Normal distributed random errors, allowing in theory any error structure (covariance matrix) that the data permit to estimate, which imply a considerable level of estimation difficulty. This model, that appears so desirable from that point of view, has been timidly incorporated to practice, even though there are from some time ago powerful tools that yield its estimation by simulation (see Munizaga and Ortúzar, 1997).

It is in this context that Mixed Logit models are developed in the last years (also known as Error Component models or Logit Kernel Probit), as an intermediate alternative that is somewhere between Logit and Probit. The main idea of this kind of models is to consider more than one random component, keeping the basic model as a Logit, but allowing to represent correlation and/or heteroscedasticity. This leads to gain generality, but the estimation simplicity that characterises Logit models is lost, and as in the Probit case, simulation is required.

It has already been said that the distribution of the random disturbance plays a fundamental role in discrete choice modelling, and that the commonly used models suppose a homoscedastic and independent Gumbel distribution. So, if the point is to incorporate models that allow more general error structures, it is important to analyse which structures would be desirable to be able to estimate and why. We are talking about the possible existence of correlation and heteroscedasticity (different variance) in the error term. The case of correlation between alternatives (present for example when the user perceive some alternatives as more similar between them than others) is assimilated under certain restrictions by the Nested Logit model, yielding a block diagonal and homoscedastic covariance matrix (see Munizaga and Ortúzar, 1999 a; b). However, many cases of correlation and heteroscedasticity, easily associable to practical situations, can not be treated properly with the traditional models (Munizaga et al, 1997). So, it seems interesting to find a more general model which adapts to more sophisticated situations.

2. THE MIXED LOGIT MODEL

2.1 Formulation

The idea of Mixed Logit (ML) models is not new, models of these characteristics have been proposed several years ago. For example, we can quote the works of Cardell and Dunbar (1980), and Boyd and Melman (1980), where a model equivalent to the current ML is described with the name of Hedonic model. Its recent re-apparition with another name and renewed force can be due to technological advances in computing and numerical methods allowing now its estimation in less time. Recently this kind of models have been used to model diverse situations (Train, 1999; Brownstone and Train, 1999; Algers *et al.,* 1999). ML models assume a utility function U_{in} conformed by a deterministic component V_{in}, a random component ε_{in} independent and identically Gumbel distributed, and one or more additional random terms. These additional error terms can be grouped together in an additive term η_{in}, that can be function of the data (attributes of alternatives), and that potentially models the presence of correlation and heteroscedasticity. So, the utility function is defined as:

$$U_{in} = V_{in} + \eta_{in} + \varepsilon_{in} \qquad (1)$$

where ε_{in} ~ Gumbel(0, λ) and $\eta_{in} \sim f(\eta/\theta^*)$, where f is a general density function and θ^* are fixed parameters that describe it (*eg* mean and variance)[i]. As ε is iid Gumbel, then the probability conditional in η of individual n choosing alternative i corresponds to the Multinomial Logit model:

$$P_n(i/\eta) = L_{in}(\eta) = \frac{e^{V_{in}+\eta_{in}}}{\sum_j e^{V_{jn}+\eta_{jn}}} \qquad (2)$$

So, the probability of choosing the alternative corresponds to the integral of the conditional probability over all the possible values of η, which depends on the parameters characterising the distribution, this is:

$$P_{in} = \int L_{in}(\beta) f(\beta / \theta^*) d\beta \quad (3)$$

As a particular case, it can be assumed a utility function with the following specification[ii]:

$$U_{in} = \underbrace{\beta^t x_{in}}_{V_{in}} + \underbrace{\mu_{in}^t z_{in}}_{h_{in}} + \varepsilon_{in} \quad (4)$$

In this expression the assumption is that the deterministic component of the utility is linear in the β parameters that multiply the attributes x_{in}. Furthermore, it is assumed that β depends of certain parameters (μ_{in}) and data observed related to alternative i (z_{in}), relation which is also supposed lineal in the parameters. An additional assumption can be that the μ term depends only on the individual, with no variation over alternatives. The latter means:

$$\beta_{in} = \mu_n^t z_{in} \quad (5)$$

This specification is the one that has been used in the mayor part of the previous studies (Ben Akiva and Bolduc, 1996; Brownstone and Train, 1999).

2.2 Covariance Matrix

Given a utility function like (4) and considering also the usual assumption (5), let z_n be the matrix of dimension K x J_n that contains the vectors z_{in} (of dimension K) for each alternative belonging to the choice set of the individual (of dimension J_n) and ε_n a random vector iid Gumbel with covariance matrix Σ_ε containing each ε_{in}. If it is assumed that each term of μ_n has a density function with zero mean and σ_k^2 variance, and that the vector has a joint covariance matrix Ω, then the covariance matrix of the model (Σ), can be written as:

$$\Sigma = z_n^t \cdot \Omega \cdot z_n + \Sigma_\varepsilon = z_n^t \cdot \Omega \cdot z_n + \sigma_\varepsilon^2 I \quad (6)$$

It is clear that the matrix is positive definite and that its dimension is well defined[iii] and from this general expression it can be concluded that the model is capable to model correlation and heteroscedasticity between alternatives. In effect, the covariance between any two alternatives $i, j \in C_n$ with $i \neq j$ is given by the expression in equation (7).

$$\text{cov}(U_{in}, U_{jn}) = \sum_{k=1}^{K} z_{kin} z_{kjn} \sigma_k^2 \quad (7)$$

It can be seen that the covariance will be different from zero if for at least one k, $\sigma_k^2 > 0$ and $z_{kin}, z_{kjn} \neq 0$. In that case, the presence of correlation between alternatives i and j can be detected by the model. The variance of the utility of any alternative, is given by the equation (8).

$$\text{var}(U_{in}) = \sum_{k=1}^{K} z_{kin}^2 \sigma_k^2 + \frac{\pi}{6\lambda^2} \quad (8)$$

So, if $\text{var}(U_{in}) \neq \text{var}(U_{jn})$, then the model represents heteroscedasticity between those alternatives.

