DRAFT

MIXED LOGIT VS. NESTED LOGIT AND PROBIT MODELS

Marcela A. Munizaga and Ricardo Alvarez-Daziano
Departamento de Ingeniería Civil, Universidad de Chile. Casilla 228-3, Santiago, Chile.
mamuniza@cec.uchile.cl, ralvarez@cec.uchile.cl
http://tamarugo.cec.uchile.cl/~dicidet/

ABSTRACT

The development of transport demand modelling can be described as a search of flexible models adapting to a greater number of practical situations. However, this search has been characterised by a flexibility-estimability trade off. In one hand, there are the traditional models of the Logit family that offer closed choice probabilities, but with restrictive assumptions that not always are properly justified. On the other hand, the Probit model allows to work with an error structure general in principle, but its estimation is quite complex and subject to identification restrictions. In this context, in addition by technological advances in term of computer's power and numerical methods, the use of simplified models has been questioned and it has appeared with force a new alternative of modelling: the Mixed Logit model.

In this paper we study both theoretically and empirically the antecedents that sustain the formulation of Mixed Logit model. Through an analysis of the covariance matrix we discuss how these models are able to model conditions in which independence and homoscedasticity are violated. This analysis is complemented with two numerical applications that allow to verify the real possibility of using this model and its capacity to adapt to practical situations. In the simulation experiments data bases are constructed so that it allows to objectively control the goodness of fit of the model, the reproduction of the calibration sample and the level of answer to changes in the attributes of the alternatives. The application with real data tries to validate the empirical study and to verify the feasibility to apply sophisticated econometric tools. Although its estimation requires simulation, it is observed that in general the model gives to a suitable reproduction of parameters and a good adjustment to the changes of policy.

We conclude that Mixed Logit models constitute an interesting and powerful alternative for discrete choice modelling. Nevertheless, as in the case of any flexible model, it is necessary to be rigorous in the construction and implementation of a particular specification, justifying suitably the any assumption done and knowing clearly its consequences previous to the estimation of the parameters.

Prepared for
5th tri-annual Invitational Choice Symposium
Workshop: Hybrid Choice Models, Formulation and Practical Issues
Asilomar, June 2001

1. INTRODUCTION

The so-called "Mixed Logit" Models have irrupted strongly in the theoretical environment of transport demand modelling in the last years (Ben Akiva and Bolduc, 1996; Brownstone and Train, 1999). It is a modelling alternative that could be located between the Logit and the Probit model. Its promoters claim it has the flexibility of Probit, keeping part of the simplicity of Logit. In this work we analyse its formulation in detail, with an optic of impartiality, verifying the consequence of its hypothesis.

In the context of discrete choice modelling, the most common approach 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 ε_{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).

Now, the models used the most are Multinomial Logit (McFadden, 1974), which is derived assuming that the error terms ε_{in} are iid Gumbel and the Nested Logit (Williams, 1977), that is derived as an extension of the last, 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 in the last years appear Mixed Logit models (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; in this way, apart from the iid Gumbel component, keeping the basic model as a Logit, other components are added, allowing to model correlation and/or heteroscedasticity. This lets to gain generality, but the estimation is not any more as simple as in the Logit case, 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 most common 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. In both cases, they can be between alternatives and between observations. 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, easy to associate 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 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 Mixed Logit 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.*, 1998).

Mixed Logit models assume a utility function U_{in} conformed by a deterministic component V_{in}, a random component ε_{in} independent and identically 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:

$$U_{in} = V_{in} + \eta_{in} + \varepsilon_{in} \tag{1}$$

where ε_{in} ~ Gumbel(0, λ) and η_{in} ~ $f(\eta/\theta^*)$, with f a general density function and θ^* are fixed parameters that describe it (*eg* mean and variance)[1]. As ε is iid Gumbel, then the probability conditional in η of individual n choosing alternative i corresponds exactly to the Multinomial Logit model:

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

So, the probability of choosing the alternative corresponds to the integral of the conditional

[1] 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 (for a specific parameter) are necessary.

probability over all the possible values of $\boldsymbol{h}$, which depends on the parameters characterising the distribution, this is:

$$P_{in} = \int L_{in}(\boldsymbol{h})\, f(\boldsymbol{h}/\boldsymbol{q}^*)\, d\boldsymbol{h} \tag{3}$$

As a particular case, it can be assumed a utility function with the following specification[2]:

$$U_{in} = \underbrace{\boldsymbol{b}^t x_{in}}_{V_{in}} + \underbrace{\boldsymbol{m}_{in}^t z_{in}}_{h_{in}} + \boldsymbol{e}_{in} \tag{4}$$

In this expression the assumption is that the deterministic component of the utility is linear in the $\boldsymbol{b}$ parameters that multiply the attributes x_{in}. Furthermore, it is assumed that $\boldsymbol{h}$ depends of certain parameters ($\boldsymbol{m}_n$) and data observed related to alternative i (z_{in}), relation which is also supposed lineal in the parameters. An additional assumption is that the $\boldsymbol{m}$ term is a property of the individual, with no variation over alternatives. The latter means:

$$\boldsymbol{h}_{in} = \boldsymbol{m}_n^t z_{in} \tag{5}$$

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 $K \times J$ that contains the vectors z_{in} for each alternative belonging to the choice set of the individual ($i \in C_n$) and $\boldsymbol{e}_n$ a random vector iid Gumbel with covariance matrix Σ_e containing each $\boldsymbol{e}_{in}$. If it is assumed that each term of $\boldsymbol{m}_n$ has a density function with zero mean and $\boldsymbol{s}_k^2$ variance, and that the vector has a joint covariance matrix Ω, then the covariance matrix of the model (Σ), can be written as:

$$\Sigma = z_n^t \cdot \Omega \cdot z_n + \Sigma_e = z_n^t \cdot \Omega \cdot z_n + \boldsymbol{s}_e^2 I \tag{6}$$

It is clear that the matrix is positive definite and that its dimension is well defined[3] and from this general expression it can be concluded that the model is capable to model correlation and heteroscedasticity between alternatives. In effect, if we obtain the covariance between two

---

[2] $\boldsymbol{b}$ 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; $\boldsymbol{m}_n$ 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, $\boldsymbol{e}_{in}$ is a random variable that represents the stochastic error.

[3] 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 $\dim \Sigma = \dim\left(z_n^t \cdot \Omega \cdot z_n + \Sigma_e\right) = J \times J$.

alternatives, for $i, j \in C_n$ with $i \neq j$:

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

which in general will be different from zero if for at least one k, $s_k^2 > 0$ and $z_{kin}, z_{kjn} \neq 0$. In that case, there will be presence of correlation between alternatives i and j.

For the variance,

$$\text{var}(U_{in}) = \sum_{k=1}^{K} z_{kin}^2 s_k^2 + \frac{p}{6 l^2} \tag{8}$$

Then if $\text{var}(U_{in}) \neq \text{var}(U_{jn})$ it will be heteroscedasticity between those alternatives.

We can see that this is a different form to justify a particular model. The usual form is to make assumptions directly over the covariance matrix of the error term e_{in}, like for example in the case of Probit. While in a Mixed Logit model an error structure is built adding terms that are source of correlation and/or heteroscedasticity.

