Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure*

by
Moshe Ben-Akiva† and Denis Bolduc‡

28 May 1996
PRELIMINARY

Abstract

The Multinomial Probit (MNP) framework is known to suffer from two serious impediments which amplify with increasing number of alternatives. The first is related to the dimensionality of the multifold normal choice probability integrals. Current practice suggests to replace the choice probabilities with easy to compute (asymptotically) unbiased efficient simulators. A unique characteristic of our MNP formulation is the presence of an additional i.i.d. Gumbel error term which makes the Multinomial Logit (MNL) model a special case. The MNL kernel feature leads to a smooth and unbiased MNP choice probability simulator. The second impediment relates to the large number of unknown parameters in the error covariance matrix of situations with large choice sets. This paper suggests different factor analytic specifications to model the inter-dependencies among the

*This work was partially supported by the Social Sciences and Humanities Research Council of Canada and the Fonds FCAR. This paper constitutes a major revision of our original working paper *Multinomial Probit with Autoregressive Error Structure* written a few years ago.

†Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Mass., 02139, U.S.A.

‡Département d'économique, Université laval, Québec, Canada, G1K 7P4.

errors. They enable one to approximate general error covariance structures with parsimonious parametric specifications. For estimation, we employ a simulated maximum likelihood estimation procedure. We test the methodology with two empirical applications. The estimation results and additional computational experiments indicate that our MNP methodology is a useful tool for discrete choice analysis.

1 Introduction

The Multinomial Probit (MNP) model provides the most general framework for inter-dependent alternatives in discrete choice analysis. The inter-dependencies are accounted for through the correlation structure of normally distributed error terms. The primary impediment to the application of the MNP model is related to the dimensionality of the multifold normal choice probability integrals of about the size of the choice set. The earliest solution to the problem has been to ignore the presence of potential inter-dependencies among the alternatives. This is the case with the widely used MultiNomial Logit (MNL) formulation in which the choice probabilities have a closed form that can be calculated easily. The Generalized Extreme-Value (GEV) models, which includes in particular the multinomial logit and the nested logit, permit inter-dependencies with closed form expressions for the choice probabilities. However, in this class of models, the inter-dependencies are captured in a very limited way.

Recent solutions to the MNP dimensionality problem replace the multifold normal integrals with smooth (asymptotically) unbiased efficient simulators computed from an underlying latent variable model. To use these probability simulators in a standard maximum likelihood framework leads to the Simulated Maximum Likelihood (SML). An alternative to the SML setting is the Method of Simulated Moments (MSM) proposed by McFadden (1989) and Pakes and Pollard (1989). MSM is often favored over SML because a given level of accuracy in model parameter estimation can be obtained with a fairly small number of replication draws. The accuracy of the SML methodology critically depends on using a large number of simulation draws because the log-likelihood function is simulated with a non-negligible downward bias. For computational reasons, in this paper, we still stick to the SML approach, mainly on the ground that the SML requires the computation of the probability of only the chosen alternative, while MSM needs all choice probabilities. With large choice sets, which is often the situation in transportation, this factor can be quite important. Our numerous experiments indicated that with a smooth unbiased simulator, to use a few hundred conventional draws is sufficient to make this bias negligible. Also, using antithetic draws significantly reduces the number of draws required to obtain a comparable level of accuracy. Another reason for our choice is that the objective function associated with SML is numerically better behaved than the MSM objective function.

A unique feature of our Logit kernel MNP formulation is the presence of an additional independently and identically distributed (i.i.d.) Gumbel error term which makes the MNL model a special case obtained by imposing restrictions on specific parameters of the error structure. The simulator that we propose for the MNP probabilities is unbiased, smooth and efficient. It arises very naturally in our suggested framework and it is based on a mixture of MNL kernels computed from standard normal distributions. To estimate the model parameters, we replace the choice probabilities in a standard Maximum Likelihood algorithm with these smooth MNL kernel simulators.

A second complication with implementing MNP concerns the number of unknown parameters to estimate in the error covariance matrix. Except for limiting cases such as those implied by equicorrelated errors or by GEV models, for example, the number of these parameters usually increases exponentially with the number of alternatives. In models with large number of alternatives, this problem can seriously limit the usefulness of MNP. To tackle this parameter estimation problem, we suggest several factor based error structure specifications. These frameworks allow one to capture fairly general error covariance structures using parsimonious parametric specifications. The main advantage of these approaches is that the number of underlying parameters increase quite independently with the size of the choice set.