It is worth pointing out that this is a different form to justify a particular model form what we are used to. In the Probit model, for example, assumptions are made directly over the covariance matrix of the error term ε_{in}, imposing the presence of correlation or heteroscedasticity; in Nested Logit models, we biuild the nests structures according to the correlation structure we expect to found. While in a ML the error structure is build by adding random terms that are source of correlation and/or heteroscedasticity, and the covariance matrix is an output of this process.

2.3 Properties of the Mixed Logit

Probably the more interesting property of this model is that under certain regularity conditions, any random utility model has choice probabilities that can be approximated as close as wished by a Mixed Logit (McFadden and Train, 1997). As a matter of fact, a Mixed Logit model with Normal random distributed parameters can approximate a Probit model.

Furthermore the Mixed Logit model, allowing the presence of correlation and/or heteroscedasticity between alternatives, is capable to release the assumption of independence of irrelevant alternatives, characteristic of the Multinomial Logit model. In other words, the substitution patterns between alternatives are flexible. In effect, given a Mixed Logit probability (9), it can be shown that the ratio between probabilities of two alternatives depends on all the set of available alternatives.

$$P_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \frac{\exp(\beta' x_{in} + \mu_n' z_{in})}{\sum_{l=1}^{J} \exp(\beta' x_{ln} + \mu_n' z_{ln})} \right\} f(\mu_{1n}) \cdot \ldots \cdot f(\mu_{Kn}) d\mu_{1n} \ldots d\mu_{Kn} \qquad (9)$$

2.4 Estimation

The choice probability of a Mixed Logit model, like the presented in equation (3), does not have a mathematical closed expression as in the Multinomial or Nested Logit. Even more, the integral can not be solved analytically and simulation must be used. Nevertheless, the fact that the conditional probability (2) has a MNL form can be exploited. If R values of η are obtained from its density function $f(\eta/\theta^*)$, then for each of this repetitions it is possible to calculate

$$P_n(i/\eta^r) = L_{in}(\eta^r) = \frac{e^{V_{in}+\eta^r{}_{in}}}{\sum_j e^{V_{jn}+\eta^r{}_{jn}}}, \qquad (10)$$

with $r=1, \ldots, R$. Accordingly to this, it is possible to obtain an average probability

$$\tilde{P}(i) = \frac{1}{R}\sum_{r=1}^{R} L_{in}(\eta^r), \qquad (11)$$

and with it to build the simulated likelihood function

$$SL = \sum_{q} \sum_{i=1}^{J} y_{qi} \ln \tilde{P}(i) \tag{12}$$

Under regularity conditions, the simulated maximum likelihood estimator is consistent and asymptotically Normal. Even though (12) is an unbiased estimator of the probability, its natural logarithm results to be biased (Brownstone and Train, 1999); nevertheless, when the number of repetitions increases faster than the square root of the number of observations, the estimator is asymptotically equivalent to the maximum likelihood estimator (Hajivassilou and Ruud, 1994).

2. MIXED LOGIT COMPARED TO NESTED LOGIT

The NL model was conceived to deal with correlation between alternatives, grouping *similar* alternatives into *nests* within which the iid assumption does hold (Williams, 1977). The aggregation into nests implies a particular structure of the covariance matrix, because if two or more alternatives are grouped in a nest, a common error term (μ_j) is considered, and the corresponding off diagonal elements will be different from zero. The model is homoscedastic by construction, as $\sigma_m^2 + \sigma_e^2 = \sigma^2$ is imposed.

$$U_{in} = V_{in} + \mu_{jn} + \varepsilon_{in} \qquad i \in nest\ j \tag{13}$$

Brownstone and Train (1999) presented a Mixed Logit model that they call “analogue” to a Nested Logit. This particular model is build by grouping the alternatives into nests; then, for each nest, a dummy variable is included in the utility function, indicating if the alternative belongs to it or not. A common random parameter is associated to each one of these variables. In this way the model has a correlation structure such that in all alternatives belonging to the same nest an off diagonal term appears. We shall identify this particular specification under the name of Nested Mixed Logit (NML) based on its nesting error structure.

The authors conclude that in that way the pattern of correlation is equal to that of the Nested Logit. We consider that the correct thing to do is to compare the covariance matrix in both models. For example, let us suppose a case where three alternatives are available for a particular individual, and two of them are perceived as more similar between them. Under a NML the covariance matrix is:

$$\Sigma = \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_m^2 + \sigma_e^2 & \sigma_m^2 \\ 0 & \sigma_m^2 & \sigma_m^2 + \sigma_e^2 \end{bmatrix} \tag{14}$$

This matrix has off diagonal terms indicating correlation between two alternatives; however, it is heteroscedastic. So the NML is not really equivalent to the NL in terms of the error structure, because the latter is homoscedastic by definition. This situation can be overcome adding an additional error component in the non nested alternative, however it is difficult

to justify the presence of such a term, and it doesn't have a direct theoretical interpretation. Also, the problem has not been overcome in fact.