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, 2001). 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 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.

$$P_{in} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \frac{\exp(b^t x_{in} + m_n^t z_{in})}{\sum_{l=1}^{J} \exp(b^t x_{\mathrm{ln}} + m_n^t z_{\mathrm{ln}})} \right\} f(m_{1n}) \cdot \ldots \cdot f(m_{Kn}) dm_{1n} \ldots dm_{Kn} \tag{9}$$

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 Multinomial Logit form can be exploited.

Then, if R values of $\boldsymbol{h}$ are obtained from its density function $f(\boldsymbol{h}/\boldsymbol{q}^*)$, then for each of this repetitions it is possible to calculate

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

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

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

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).

3. MIXED LOGIT COMPARED TO NESTED LOGIT

A subject that has been matter of confusion is that a particular Mixed Logit specification could be equivalent to a Nested Logit model. This last 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, the corresponding off diagonal elements will be different from zero.

Brownstone and Train (1999) present a Mixed Logit model that they call “analogue” to a Nested Logit. This particular model is built grouping the alternatives into nests; then, in the utility function a dummy variable is added for each nest indicating if the alternative belongs or not to it. A common random parameter is associated to each one of these variables. In this way the model has a correlation structure such that in the alternatives belonging to the same nest an off diagonal term appears. The authors conclude that in that way the pattern of correlation is equal to that of the Nested Logit. Nevertheless, 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. These alternatives are car, bus and metro (underground). Let us also suppose that bus and metro are correlated, because of being perceived as more similar between them than car. This case, that corresponds to a Nested Logit with a public transport nest could be modelled as a Mixed Logit with the following specification according to Brownstone and Train (1999):

$$\begin{aligned} U_{car} &= V_{car} + e_{car} \\ U_{bus} &= V_{bus} + m + e_{bus} \\ U_{metro} &= V_{metro} + m + e_{metro} \end{aligned} \tag{13}$$

where m is a random term with zero mean and variance s_m^2, and $e_.$ is an iid Gumbel term with variance s_e^2. It is easy to see that the covariance matrix of this model is:

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

This matrix has off diagonal terms indicating correlation between bus and metro alternatives; however, it is heteroscedastic. So this model is not really equivalent to the Nested Logit in terms of error structure, because the latter is homoscedastic by definition.

The correlation between bus and metro alternatives is given by:

$$corr(U_{bus}, U_{metro}) = r_{bus,metro} = \frac{s_m^2}{s_m^2 + s_e^2} = \frac{1}{1 + \dfrac{s_e^2}{s_m^2}} \tag{15}$$

Then,

- If $s_m^2 << s_e^2$, then $r_{bus,metro} \to 0$
- If $s_m^2 = s_e^2$, then $r_{bus,metro} = 0.5$
- If $s_m^2 >> s_e^2$, then $r_{bus,metro} \to 1$

From the shown cases, it is clear that larger the deviation of m compared to that of the iid Gumbel error, larger will be the correlation obtained. This is a reasonable result, because m is the common term that imposes the presence of correlation between alternatives bus and metro.

The covariance matrix shows terms outside the diagonal indicating correlation between the alternative bus and metro. However, it is heteroscedastic. Therefore, this model is not in fact equivalent to the Nested Logit in terms of the error structure, since the latter is homoscedastic by definition. It is necessary to notice that this situation can be overcome adding an additional error component in the not nested alternative, that is to say:

$$\begin{aligned} U_{car} &= V_{car} + m_1 + e_{car} \\ U_{bus} &= V_{bus} + m_2 + e_{bus} \quad , \\ U_{metro} &= V_{metro} + m_2 + e_{metro} \end{aligned} \tag{17}$$

where m_1, m_2 are iid N(0,s^2_m).

Accordingly to this:

$$\Sigma = \begin{bmatrix} s_m^2 + s_e^2 & 0 & 0 \\ 0 & s_m^2 + s_e^2 & s_m^2 \\ 0 & s_m^2 & s_m^2 + s_e^2 \end{bmatrix} \tag{18}$$

However, this additional term is difficult to justify and it doesn't have a direct theoretical interpretation. Also, the problem has not been overcome in fact. Indeed, let us suppose the presence of a new alternative, for example car companion - and let us think, then that car refers to car driver -. In practical terms one can argue car driver and car companion are considered as similar alternatives for the individual. This situation can be modelled by a Nested Logit:

$$\begin{aligned} U_{car} &= V_{car} + m_1 + e_{car} \\ U_{car-comp} &= V_{car-comp} + m_1 + e_{car-comp} \\ U_{bus} &= V_{bus} + m_2 + e_{bus} \\ U_{metro} &= V_{metro} + m_2 + e_{metro} \end{aligned} \tag{19}$$

where $e_{auto\text{-}ch}$, $e_{auto\text{-}ac}$ are iid Gumbel(0,l_1) – with variances $\sigma^2_{\varepsilon 1}$ – and e_{bus}, e_{metro} are iid Gumbel(0, l_2) – with variance $\sigma^2_{\varepsilon 2}$ –. m_1 y m_2 distributes according to the suppositions of Nested Logit[4], with equal variances a s^2_{m1} y s^2_{m2}, respectively. Scale parameters l_1 y l_2 must be chosen so that $s^2_{m1} + \sigma^2_{\varepsilon 1} = s^2_{m2} + \sigma^2_{\varepsilon 2} = p^2/6b$, that is the joint variance associated to an error term x_{in} iid Gumbel(0,b). That yields the following covariance structure:

$$\Sigma_{NL} = \begin{bmatrix} s_{m1}^2 + s_{e1}^2 & s_{m1}^2 & 0 & 0 \\ s_{m1}^2 & s_{m1}^2 + s_{e1}^2 & 0 & 0 \\ 0 & 0 & s_{m2}^2 + s_{e2}^2 & s_{m2}^2 \\ 0 & 0 & s_{m2}^2 & s_{m2}^2 + s_{e2}^2 \end{bmatrix} = \begin{bmatrix} s_x^2 & s_{m1}^2 & 0 & 0 \\ s_{m1}^2 & s_x^2 & 0 & 0 \\ 0 & 0 & s_x^2 & s_{m2}^2 \\ 0 & 0 & s_{m2}^2 & s_x^2 \end{bmatrix} \tag{20}$$

If we define $f_1 = b/l_1$ y $f_2 = b/l_2$, then:

$$\Sigma_{NLJ} = \begin{bmatrix} s_x^2 & (1-f_1^2)s_x^2 & 0 & 0 \\ (1-f_1^2)s_x^2 & s_x^2 & 0 & 0 \\ 0 & 0 & s_x^2 & (1-f_2^2)s_x^2 \\ 0 & 0 & (1-f_2^2)s_x^2 & s_x^2 \end{bmatrix} \tag{21}$$

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.