Indeed, let us suppose a case with four alternatives: car driver, car companion, bus and metro (underground). In practical terms it can be argued the presence of two nests: one defined by car driver and car companion and another one nesting bus and metro. This situation can be modelled by a Nested Logit:

$$\begin{aligned} U_{car} &= V_{car} + \mu_1 + \varepsilon_{car} \\ U_{car-comp} &= V_{car-comp} + \mu_1 + \varepsilon_{car-comp} \\ U_{bus} &= V_{bus} + \mu_2 + \varepsilon_{bus} \\ U_{metro} &= V_{metro} + \mu_2 + \varepsilon_{metro} \end{aligned} \qquad (15)$$

where $\varepsilon_{auto\text{-}ch}$, $\varepsilon_{auto\text{-}ac}$ are iid Gumbel(0, λ_1) with variance $\sigma^2_{\varepsilon 1}$, and ε_{bus}, ε_{metro} are iid Gumbel(0, λ_2) with variance $\sigma^2_{\varepsilon 2}$. On the other side, μ_1 and μ_2 distribute according to the assumptions of Nested Logit[iv], with variances $\sigma^2_{\mu 1}$ and $\sigma^2_{\mu 2}$, respectively. The scale parameters λ_1 and λ_2 must satisfy $\sigma^2_{\mu 1} + \sigma^2_{\varepsilon 1} = \sigma^2_{\mu 2} + \sigma^2_{\varepsilon 2} = \pi^2/6\beta$, that is the joint variance associated to an error term ξ_{in} iid Gumbel(0,β). That yields the following covariance structure:

$$\Sigma_{NL} = \begin{bmatrix} \sigma^2_{\mu 1} + \sigma^2_{\varepsilon 1} & \sigma^2_{\mu 1} & 0 & 0 \\ \sigma^2_{\mu 1} & \sigma^2_{\mu 1} + \sigma^2_{\varepsilon 1} & 0 & 0 \\ 0 & 0 & \sigma^2_{\mu 2} + \sigma^2_{\varepsilon 2} & \sigma^2_{\mu 2} \\ 0 & 0 & \sigma^2_{\mu 2} & \sigma^2_{\mu 2} + \sigma^2_{\varepsilon 2} \end{bmatrix} = \begin{bmatrix} \sigma^2_{\xi} & \sigma^2_{\mu 1} & 0 & 0 \\ \sigma^2_{\mu 1} & \sigma^2_{\xi} & 0 & 0 \\ 0 & 0 & \sigma^2_{\xi} & \sigma^2_{\mu 2} \\ 0 & 0 & \sigma^2_{\mu 2} & \sigma^2_{\xi} \end{bmatrix} \qquad (16)$$

If $\theta_1 = \beta / \lambda_1$ and $\theta_2 = \beta / \lambda_2$ are defined, then:

$$\Sigma_{NLJ} = \begin{bmatrix} \sigma^2_{\xi} & (1-\theta_1^2)\sigma^2_{\xi} & 0 & 0 \\ (1-\theta_1^2)\sigma^2_{\xi} & \sigma^2_{\xi} & 0 & 0 \\ 0 & 0 & \sigma^2_{\xi} & (1-\theta_2^2)\sigma^2_{\xi} \\ 0 & 0 & (1-\theta_2^2)\sigma^2_{\xi} & \sigma^2_{\xi} \end{bmatrix} \qquad (17)$$

It can be seen that the matrix is homoscedastic, that is possible to identify two nests and that the correlation within each nest does not have to be the same among different nests.

If we model the same situation with a NML, then

$$\begin{aligned} U_{car} &= V_{car} + \mu_1 + \varepsilon_{car} \\ U_{car-comp} &= V_{car-comp} + \mu_1 + \varepsilon_{car-comp} \\ U_{bus} &= V_{bus} + \mu_2 + \varepsilon_{bus} \\ U_{metro} &= V_{metro} + \mu_2 + \varepsilon_{metro} \end{aligned} , \qquad (18)$$

where $\mu_1 \sim N(0, \sigma_{\mu 1}^2)$, $\mu_2 \sim N(0, \sigma_{\mu 2}^2)$ and ε_{in} is iid Gumbel(0,λ) with variance σ_{ε}^2. The covariance matrix associated to this model is:

$$\Sigma_{ML} = \begin{bmatrix} \sigma_{m1}^2 + \sigma_e^2 & \sigma_{m1}^2 & 0 & 0 \\ \sigma_{m1}^2 & \sigma_{m1}^2 + \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_{m2}^2 + \sigma_e^2 & \sigma_{m2}^2 \\ 0 & 0 & \sigma_{m2}^2 & \sigma_{m2}^2 + \sigma_e^2 \end{bmatrix} \quad (19)$$

This matrix would be of a heteroscedastic nature, unless we assume $\sigma_{m1}^2 = \sigma_{m2}^2 = \sigma_m^2$. Nonetheless, under this additional supposition, the correlation within each nest must be the same. That is, we will consider an equivalent matrix with that one of NL, only if $\phi_1 = \phi_2$. Therefore, in this NML structure there is a clear trade off between correlation and heteroscedasticity that it is not observed in the NL. Again, it is possible to consider additional independent error terms, seeking a homoscedastic matrix. However, the structure obtained is even more complicated and less intuitive than the previous example. Also, there are identifiability restrictions that must be considered.

3. A NOTE ON IDENTIFIABILITY

The analysis of identifiability is usually restricted to setting the scale factor. This is enough for models with a simple covariance structure (Ben-Akiva and Lerman, 1985), however when more flexible structures are used the analysis is more complex. In order to determinate how many or which parameters can be estimated, it is essential to study the differentiated covariance matrix of any particular situation. This is due to the fact that utility maximisation depends on utility differences; so, when estimating the parameters, the model can distinguish only the different terms that appear in the differentiated covariance matrix. The identifiability issue has been well studied by Bunch (1991) and Bolduc (1992) and, for the particular case of Mixed Logit, by Ben-Akiva *et al* (2001).

If a model is specified over utility differences, with respect to a reference alternative, then there are less cells in the covariance matrix in contrast with the original one. This imposes the first restriction, known as **order condition.** Let us consider an universal choice set (C, with J alternatives), so the order condition limits the maximum number of identifiable elements to:

$$s^* = \frac{J(J-1)}{2} - 1 \quad (20)$$

Note that this number is equal to the order of the differentiated matrix, less one element which is fixed to set the scale of the model. These are the elements that the model can recognise when estimating the parameters. So it is important to analyse which of these the model can distinguish.