[4] A distribution so that $m_{kn} + \max_i e_{in} \sim$ Gumbel $(0,b)$

If we model the same situation with a Mixed Logit structure, then

$$
\begin{aligned}
U_{car} &= V_{car} + m_1 + e_{car} \\
U_{car-comp} &= V_{car-comp} + m_1 + e_{car-comp} \\
U_{bus} &= V_{bus} + m_2 + e_{bus} \\
U_{metro} &= V_{metro} + m_2 + e_{metro}
\end{aligned}
, \qquad (22)
$$

where $m_1 \sim N(0, s_{m1}^2)$, $m_2 \sim N(0, s_{m2}^2)$ and e_{in} is iid Gumbel(0,λ) with variance s_e^2. The covariance matrix associated to this model is:

$$
\Sigma_{ML} = \begin{bmatrix} s_{m1}^2 + s_e^2 & s_{m1}^2 & 0 & 0 \\ s_{m1}^2 & s_{m1}^2 + s_e^2 & 0 & 0 \\ 0 & 0 & s_{m2}^2 + s_e^2 & s_{m2}^2 \\ 0 & 0 & s_{m2}^2 & s_{m2}^2 + s_e^2 \end{bmatrix} \qquad (23)
$$

This matrix would be of a heteroscedastic nature, unless we assume $s_{m1}^2 = s_{m2}^2 = s_m^2$. Nonetheless, under this additional supposition, the correlation within each nest must be the same. That is, we will consider an equivalent matrix with the one of NL, only if $f_1 = f_2$. Therefore, in the described Mixed Logit structure there is a clear trade off between correlation and heteroscedasticity that it is not observed in the Nested Logit. 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. By the way, it can be demonstrated that if we consider the matrix presented in (), only one parameter can be estimated. If we differentiate the model, the covariance matrix is so that only the sum $s_{m1}^2 + s_{m2}^2$ can be identified. This complicate the analysis of a Nested Mixed Logit with two nests.

4. MIXED LOGIT COMPARED TO PROBIT

As mentioned before, the Mixed Logit model is built 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 is assumed with a general covariance matrix. In effect, a multinomial Probit model is derived assuming that given a utility function $U_{in} = V_{in} + e_{in}$, the vector $e_n = (e_{1n}, \ldots, e_{Jn})^t$ 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 used estimation method is the simulated maximum likelihood with the Geweke-Hajivassilou-Keane (Börsch-Supan and Hajivassilou, 1993) simulator, which recursively reduce the dimension of the integral up to an equivalent problem where repetitions of a truncated unidimensional 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[5] and *K* for case of Mixed Logit. In this way, if *K<J-1* 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 Mixed Logit model is smaller than the number of alternatives.

5. SIMULATION ANALYSIS

5.1 Experimental Design

Following the methodology of Williams and Ortúzar (1982), it was carried out a simulated experiment with the purpose of checking the real feasibility of application of the model. We elaborate different synthetic databases considering means and deviations of the attributes from a real database for Santiago de Chile. It was considered a situation of modal choice with four alternatives (car, bus, metro and taxi) and four explanatory variables (travel cost, travel time, access time, income dummy).

Table 1: Taste Parameters

Car	Metro	Taxi	Travel Cost	Travel Time	Access Time	Income Dummy
-0.40	0.20	-0.45	-0.005	-0.08	-0.16	1.2

It is sought to model the case where the alternative bus and metro are considered similar. To build the stochastic part of the utility function we worked with the Nested Mixed Logit described above and outlined by Brownstone and Train (1999). It was considered an error term iid Gumbel(0,l) and, additionally, an error term m_n distributed Normal(0,s_m^2) with the purpose of modelling correlation. In the first place it was only considered this last one in the alternative bus and metro, obtaining a heteroscedastic covariance matrix. Also we considered iid Normal errors for the non nested alternatives seeking to obtain a homoscedastic matrix corresponding to a structure theoretically modelable with a NL. If we assume that m_n ~ N(0,s_m^2), then it is possible to say that $m_n = s_n s_m$ with s_n standard Normal distributed. As s_m is unknown, then we shall estimate its value, with which its distribution is completely described. Notice that as there is a Gumbel error term, the estimate parameter will be scaled so that $\hat{s}_m = l s_m$.

For the estimation of the Multinomial and Nested Logit Models, as well as Probit, we used a self-made code programmed in Gauss (Aptech Systems, 1994) based on the maximum likelihood routine. For the estimation of Mixed Logit we used a flexible code programmed in Gauss by Kenneth Train, available in his web page[6].The Mixed Logit estimations where made with 200 repetitions using numbers based on Halton sequences (Train, 1999; Bhat, 2000). For the Probit case, we considered 10 repetitions of the GHK simulator. The reported values correspond to runs in a personal computer with a 450 MHz Pentium II processor and 64 MB RAM.

[5] Because it's based on the differences $e_j - e_i$, with *i* the chosen alternative and *j* each of the rest *J-1* alternatives.

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

In Table 2 the policy changes considered for the response analysis of the models are reported (Williams and Ortúzar, 1982); it may 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 Chi squared index (Gunn and Bates, 1982) is as a measure of error for each policy change; it is calculated as $c^2 = \sum_i \frac{(\hat{N}_i - N_i)^2}{N_i}$, where $\hat{N}_i$ is the number of individuals that choose alternative i according to the prediction made by the model, and N_i is the number of individuals choosing alternative i according to the simulation model.

5.2 Influence of the number of repetitions

As we described Mixed Logit and Probit both require simulation for their estimate. This motivates a convergence analysis, in the sense of observing the behaviour of the estimates considering variable the number of repetitions for the simulation. To make operative this comparison in a context of correlated alternatives, it was considered a database composed by 4,000 individuals and 4 alternatives, where two of those (specifically the alternatives 2 and 3) present a correlation coefficient equal to 0.5. The database was built assuming a Nested Mixed Logit, with a homoscedastic covariance matrix. The number of repetitions considered for the simulation took the following values: 5, 10, 25, 50, 100, 200, 250, 500, 750, 1000.

Fundamental aspects for the comparison are: time for convergence, loglikelihood, number of iterations. It is also interesting to observe what happens to the level of reproduction of the parameters with those the sample was created. Note that the dimension of integration of both Probit and Mixed Logit are the same and equals three (The number of alternatives - 1 for Probit; 3 additional error components for Mixed Logit: one shared that induces correlation and two independent to achieve homoscedasticity with the other alternatives). To be able to compare the predictive power of the models, it is considered a strong change in the value of certain attributes. Then it is possible to calculate the predicted market shares by each one of the models and to compare them with the observed (modelled) shares. To define the attribute changes we considered P4 from the policy plan defined. For the case of Mixed Logit we considered both, estimation with pseudorandom numbers (MLR) and based on Halton sequences (MLH).

The results for Probit are reported in Table 3, varying the number of repetitions for the GHK simulator. In general terms, the parameters stay stable even for a low number of repetitions.

Table 3: Probit, 4000 Observations with correlated alternatives. r = 0.5

	Target	5	10	25	50	100	200	250	500	750	1000
Car	-0.40	-0.2148 (-2.674)	-0.2747 (-3.298)	-0.3017 (-3.823)	-0.3292 (-3.753)	-0.3179 (-3.597)	-0.3260 (-3.746)	-0.3287 (-3.720)	-0.3279 (-3.656)	-0.3273 (-3.886)	-0.3287 (-3.753)
Metro	0.20	0.2370 (3.965)	0.2263 (3.662)	0.2533 (3.880)	0.2560 (3.889)	0.2542 (3.913)	0.2539 (3.900)	0.2541 (3.878)	0.2544 (3.874)	0.2530 (3.885)	0.2554 (3.894)
Taxi	-0.45	-0.2292 (-3.193)	-0.2637 (-3.621)	-0.2943 (-4.525)	-0.3247 (-4.250)	-0.3097 (-4.042)	-0.3203 (-4.214)	-0.3229 (-4.242)	-0.3227 (-4.183)	-0.3204 (-4.358)	-0.3234 (-4.286)
TCOST	-0.005	-0.0033 (-5.006)	-0.0033 (-4.888)	-0.0036 (-4.955)	-0.0037 (-4.894)	-0.0037 (-4.946)	-0.0037 (-4.912)	-0.0037 (-4.964)	-0.0038 (-4.991)	-0.0037 (-4.970)	-0.0038 (-4.989)
TTIME	-0.08	-0.0584 (-14.960)	-0.0616 (-15.487)	-0.0638 (-18.368)	-0.0655 (-16.022)	-0.0647 (-15.942)	-0.0652 (-16.312)	-0.0656 (-16.323)	-0.0656 (-16.122)	-0.0654 (-16.874)	-0.0656 (-16.582)
ATIME	-0.16	-0.1230 (-15.027)	-0.1283 (-15.848)	-0.1357 (-18.709)	-0.1395 (-16.927)	-0.1376 (-16.313)	-0.1388 (-16.748)	-0.1394 (-17.028)	-0.1395 (-16.796)	-0.1391 (-17.460)	-0.1397 (-17.334)
Income Dummy	1.2	0.9480 (10.063)	1.0156 (10.422)	1.0721 (11.662)	1.1193 (10.671)	1.1000 (10.469)	1.1126 (10.472)	1.1165 (10.704)	1.1176 (10.578)	1.1140 (10.964)	1.1183 (10.839)
s_m	0.9069	0.5263 (3.564)	0.6254 (4.749)	0.7382 (7.246)	0.8053 (6.854)	0.7746 (6.292)	0.7920 (6.580)	0.8024 (6.864)	0.8030 (6.767)	0.7947 (6.967)	0.8049 (7.010)
N° Iter.		10	7	6	6	6	6	6	6	6	6
Loglik.		-1.05450	-1.04922	-1.04560	-1.04472	-1.04586	-1.04604	-1.04533	-1.04516	-1.04525	-1.04519
Time for convrg.		25.06983	15.48433	23.95383	41.51167	72.73317	130.4957	136.2673	325.0023	438.8447	684.3543

The model presents certain difficulty to reproduce the parameter associated to correlation; however, it detects its presence for a considerably low number of repetitions, demonstrating the power of the Probit model.

The results for Mixed Logit are shown in Tables 4 and 5, considering simulation with Pseudo Monte Carlo method (MLR) and Quasi Monte Carlo (MLH), respectively.

With a low number of repetitions correlation is practically not detected. As a matter of fact, for very low values of the repetitions, the parameters cannot be compared directly with the target values that appear in the respective tables. This is explained by the fact that we considered a total variance such that if it is considered only an error Gumbel iid to explain it, then the scale parameter equals one. However, under the assumption of Mixed Logit, the error term has been divided into a Normal distributed component plus the Gumbel error term. Thus, the Gumbel term for the Mixed Logit explains a smaller portion of the total variance and, therefore, its scale parameter greater than one.

Table 4: MLR, 4000 Observations with correlated alternatives. r = 0.5

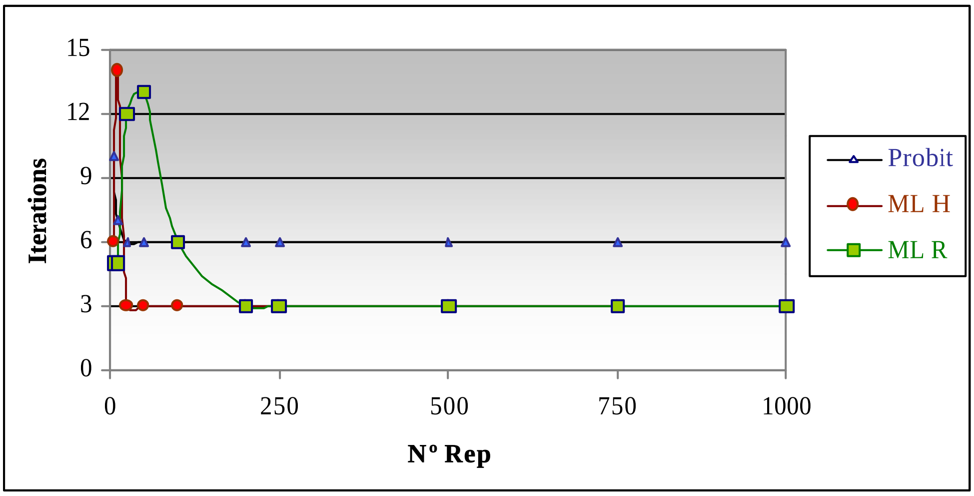
	Target	5	10	25	50	100	200	250	500	750	1000
Car	-0.5657	-0.2448 (-2.668)	-0.2456 (-2.676)	-0.2503 (-2.685)	-0.3769 (-2.833)	-0.4420 (-3.571)	-0.5063 (-3.908)	-0.5166 (-3.952)	-0.5235 (-3.975)	-0.5241 (-3.971)	-0.5231 (-3.969)
Metro	0.2828	0.3857 (4.574)	0.3856 (4.572)	0.3853 (4.556)	0.3769 (4.261)	0.3783 (4.011)	0.3808 (3.885)	0.3830 (3.888)	0.3843 (3.882)	0.3856 (3.885)	0.3845 (3.874)
Taxi	-0.6	-0.1777 (-2.510)	-0.1785 (-2.520)	-0.1826 (-2.528)	-0.2640 (-2.662)	-0.3904 (-3.696)	-0.4697 (-4.239)	-0.4823 (-4.297)	-0.4931 (-4.348)	-0.4978 (-4.370)	-0.4959 (-4.364)
TCOST	-0.0071	-0.0041 (-5.268)	-0.0041 (-5.262)	-0.0041 (-5.242)	-0.0044 (-4.918)	-0.0050 (-4.918)	-0.0054 (-4.936)	-0.0054 (-4.914)	-0.0055 (-4.936)	-0.0056 (-4.952)	-0.0056 (-4.954)
TTIME	-0.1131	-0.0820 (-22.066)	-0.0821 (-22.046)	-0.0824 (-21.373)	-0.0877 (-15.310)	-0.0955 (-16.367)	-0.1000 (-16.739)	-0.1005 (-16.752)	-0.1011 (-16.738)	-0.1014 (-16.718)	-0.1014 (-16.704)
ATIME	-0.2263	-0.1660 (-31.079)	-0.1661 (-31.041)	-0.1668 (-29.073)	-0.1787 (-15.923)	-0.1978 (-17.160)	-0.2094 (-17.603)	-0.2108 (-17.574)	-0.2123 (-17.534)	-0.2132 (-17.502)	-0.2131 (-17.502)
Income Dummy	1.6971	1.2229 (14.685)	1.2229 (14.673)	1.2315 (14.276)	1.3408 (10.473)	1.5239 (10.812)	1.6381 (10.845)	1.6518 (10.833)	1.6672 (10.801)	1.6738 (10.779)	1.6746 (10.774)
s_m	1.2825	0.0185 (0.361)	0.0486 (0.610)	0.1379 (0.770)	0.5799 (0.2358)	0.9624 (5.501)	1.1639 (6.930)	1.1889 (7.056)	1.2124 (7.170)	1.2251 (7.228)	1.2235 (7.222)
Nº Iter.		5	5	12	13	6	3	3	3	3	3
Loglik.		-1.04785	-1.04783	-1.04780	-1.04770	-1.04650	-1.04507	-1.04492	-1.04488	-1.04483	-1.04485
Time for convrg.		0.43567	0.769	8.919	24.68983	24.6615	32.0005	32.23033	55.1425	94.5285	113.0605