The visual analysis of the differenciated covariance matrix can show which elements are identifiable, but there is a formal **rank condition** (Bunch, 1991) which establishes the maximum number of parameters of the covariance matrix that can be estimated. This number corresponds to the rank of a matrix built as the jacobian of a column vector that contains all the different elements of the deviated covariance matrix.

As an example, in the last case presented in the previous section (Nested mixed logit with two nests), it can be easily shown that is not possible to identify s^2_{m1} and s^2_{m2} separately. This is because only the sum of the variances are distinguishable in the differentiated model, and it is true for any NML with two nests.

3. MIXED LOGIT COMPARED TO PROBIT

As mentioned before, the Mixed Logit model is build assuming additional error terms that may imply a heteroscedastic and correlated covariance matrix. On the other side, in the case of Probit only one error term, with a general covariance matrix, is assumed. In effect, a multinomial Probit model is derived assuming that given a utility function $U_{in} = V_{in} + \varepsilon_{in}$, the vector $\varepsilon_n = (\varepsilon_{1n}, \ldots, \varepsilon_{Jn})^t$ distributes multivariate Normal with general Σ covariance matrix.

The Probit model does not have a closed expression of the choice probability either, so it becomes necessary to use some kind of approximation or simulation. The more commonly used estimation method is the simulated maximum likelihood with the Geweke–Hajivassiliou-Keane simulator (Börsch-Supan and Hajivassilou, 1993), which recursively reduce the dimension of the integral up to an equivalent problem where repetitions of a truncated unidimentional normal are required. The simulated probabilities of this form are unbiased, continuous and differentiable.

The simulation for the Probit and Mixed Logit models have different dimension: *J-1* for the case of Probit[v] and *K* for case of Mixed Logit. In this way, if *K*<*J-1* then there is an advantage over Probit, because the simulation has a smaller dimension. This will happen when the number of random parameters incorporated to the ML is smaller than the number of alternatives minus one.

4. EMPIRICAL RESULTS

4.1 Simulation experiments

As a first step to evaluate the performance of the models mentioned before, a simulation experiment was carried out, following the methodology of Williams and Ortúzar (1982). This means to evaluate the models in a controlled case, where a database is generated with known choice rules, known taste parameters and known distribution of the random component(s). So, it is possible to evaluate if the parameters are properly recovered by the different models, and to make an evaluation of the prediction capabilities by simulating choices under modified conditions (different values of explanatory variables).

Different synthetic databases where elaborated according to the maximum utility principle, considering realistic means and deviations of the attributes (taken from a real database for Santiago de Chile). The choice situation emulated was mode choice within four alternatives available (car, bus, metro and taxi) and four explanatory variables (travel cost, travel time, access time, income dummy).

For the estimation of MNL and NL, as well as for Probit, a self-made code programmed in Gauss (Aptech Systems, 1994) was used. For the estimation of ML we used a flexible code programmed in Gauss by Kenneth Train, available from his web page[vi].

As a sample of the simulation experiments carried out, we show two cases with 8000 observations, generated considering the presence of correlation (correlation coefficient equal to 0.5 between alternatives bus and metro), where one of the cases also includes heteroscedasticity and the other one is homoscedastic. The total error variance for the simulation process was chosen so that the scale parameter of an iid Gumbel disturbance is equal to unity, and verifying that the experiment is not completely deterministic (choice mainly determined by the deterministic part of the utility function) nor completely random (choice mainly determined by the random component).

The results for the estimation process of the Multinomial Logit (MNL), Nested Logit (NL), Probit and Mixed Logit (ML) are shown in Table 1. The reference values reported (target) correspond to those values used to generate the database, and that we expect to recover by estimating models. The table shows the parameters estimates, the t statistic against zero and the t test against the reference value for the model which is fully consistent with the data generation process (ML).

Table 1: Simulation Results

Parameters' estimates, () t-value against zero, [ ] t-value against target

	Heteroscedastic Database					Homoscedastic Database				
	Target	MNL	NL	Probit	ML	Target	MNL	NL	Probit	ML
Car	-0.40	-0.3402 (-4.791)	-0.3872 (-5.469)	-0.2434 (-4.743)	-0.2946 (-5.168) [1.848]	-0.40	-0.1489 (-2.143)	-0.2843 (-4.269)	-0.1921 (-3.2748)	-0.4619 (-7.982) [-1.070]
Metro	0.20	0.3636 (5.848)	0.3171 (5.194)	0.2267 (5.322)	0.2512 (5.263) [1.072]	0.20	0.4884 (7.982)	0.3605 (6.530)	0.2982 (6.768)	0.1731 (3.666) [-0.569]
Taxi	-0.45	-0.7007 (-12.557)	-0.7698 (-13.065)	-0.4831 (-12.957)	-0.5218 (-12.116) [1.815]	-0.45	-0.1166 (-2.274)	-0.3101 (-5.905)	-0.2043 (-4.159)	-0.4610 (-10.967) [-0.261]
TCOST	-0.005	-0.0070 (-11.359)	-0.0070 (-11.279)	-0.0049 (-10.912)	-0.0055 (-11.011) [-0.995]	-0.005	-0.0053 (-8.901)	-0.0052 (-8.453)	-0.0041 (-8.067)	-0.0049 (-9.721) [0.232]
TTIME	-0.08	-0.1044 (-36.560)	-0.1005 (-32.035)	-0.0702 (-30.768)	-0.0804 (-31.172) [-0.148]	-0.08	-0.0835 (-31.464)	-0.0760 (-27.932)	-0.0614 (-21.918)	-0.0791 (-30.845) [0.344]
ATIME	-0.16	-0.2012 (-47.029)	-0.1954 (-41.264)	-0.1379 (-35.499)	-0.1563 (-36.078) [0.860]	-0.16	-0.1765 (-44.919)	-0.1643 (-38.757)	-0.1323 (-23.712)	-0.1596 (-36.698) [0.103]
Income Dummy	1.2	1.4928 (24.094)	1.4755 (23.953)	1.0686 (21.306)	1.1776 (21.464) [-0.409]	1.2	1.2454 (21.117)	1.2174 (20.828)	0.9998 (15.243)	1.1866 (21.842) [-0.247]
f	0.7071		0.8945 (23.953)			0.7071		0.7458 (22.558)		
s_m	0.9069			0.5100 (4.597)	0.7601 (8.352) [-1.613]	0.9069			0.5441 (5.880)	0.8472 (9.350) [-0.658]
Iter.		5	5	6	3		5	5	7	2
l(θ)		-0.93469	-0.93426	-0.93688	-0.93291		-1.03180	-1.02919	-1.03138	-1.02867
CPU Time [min]		0.6	0.8	35.5	42.5		0.7	0.8	35.2	152.5