Table 5: MLH, 4000 Observations with correlated alternatives. r = 0.5

	Target	5	10	25	50	100	200	250	500	750	1000
Car	-0.5657	-0.2440 (-2.660)	-0.3010 (-2.879)	-0.4861 (-3.855)	-0.5220 (-3.973)	-0.5337 (-4.017)	-0.5321 (-4.028)	-0.5351 (-4.033)	-0.5375 (-4.048)	-0.5375 (-4.047)	-0.5375 (-4.047)
Metro	0.2828	0.3868 (4.587)	0.3798 (4.362)	0.3788 (3.903)	0.3887 (3.926)	0.3882 (3.883)	0.3877 (3.884)	0.3880 (3.878)	0.3885 (3.880)	0.3884 (3.877)	0.3885 (3.879)
Taxi	-0.6	-0.1777 (-2.510)	-0.2335 (-2.767)	-0.4516 (-4.231)	-0.4929 (-4.341)	-0.5122 (-4.455)	-0.5083 (-4.492)	-0.5130 (-4.495)	-0.5138 (-4.512)	-0.5154 (-4.519)	-0.5155 (-4.520)
TCOST	-0.0071	-0.0041 (-5.277)	-0.0043 (-5.105)	-0.0054 (-5.047)	-0.0055 (-4.902)	-0.0057 (-4.948)	-0.0056 (-4.968)	-0.0057 (-4.966)	-0.0057 (-4.961)	-0.0057 (-4.962)	-0.0057 (-4.962)
TTIME	-0.1131	-0.0820 (-22.055)	-0.0858 (-18.147)	-0.0991 (-17.130)	-0.1009 (-16.735)	-0.1022 (-16.786)	-0.1021 (-16.951)	-0.1023 (-16.895)	-0.1023 (-16.932)	-0.1024 (-16.921)	-0.1024 (-16.921)
ATIME	-0.2263	-0.1661 (-31.010)	-0.1745 (-20.783)	-0.2067 (-18.510)	-0.2124 (-17.389)	-0.2154 (-17.573)	-0.2150 (-17.881)	-0.2156 (-17.782)	-0.2158 (-17.820)	-0.2160 (-17.818)	-0.2161 (-17.818)
Income Dummy	1.6971	1.2226 (14.678)	1.2971 (12.409)	1.6039 (11.310)	1.6567 (10.772)	1.6975 (10.792)	1.6892 (10.905)	1.6964 (10.866)	1.6985 (10.877)	1.7003 (10.876)	1.7002 (10.877)
s_m	1.2825	0.0299 (0.268)	0.4666 (2.552)	1.1105 (7.186)	1.2121 (7.072)	1.2632 (7.433)	1.2514 (7.625)	1.2621 (7.609)	1.2651 (7.657)	1.2686 (7.676)	1.2683 (7.673)
Nº Iter.		6	14	3	3	3	3	3	3	3	3
Loglik.		-1.04786	-1.04735	-1.04485	-1.04495	-1.04449	-1.04437	-1.04444	-1.04434	-1.04434	-1.04434
Time for convrg.		0.5145	2.11733	1.04633	2.0185	14.45367	24.03817	28.67117	91.88383	141.4167	168.4447

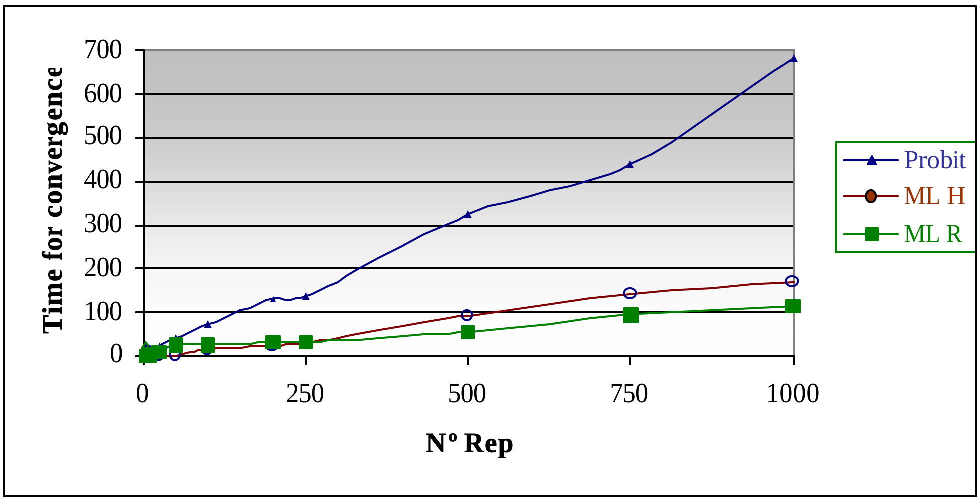
The number of iterations stabilises in 6 for Probit, starting from 25 repetitions. The Mixed Logit that considers Halton sequences (MLH) also stabilises from 25 repetitions, but this time in a value equals to 3. The same number of iterations is the one that we can observe for a Mixed Logit with random numbers (MLR), but now it stabilises in this number starting from 200 repetitions.

Graph 1: Number of iterations vs repetitions



Time for convergence is notoriously bigger for Probit. In fact, only considering 5 repetitions, this takes approximately half an hour, in comparison to the less than a minute that is observed for MLH and MLR. Although times for Probit are high, they do not discard their use, except for a very high number of repetitions, case for which the time for convergence overcomes ten hours.

Graph 2: Convergence time vs repetitions

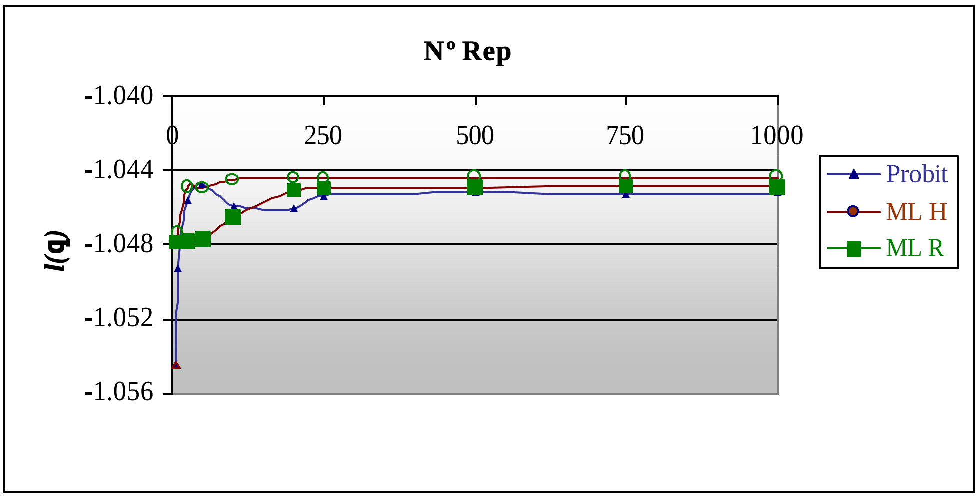


For a reduced number of repetitions, the lowest times of convergence are associated to the MLH. However, as the number of repetitions increases the convergence of the MLH becomes slower in comparison to the MLR. A tentative explanation is that the storage of the Halton sequences occupies an important amount of the memory dedicated to carry out the calculations.