It can be seen that the ML model allows to recover properly all the values of the taste parameters, which is shown by the t statistic against target, that is less than 1.96 in all cases, indicating that they are not significantly different.

On the other side, the processing time is considerable but not prohibitive both for Probit and ML. The ML estimations presented here were made with 200 repetitions randoms using numbers based on Halton sequences (Train, 1999). For the Probit case, we considered 10 repetitions of the GHK simulator. The CPU time values, reported in the last row of Table 1, correspond to runs in a personal computer with a 450 MHz Pentium II processor and 64 MB RAM. The largest time was required for the estimation of the homoscedastic ML.

It is worth noting that in the heteroscedastic case, there is a clear relation between the taste parameters estimated with NL model and those of the ML. The ratio between both parameters, is relatively constant. This effect has been found in several of our simulation experiments, and the ratio happens to be larger in cases of more correlation. This can be explained because the presence of heteroscedasticity affects the scale factor that multiplies the parameters. In the case of the ML model the common error component (μ) is fixed to a certain value on each repetition of the simulation, so, the scale factor of the Gumbel distribution corresponds to the ε random term only $\lambda=\pi/\sqrt{6}\,\sigma_\varepsilon$. While, in the case of the NL model, even dismissing the heteroscedasticity, it is the addition of both error components that is implicitly supposed to be Gumbel distributed, and in that case the scale is smaller. If all the alternatives have had the same variance of the error term, then the scale factor of the NL would have been $\lambda=\pi/\sqrt{6(\sigma_\varepsilon^2+\sigma_\mu^2)}$.

The response analysis was carried out by modifying the attributes levels in the magnitudes reported in Table 2. It can be seen that the defined policies correspond to strong changes in the attribute values, increasing up to double or diminishing to a half some values on each case.

Table 2: Policy changes

	Travel Cost				Travel Time				Access Time			
	Auto	Bus	Metro	Taxi	Auto	Bus	Metro	Taxi	Auto	Bus	Metro	Taxi
P1	2.0											
P2			2.0						1.5			
P3		2.0				0.5						
P4		0.5	1.5			2.0	0.3					
P5									1.5			2.0
P6	2.0		0.5		0.5	1.5				2.0		

The evaluation of the response analysis is carried out through the analysis of the difference between model predictions and simulated shares shown in Table 3, and is benefited from the use of the Chi squared index (Gunn and Bates, 1982), that measures that difference, for each policy change. It is calculated as $\chi^2=\sum_i \frac{(\hat{N}_i-N_i)^2}{N_i}$, where $\hat{N}_i$ is the number of individuals that choose alternative *i* according to the prediction made by the model (sum of

the predicted probabilities over the sample), and N_i is the number of individuals choosing alternative *i* according to the simulation (observations where the simulated utility of that alternative happens to be larger than all the others). The values of the Chi squared index are reported in Table 4.

Table 3: Market Shares

		Heteroscedastic Database							Homoscedastic Database						
	Alt.	Base	P1	P2	P3	P4	P5	P6	Base	P1	P2	P3	P4	P5	P6
BASE	1	3225	2482	2556	2754	2961	2837	3818	3244	2542	2596	2837	2973	2919	3665
	2	1056	1146	1190	2262	157	1341	157	918	1090	1113	2050	162	1290	158
	3	2625	3017	2862	2103	3858	3327	2897	2498	2843	2729	1975	3626	3171	2799
	4	1094	1355	1392	881	1024	495	1128	1340	1525	1562	1138	1239	620	1378
MNL	1	3227	2445	2495	2675	2941	2713	3806	3244	2512	2548	2820	3014	2836	3703
	2	1055	1232	1266	2281	193	1393	162	918	1065	1093	1869	193	1239	159
	3	2624	3034	2908	2166	3882	3398	2865	2498	2867	2758	2165	3557	3289	2717
	4	1094	1289	1331	878	983	496	1166	1340	1556	1601	1145	1236	636	1421
NL	1	3217	2438	2500	2681	2927	2717	3743	3242	2510	2577	2847	3005	2859	3588
	2	1058	1234	1269	2317	182	1391	157	919	1066	1097	1929	168	1228	145
	3	2628	3037	2901	2116	3907	3391	2931	2499	2869	2735	2066	3594	3262	2843
	4	1096	1291	1329	887	984	502	1169	1340	1556	1591	1159	1232	650	1425
PROBIT	1	3238	2466	2513	2725	3000	2791	3752	3263	2545	2589	2860	3068	2895	3643
	2	1067	1239	1275	2305	175	1405	150	932	1082	1115	1938	165	1270	139
	3	2610	2984	2859	2103	3830	3324	2928	2478	2826	2709	2080	3528	3227	2784
	4	1090	1318	1361	875	997	486	1170	1332	1553	1594	1129	1240	614	1435
ML	1	3224	2439	2491	2715	2945	2780	3746	3245	2504	2579	2861	3013	2884	3570
	2	1057	1221	1257	2297	180	1369	158	920	1065	1095	1927	167	1227	145
	3	2625	3008	2877	2092	3889	3345	2948	2498	2864	2724	2059	3591	3255	2854
	4	1094	1332	1375	896	986	506	1148	1338	1568	1602	1153	1228	633	1431

Table 4: c^2 Index

	Heteroscedastic Database				Homoscedastic Database			
	MNL	NL	Probit	ML	MNL	NL	Probit	ML
Base	0.00	0.03	0.40	0.00	0.00	0.00	0.53	0.01
P1	10.21	10.62	9.00	6.01	1.78	1.79	0.67	2.52
P2	9.65	9.81	7.50	5.65	2.52	0.91	0.81	1.42
P3	4.32	3.41	1.16	1.41	34.34	11.79	11.88	11.37
P4	10.35	6.64	3.41	8.03	11.63	5.44	13.46	5.41
P5	8.99	8.24	3.96	5.16	9.23	8.25	1.54	6.01
P6	1.87	3.40	3.33	2.64	4.15	5.01	4.91	6.64

It can be seen that the biggest differences between model predictions and simulation (virtual reality) are obtained for the MNL in the heteroscedastic database. This should not be surprising because it is an inappropriate model for that case. The predictions of the NL and ML are quite similar and practically indistinguishable, for the homoscedastic case, being both very similar to the virtual reality. This might be surprising, given that the parameters estimates look different in both models. On the other side, if the database is heteroscedastic some differences among predictions of both models are observed.

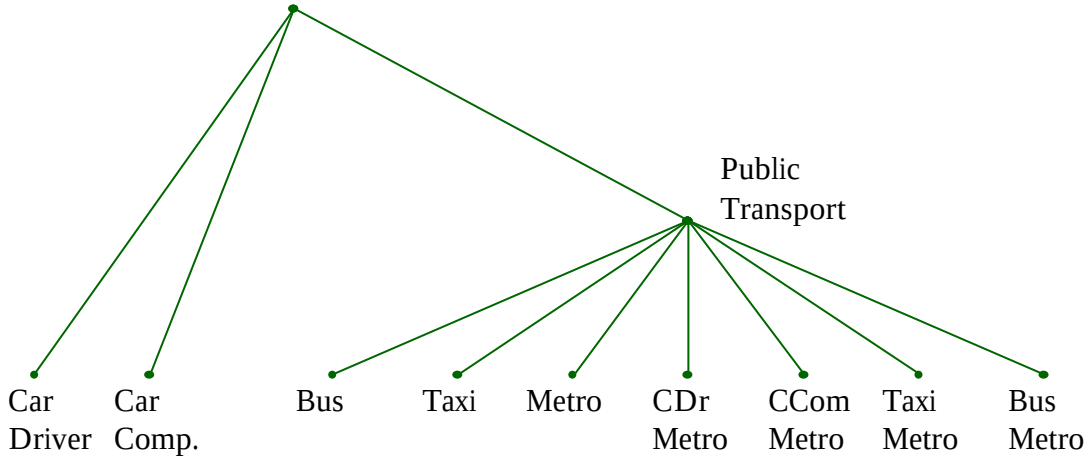
4.2. REAL DATA

The real data study was carried out using the Las Condes - CBD corridor database (Ortúzar and Donoso, 1983). The sample consists in 697 choice observations over nine alternatives. We worked with the following level of

service variables: TT (travel time), WALKT (walking time), WAITT (waiting time), C/w (cost divided by salary rate) and LICENCE (number of cars divided by number of driving licences in the household).

It can be argued that there is correlation between all the alternatives having a public transport component. That yields to a common nest for these alternatives as shown in Figure 1. For this specific case we estimated MNL, Independent Probit, NL, homoscedastic Probit and two nesting Mixed Logit specifications. The first one (HeNML), a heteroscedastic Nested Mixed Logit that considers an additional term inside the nest; while the second one (HoNML) is a homoscedastic model that also considers independent additional error terms in the non nested alternatives.

FIGURE 1: Nested Structure 1



For the NL model, the correlation coefficient within each nest can be obtained from the structural parameter f, while for Probit it corresponds to a parameter to be estimated. In the case of ML, it can be demonstrated that:

$$\hat{r}_{ML} = \frac{6\hat{s}_m^2}{6\hat{s}_m^2 + p^2} \qquad (21)$$

The estimation results are reported in Table 5. It can be seen that there are not major differences in the taste parameters ratios among models. However, there are important differences in the correlation estimated by the different models. While NL and Probit models detect very low correlation, the HeNML obtains a considerably higher value. This value practically decreases to a half when a homoscedastic structure is imposed. Two possible explanations can be enunciated for this strange result. On one hand it can be postulated that the base is in fact heteroscedastic and correlated. So when imposing homoscedasticity the correlation is underestimated. On the other hand it can be that the number of observations of the sample is not enough to be able to recover in a correct way the complex effects of the covariance structure. Processing time to convergence is also reported for the different models. Of course, processing time depends on the starting point and the algorithm used for the maximisation routine, but we thought it was worth reporting some values, as a reference. It can be seen that, under this implementation, ML models are comparable to NL in computing time, while Probit models

including an analogue covariance structure takes considerable more time (in this particular case, Probit has larger dimension of the integral that must be evaluated by simulation).