For Probit a curious situation is observed. The highest value in the average log-likelihood is obtained for 50 repetitions (-1.04472), lowering for further repetitions and being stabilised in a relatively smaller value to the reached maximum (-1.04519 for 1000 repetitions).

The MLH achieves loglikelihood values bigger than -1.045 for 25 repetitions, coming closer to -1.044 as these increase. On the other hand, the MLR reaches values bigger than -1.045 starting from 250 repetitions.

Graph 3: Average log-likelihood vs repetitions



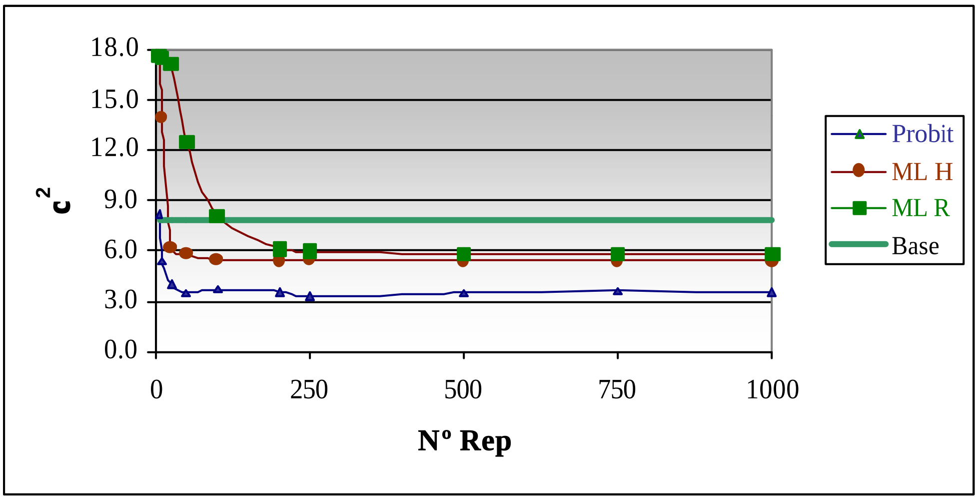
In the response analysis, the Probit achieves values under 10 of the Chi squared index starting from 5 repetitions (see Table 6). A 10 value is considered the threshold for acceptable predictions, given the inherent randomness of the process. Even from 10 repetitions it achieves values under the critical value of the index ($\chi^2_{95\%,3}$ = 7,815). Even more it is quickly stabilised in very low values, near 3.5. On the other hand, the MLH achieves values under the critical index starting from 25 repetitions, while MLR do it from 200 repetitions. By the way, the MLH is stabilised in an index 5,4 (100 repetitions) and the MLR makes it in 5,8 (500 repetitions).

Table 6: χ^2 index 4000 Observations, correlated alternatives . ρ = 0.5

	5	10	25	50	100	200	250	500	750	1000
Probit	8.20	5.39	4.02	3.49	3.70	3.57	3.26	3.46	3.57	3.73
MLR	17.60	13.96	6.16	5.81	5.43	5.40	5.44	5.40	5.44	5.44
MLH	17.67	17.52	17.16	12.49	8.12	6.12	5.96	5.83	5.82	5.76

This situation is graphically represented in Graph 4. Note how Probit responses quite well to the considered policy. ML has an adequate behaviour but it requires more repetitions than Probit. Certainly both simulators are different but they are inspired in Monte Carlo methods. The GHK simulator reduces the dimension of the integral up to an equivalent problem and requires repetitions of a truncated unidimensional normal deviate. The simulation required for ML is based on random draws that permit to calculate a well behaved function (the Logit expression for the probability).

Graph 4: χ^2 index vs repetitions



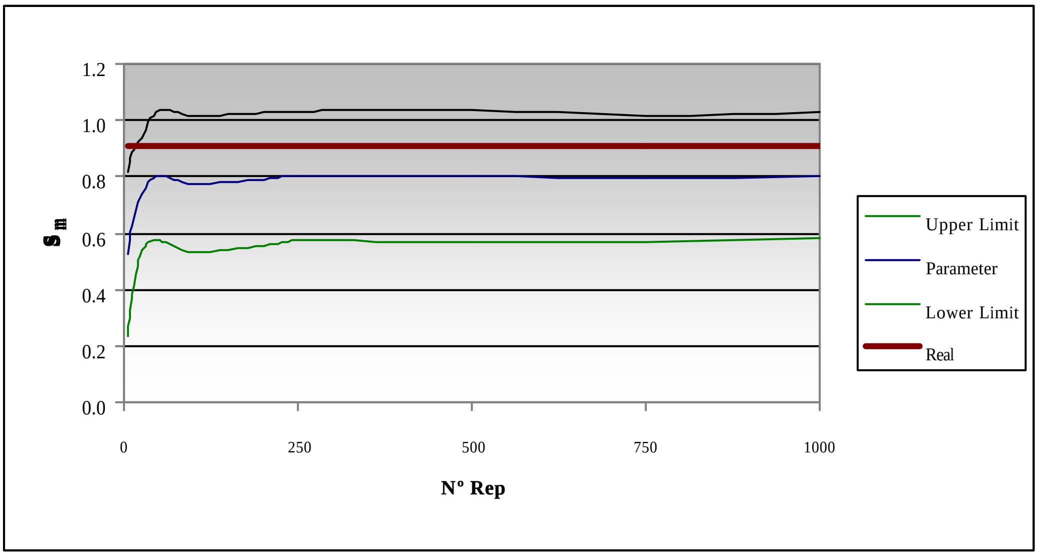
The Market Shares for P4 are shown in Table 7. All the models present only low differences between different number of repetitions for the simulation. Note that the χ^2 index reported above is calculated from the Market Shares presented in this table.

Table 7: Market Shares, 4000 Observations, correlated alternatives. ρ = 0.5

Modelo	Modos	5	10	25	50	100	200	250	500	750	1000
BASE	Auto	1548	1548	1548	1548	1548	1548	1548	1548	1548	1548
	Bus	68	68	68	68	68	68	68	68	68	68
	Metro	1761	1761	1761	1761	1761	1761	1761	1761	1761	1761
	Taxi	623	623	623	623	623	623	623	623	623	623
Probit	Auto	1571	1557	1550	1540	1539	1543	1541	1538	1537	1538
	Bus	90	86	84	83	84	83	83	83	83	84
	Metro	1720	1734	1745	1750	1751	1750	1751	1754	1755	1755
	Taxi	622	625	627	623	626	624	624	625	625	625
MLH	Auto	1527	1526	1525	1525	1526	1527	1525	1525	1526	1525
	Bus	102	98	87	87	86	86	86	86	86	86
	Metro	1755	1760	1774	1773	1772	1772	1773	1773	1773	1773
	Taxi	616	616	614	615	615	615	615	615	615	615
MLR	Auto	1526	1526	1526	1524	1523	1525	1526	1525	1524	1525
	Bus	102	102	102	97	91	88	87	87	87	87
	Metro	1756	1756	1756	1765	1772	1772	1772	1773	1774	1773
	Taxi	616	615	615	615	614	616	615	615	615	615