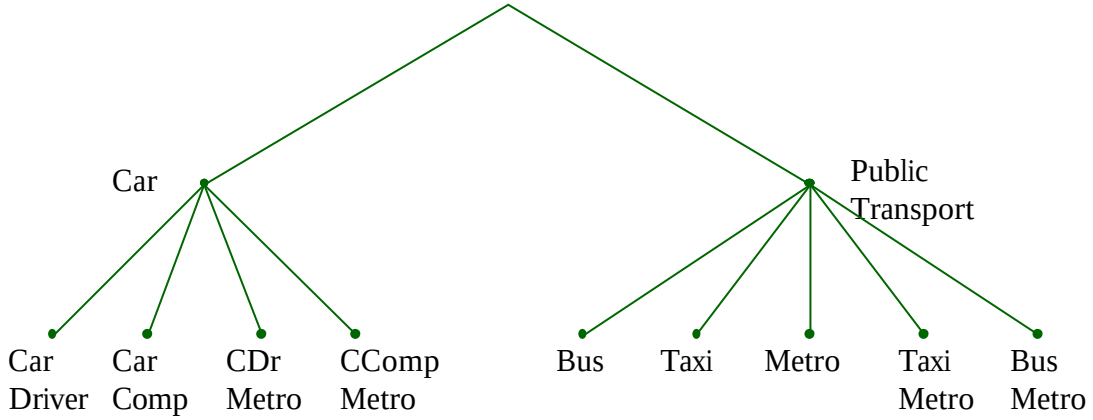
Table 5: Nested structure 1

Parameters' estimates, () t-value against zero

	MNL	Ind Probit	NL	Probit	HeNML	HoNML
TT	-0.0823 (-4.743)	-0.0554 (-4.031)	-0.0907 (-4.002)	-0.0550 (-1.774)	-0.0951 (-4.936)	-0.0909 (-4.783)
WALKT	-0.1610 (-8.625)	-0.1077 (-8.662)	-0.1531 (-7.019)	-0.1067 (-7.702)	-0.1904 (-7.488)	-0.1807 (-7.007)
WAITT	-0.2359 (-2.238)	-0.1475 (-2.028)	-0.2170 (-1.966)	-0.1484 (-0.800)	-0.2741 (-2.498)	-0.2641 (-2.426)
C/w	-0.0244 (-3.647)	-0.0143 (-2.028)	-0.0228 (-2.854)	-0.0142 (-1.979)	-0.0253 (-3.318)	-0.0267 (-3.424)
SEX	-0.2951 (-1.361)	-0.1531 (-1.169)	-0.2627 (-1.269)	-0.1479 (-1.049)	-0.2830 (-1.273)	-0.2923 (-1.305)
LICENCE	2.3606 (5.786)	1.4889 (5.902)	2.2018 (4.842)	1.4736 (1.730)	2.5321 (5.690)	2.5308 (5.496)
f			0.9181 (6.575)			
s_m					1.6061 (3.026)	0.8974 (2.191)
r			0.1571	0.0730 (0.299)	0.6106	0.3287
Iterations	6	9	5	27	6	8
Log likelihood	-1.36456	-1.37835	-1.36439	-1.37828	-1.35857	-1.35989
Time to convergence [min]	5.4	3.8	39.8	226.4	37.4	55.7

Another structure was tested, considering an additional nest for car alternatives (car driver and car companion), as shown in Figure 2.

FIGURE 2: Nested Structure 2



The results of the estimation for nesting structure 2 are presented in Table 6. Once again, there are similar taste parameters' rations among models (subjective values); but, strange results are found for the correlation

parameters. NL detects no correlation within the car nest (f_1 not significantly different from one) and significant correlation in the public transport nest. Practically the same happens with Probit (with lower values for both correlations). On the other side, the HeNML model, indicates exactly the opposite (no correlation in the public transport nest, significant correlation in the car nest). The HoNML, that imposes the same level of correlation in both nests, detects a value which is half the way between the extreme values detected by the other models. However we must note that the nested structure 2 corresponds to a two nests situation; so, as said in section 3, that correlation structure is unidentifiable. It can be shown that only the sum of both covariances can be identified. A warning must be made: An unidentifiable model can be estimated and there are no clear signals of an eventual problem. There is no way to detect this problem unless we study the covariance structure of the differentiated model prior to the estimation.

Table 6: Nested structure 2

	NL	Probit	HoNML	HeNML
TT	-0.0886 (-4.562)	-0.0581 (-4.533)	-0.0955 (-4.990)	-0.0962 (-5.010)
WALKT	-0.1292 (-6.486)	-0.0978 (-7.240)	-0.1809 (-8.081)	-0.1813 (-8.082)
WAITT	-0.1981 (-2.308)	-0.1464 (-2.378)	-0.3142 (-2.730)	-0.3149 (-2.731)
C/w	-0.0228 (-2.801)	-0.0141 (-2.786)	-0.0224 (-3.143)	-0.0226 (-1.167)
SEX	-0.2356 (-1.378)	-0.1366 (-1.164)	-0.2667 (-1.173)	-0.2660 (-1.167)
LICENCE	2.2306 (5.292)	1.5845 (5.468)	2.7949 (5.497)	2.7959 (5.476)
f_1	0.9369 (6.888)			
f_2	0.6309 (6.012)			
s_m			1.0358 (3.805)	1.4847 (3.781)
s_m				0.0219 (0.006)
r_1	0.1222	0.0333 (0.132)	0.3948	0.5727
r_2	0.6020	0.4840 (2.688)	0.3948	0.0003
Iterations	6	32	8	9
Log likelihood	-1.35909	-1.37494	-1.35698	-1.35671

5. CONCLUSIONS

From our point of view the ML model is an interesting, very flexible and useful modelling alternative, allowing to model and estimate correlation and

heteroscedasticity with a personal computer in a moderate time. In this context, it can become a real competitor to Probit, usually considered as the only or principal way to make more flexible the modelling of discrete choices. Nevertheless it is quite important to have a clear assessment of its properties and to justify properly any specific structure over the basis of theoretical considerations prior to the estimation of the parameters.

The covariance matrix associated to ML depends on the specification given to the additional error terms and it can be as general as desired. In this sense, it offers a more flexible structure that other models, in particular it has the capacity of recognising correlated alternatives and taste variations expressed through random parameters.

In this work, two numerical applications where presented, one based on simulation experiments and another one with real data, both in the context of similar alternatives. The latter motivates a nesting error structure. It is shown analytically that the Nested Mixed Logit is not equivalent to Nested Logit at least considering its covariance structure. However, for the reported correlation level, if the ML is not adjusted to obtain a homoscedastic covariance matrix, then the predicted market shares for both do not present severe differences. So, we understand that these models could approximate a situation like the one presented here.

The fact that the nested structure for Mixed Logit commits homoscedasticity when defining correlation, could be seen as a problem, if you want to compare it with Nested Logit, or as an advantage from the point of view that it allows to represent another covariance structure, that includes correlation and heteroscedasticity.

Acknowledgements

This study was partially funded by Fondecyt, and the Millennium program.

References

Algers, S., Bergstrom, P., Dahlberg, M. and Dillen, J. (1999) Mixed logit estimation of the value of travel time. Working Paper, Department of Economics, Uppsala University.

Aptech Systems (1994) *GAUSS User's Manual*. Maple Valley, CA.

Ben-Akiva, M.E. and Bolduc, D. (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Department d'Economique, Université Laval.

Ben-Akiva, M., Bolduc, D. and Walker, J. (2001) Specification, Identification, and Estimation of the Logit Kernel (or Continous Mixed Logit) Model. Working Paper presented at the 5th Invitational Choice Symposium, Asilomar, California.

Ben-Akiva, M.E. and Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Mass.

Bolduc, D (1992) Generalized autoregressive error in the multinomial probit model**.** *Transportation Research* **26B**, 155-170.

Börsch-Supan, A. and Hajivassiliou, V.A. (1993) Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* **58**, 347-368.

Boyd, J. and Melman, R. (1980) The effect of fuel economy standards on the US automotive market: an hedonic demand analysis. *Transportation Research* **14A**, 367-378.

Brownstone, D. and Train, K. (1999) Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* **89**,109-129.

Cardell, N. and Dunbar, F. (1980) Measuring the societal impacts of car mobile downsizing*. Transportation Research* **14A**, 423-434.

Daganzo, C.F. (1979) *Multinomial Probit: The Theory and its Applications to Travel Demand Forecasting*. Academic Press, New York.

Gunn, H.F. and Bates, J.J. (1982) Statistical aspects of travel demand modelling. *Transportation Research* **16A**, 371-382.

Hajivassiliou, V. and Ruud, P. (1994) Classical estimation methods for LDV models using simulation. In R. Engle and D. McFadden (Eds.) *Handbook of Econometrics*, **Vol. IV**, Elsevier, New York.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.) *Frontiers in Econometrics*. Academic Press, New York.

McFadden, D. and Train, K. (1997) Mixed MNL models for discrete response. Working Paper, Department of Economics, University of California at Berkeley.

Munizaga, M.A. and Alvarez-Daziano, R. (2000) Modelos mixed logit: uso y potencialidades. *Actas del XI Congreso Panamericano de Ingeniería de Tránsito y Transporte* , Noviembre, Gramado, Brasil, 523-535.

Munizaga, M.A., Heydecker, B.G. and Ortúzar. J.de D. (1997) On the error structure of discrete choice models. *Traffic Engineering and Control* **38**, 593-597.

Munizaga, M.A. and Ortúzar, J.deD. (1997) On the applicability of the multinomial probit model. *Proceedings of the 25th European Transport Forum, Vol. P415*. PTRC Education and Research Services Ltd.. Londres.

Munizaga, M.A. and Ortúzar, J.deD. (1999a) Correlación entre alternativas: el modelo Logit Jerárquico en profundidad. *Actas del IX Congreso Chileno de Ingenieria de Transporte*. 109-120. Sociedad Chilena de Ingenieria de Transporte. Santiago.

Munizaga, M.A. and Ortúzar, J.deD. (1999b) Nested logit modelling: some hard facts. *27th European Transport Conference. Vol.P324.* PTRC Education and Research Services Ltd.. London. 25-36.

Ortúzar, J.deD. and Donoso, P.C.F. (1983) Survey design, implementation, data coding and evaluation for the estimation of disaggregate choice models in Santiago, Chile. *2nd International Conference on Survey Methods in Transport*, September, Hungerford Hill Village, Australia.

Ortúzar, J.deD. and Willumsen, L.G. (1994) *Modelling Transport*. Second Edition. John Wiley and Sons. Chichester.

Revelt, D. and Train, K.E. (1998) Mixed logit with repeated choices: household's choices of appliance efficiency level. *Review of Economics and Statistics* **80**. 647-657.

Train, K. (1999) Halton sequences for mixed logit. Working paper. Department of Economics. University of California at Berkeley.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* **9A**. 285-344.

Williams, H.C.W.L. and Ortúzar, J.deD. (1982) Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research* **16B**. 167-219.

---

[i] In practical terms, the distribution of the random terms is usually assumed Normal, existing a variety of justifications behind this assumption. Another distribution that has been used is the log-normal, specially in those cases where sign restrictions are necessary (for a specific parameter).

[ii] β is a vector of parameters of dimension *L* (there are explanatory variables *L* in the deterministic component of the utility function); x_{in} is a vector of attributes of dimension *L*; η_{in} is a random vector of dimension *K* which components have zero mean and covariance matrix Ω; z_{in} is a vector of attributes associated with alternative *i* and individual *n*, and has dimension *K*; finally, ε_{in} is a random variable that represents the stochastic error.

[iii] The covariance matrix is of dimension $J \times J$. In effect, as Ω is of dimension $K \times K$ (with *K* the number can of random components), and z_n has dimension $K \times J$, then $z_n^t \cdot \Omega \cdot z_n$ is a matrix of dimension $J \times J$; Then adding this last to Σ_e, which is of dimension $J \times J$, finally $\dim \Sigma = \dim\left(z_n^t \cdot \Omega \cdot z_n + \Sigma_e\right) = J \times J$.

[iv] A distribution so that $\eta_{kn} + \max_i \varepsilon_{in} \sim$ Gumbel $(0, \beta)$

[v] Because it's based on the differences $\varepsilon_j - \varepsilon_i$, with *i* the chosen alternative and *j* each of the rest *J-1* alternatives.

[vi] http://elsa.berkeley.edu/~train/software.html