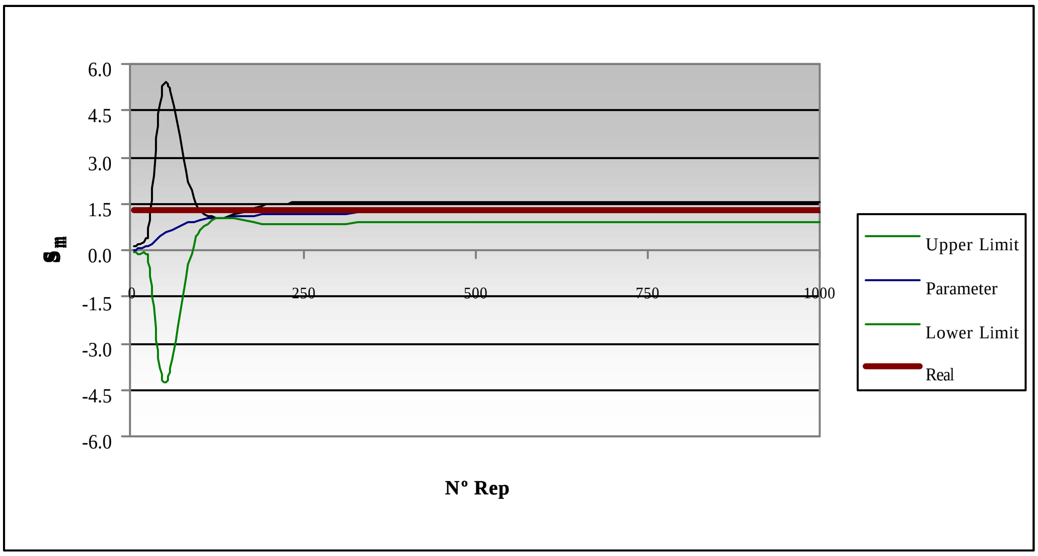
For the specific case of Probit skewed punctual parameters are obtained, in the sense that even increasing the number of repetitions, the parameters are stabilised in a value different from the real parameter. This is observed clearly in Graph 5, where the recovery of the parameter s_m can be appreciated.

Graph 5: Correlation parameter recovery. Probit Model



However, the confidence intervals appear appropriate (the real parameter is contained in the interval) starting from 50 repetitions. On the other hand, an excellent behaviour of the model is observed for a quite low number of repetitions (starting from 10 repetitions). It can be concluded for Probit that an excessively high number of repetitions is not required.

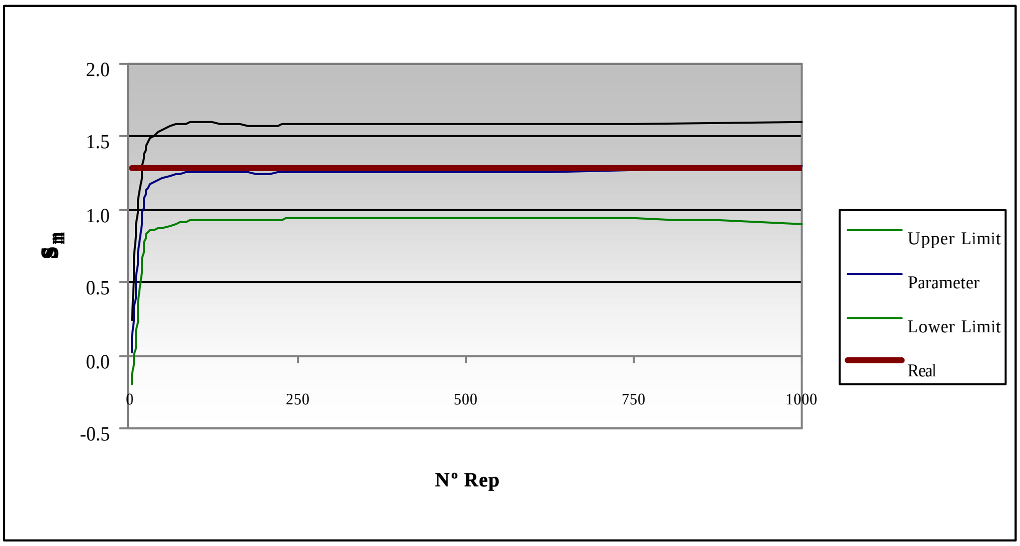
Graph 6: Correlation parameter recovery. MLR



When considering a Mixed Logit with simulation based on pseudorandom numbers (MLR), the estimation of s_m is appropriate starting from 100 repetitions (see Graph 6). By the way, it should be noticed that although in general the confidence interval for this same parameter is adequate, it presents a strange behaviour for 50 repetitions, motivated by an increase in the t-values. For the response analysis, good results are obtained from 200 repetitions.

When using Halton sequences for the simulation of the Mixed Logit (MLH), it is observed that the parameter that induces correlation is unbiased, in the sense that when increasing the number of repetitions, it is stabilised in a quite near value to the real parameter (Graph 7). These results are observed starting from an inferior number of repetitions in comparison with the MLH (25 repetitions against 100 of the MLH). Also, the model responds appropriately to policy changes starting from 25 repetitions.

Graph 7: Correlation parameter recovery. MLH



5.3 Detailed Analysis for correlated alternatives

A case with 8000 observations is presented considering a correlation coefficient equals to 0.5 between alternatives bus and metro[7], where $s_m = s_e$ (Munizaga and Alvarez, 2000). First we consider a heteroscedastic database, while the second database is homoscedastic. The total error variance was chosen so that the scale parameter when considering only an iid Gumbel disturbance equals 1, assuring an experiment not completely deterministic nor completely random.

The results for the estimation process of the Multinomial Logit (MNL), Nested Logit (NL), Probit and Mixed Logit (ML) are shown in Table 8, where the reference values are also reported. The

[7] We also considered higher correlation coefficients, but we prefer to report this particular experiment, because practical correlation is not substantially high.

table shows the estimations of the parameters for each model, the t statistic of significance and the t test over the reference value of the parameter for the ML model. For the NL, the reference value of f is calculated from the simulated correlation.

The ML model allows to recover properly all the values of the taste parameters with which the database was generated, which is shown by the t statistic, that is less than 1.96 in all cases (see Table 8, () t-value against zero, [] t-value against target).

Table 8: Simulation Results

	Heteroscedastic Database					Homoscedastic Database				
	Target	MNL	NL	Probit 10 Rep	ML 200 Rep	Target	MNL	NL	Probit 10 Rep	ML 200 Rep
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
$l(\theta)$		-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 is worth noting that there is a relation between the parameters estimated with NL model and those of the ML. The ratio between both parameters in each database, is relatively constant, and is larger in the case 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 is corresponding 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 sum of both error components that is supposed to be Gumbel distributed, and in that case the scale is smaller; if all the alternatives had the same variance of the error term, then the scale factor of the NL would be $\lambda = \pi / \sqrt{6(\sigma_e^2 + \sigma_m^2)}$ (Munizaga and Alvarez, 2000).

Table 9: 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

The biggest differences between the predictions of the models and the simulated ones (virtual reality) are obtained especially for the MNL in the heteroscedastic database. 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. However, if the database is heteroscedastic we observed some small differences among the predictions of both models.

Table 10: χ^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

6. REAL DATA

As a way to validate the simulation analysis, an empirical study of a real database for the Las Condes - CBD corridor was carried out (Ortúzar and Donoso. 1983). Theis database was chosen for its quality as well as for the fact that it has been broadly studied. The sample consists in 697 observations and 9 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). The correlation structures supposed are presented in Figure 1 and 2. We estimate MNL, Independent Probit, NL, homoscedastic Probit and two Nested Mixed Logit specifications. The first one (HeNML), a heteroscedastic Nested Mixed Logit that considers only one additional term inside the nest; while the second one (HoNML) is a homoscedastic model that considers also independent error terms in the non nested alternatives. The correlation coefficient within each nest can be obtain for the NL through the structural parameter f, while for the Probit it corresponds to a parameter to be estimated. In the case of ML, it can be demonstrated that:

$$\hat{r}_{ML} = \frac{6\hat{s}_m^2}{6\hat{s}_m^2 + p^2} \qquad (24)$$

The results are reported in Tables 11 and 12. It is possible to verify that the parameters do not possess remarkable variations among models. However, it is possible to observe important differences in the correlation estimated by the different models.

Table 11: Nested estructure 1

	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

While with the NL and Probit very low correlation is obtained, the HeNML obtains a considerably high value. This value practically reduces 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 database 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 sufficiently high as to be able to recover in a correct way the covariance structure.

Table 12: Nested estructure 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

The results presented in Table 12 are quite strange, specially the one obtained for HeNML. Nonetheless, this structure has related identification issues, as discussed above.

7. CONCLUSIONS

From our point of view the ML model is an interesting, very flexible and useful modelling alternative, permitting 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 know its properties and limitations 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 that and it can be as general as desired subject to identification restrictions. 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 numeric applications are presented, one based on simulation experiments (including a convergence analysis) and another one with real data, both in a context of similar alternatives, implying a nesting error structure.

It is shown both empirically and theoretically 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. We concluded that the nested structure for the Mixed Logit in theoretical terms always commits homoscedasticity when defining correlation. This could be seen as a problem if you want to compare it with Nested Logit, or as an advantage for the gain in flexibility.

Acknowledgements

This study was partially funded by Fondecyt.

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.

Bhat, C. (2000) Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. **9th International Association for Travel Behaviour Research Conference (IATBR)**, Queensland, Australia

Börsch-Supan, A. And Hajivassilou, V.A. (1993) Smooth unbiased multivariate probability simulators for maximum likelihod 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,** Vol 14A, 367-378.

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

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

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

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

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

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

McFadden, D. and Train, K. (2001) Mixed MNL models for discrete response. **Journal of Applied Econometrics** (in press).

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

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

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.

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

Ortúzar, J. de D. and Willumsen, L.G. (1994) **Modelling Transport**. Segunda Edición, John Wiley and Sons, Chichester.

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**, Vol 9A, 285-344.

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

APPENDIX

Let i be an alternative belonging to nest k. Let us consider the utility of this alternative, using a Nested ML structure: $U_{in} = V_{in} + m_{kn} + e_{in}$, where e_{in} ~ Gumbel(0, l) and $m_{kn} \sim f(0, s_m^2)$. It is easy to see that Var(U_{in}) = $s_m^2 + s_e^2$ and that Cov(U_{in},U_{jn}) = s_m^2 iif $j \in k$. This kind of covariance structure implies the following correlation level:

$$r = \frac{s_m^2}{s_m^2 + s_e^2} \tag{A.1}$$

If $m_{kn} \sim N(0, s_m^2)$, then $m_{kn} = s_{kn} s_m$ with s_{kn} a standard Normal deviate and the estimated parameter shall be so that $\hat{s}_m = l s_m$. Let us consider (A.1) and the relation between the scale parameter and the variance of the Gumbel distribution, then it is direct to demonstrate (24).

$$\hat{s}_m = \frac{p}{\sqrt{6}s_e} s_m \Leftrightarrow \frac{s_e^2}{s_m^2} = \frac{p^2}{6\hat{s}_m^2} \Leftrightarrow \frac{s_m^2}{s_m^2 + s_e^2} = \frac{6\hat{s}_m^2}{6\hat{s}_m^2 + p^2} \tag{A.2}$$

FIGURES

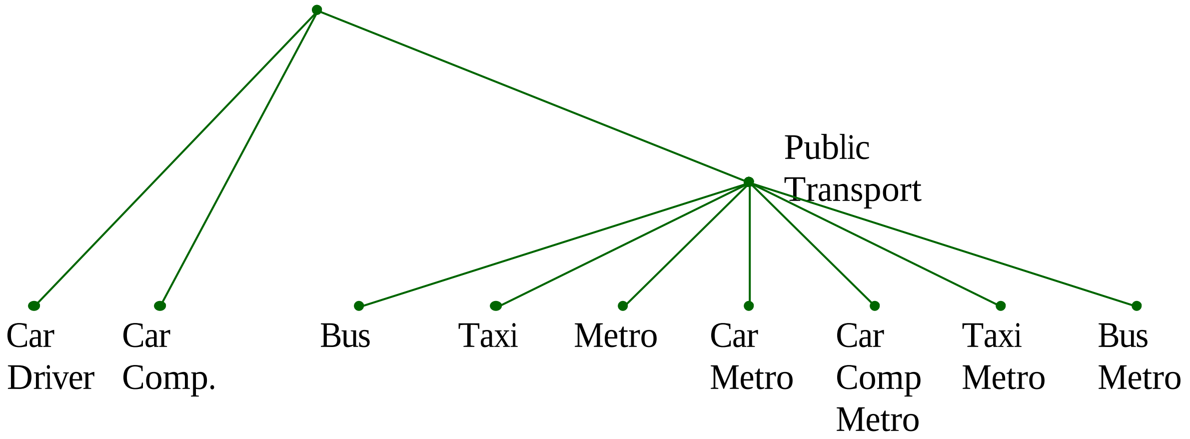


Figure 1: Nested structure 1

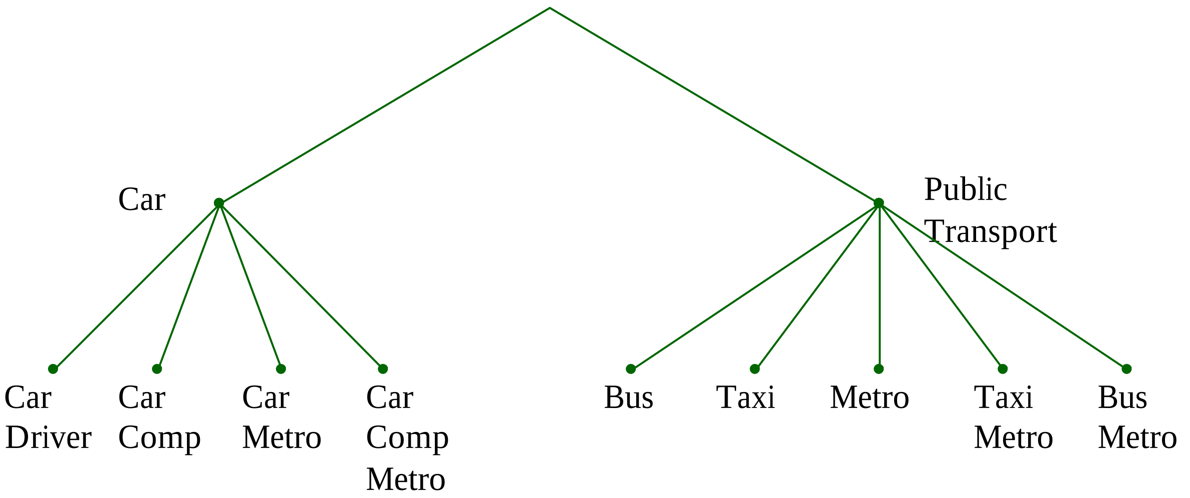


Figure 2: Nested structure 2