The paper is organized as follows. In section 2, we introduce the logit kernel MNP model. In section 4, we discuss several interesting special cases all based on a factor analytic representation of the error covariance structure. Subsequently, we focus on the implementation of the proposed approach in the (simulated) maximum likelihood context. In the final section of the paper, we test our methodology with two empirical applications. The first one involves a trinomial choice based on synthetic data and the second one concerns a choice among five residential telephone service options using survey data collected to predict residential telephone demand.

2 The Model

2.1 The Discrete Choice Model

Consider the following discrete choice model. For a given individual $n, n = 1, \ldots, N$ and an alternative $i, i = 1, \ldots, J_n$ where J_n is the number of alternatives in the choice set C_n specific to individual n, the model is written as:

$$\begin{aligned} y_{in} &= \begin{cases} 1 & \text{if } U_{in} \geq U_{jn}, \text{ for } j = 1, \ldots, J_n \\ 0 & \text{otherwise, and} \end{cases} \\ U_{in} &= X_{in}\beta + \varepsilon_{in}, \end{aligned} \tag{1}$$

where y_{in} indicates the observed choice, U_{in} is the utility of alternative i as perceived by individual n, X_{in} is a $(1 \times K)$ vector of attributes describing individual n and alternative i. This vector can contain an alternative specific dummy as well as generic and alternative specific variables. It is well-known that one of the alternative specific constants cannot be estimated. As usual, we drop form the model the constant associated with the last alternative. Any alternative could be selected as the referent without affecting the results. β is a $(K \times 1)$ vector of fixed coefficients and ε_{in} is a random disturbance.

2.2 The Logit Kernel MNP Model

Distributional assumptions on the error term
The ε_{in} random utility term is defined as follows:

$$\varepsilon_{in} = \xi_{in} + \nu_{in} \tag{2}$$

where:

- ξ_{in} captures the interdependencies among the alternatives. It is assumed to arise from a multivariate normal distribution.

- ν_{in} is an i.i.d. standard Gumbel random variate.

Remark 1 *Given $\xi_n = (\xi_{1n}, \ldots, \xi_{J_n n})'$, the model corresponds to a multinomial logit formulation as follows:*

$$\Lambda(i \mid C_n, \xi_n) = \frac{e^{X_{in}\beta + \xi_{in}}}{\sum\limits_{j \in C_n} e^{X_{jn}\beta + \xi_{jn}}}, \tag{3}$$

where $\Lambda(i \mid C_n, \xi_n)$ is the probability that the choice is i when the ξ_{jn}'s are known.

2.3 The Model in Vector Form

In a more compact vector form, the model can be written as follows:

$$y_n = [y_{1n}, ..., y_{J_n n}]', \tag{4}$$

$$U_n = X_n \beta + \varepsilon_n, \tag{5}$$

where y_n, U_n and ε_n are $(J_n \times 1)$ vectors and X_n is a $(J_n \times K)$ matrix. The error vector is decomposed as follows:

$$\varepsilon_n = \xi_n + \nu_n, \tag{6}$$

where by assumption, $\xi_n \sim N(0, \Xi_n)$, which implies that the covariance matrix can potentially change across observations.

2.4 Generalized Factor Analytical Representation

To capture the interdependencies among the alternatives, we use a factor analytic structure:

$$\xi_n = F_n \zeta_n, \tag{7}$$

where:

- ζ_n is a $(M \times 1)$ vector of factors. By assumption, the factors are i.i.d. standard normal distributed terms. The number of factors can be less, equal or greater than the number of alternatives.

- F_n is a $(J_n \times M)$ matrix of loadings that map the factors to the ξ_n error vector.

Remark 2 *The error covariance structure captured by (7) is very general. In the most flexible situation, ζ_n contains $J = \max_n J_n$ factors and F_n is a $(J_n \times J)$ matrix obtained from a $(J \times J)$ lower triangular Cholesky matrix $\mathbf{L}_n$ in removing the rows associated with the unavailable alternatives. $\mathbf{L}_n$ is the matrix such that $\mathbf{L}_n \mathbf{L}_n' = \Sigma_n$, where Σ_n describes the covariance structure among the J possible error terms.*

Remark 3 *Different interesting special cases covered by our model formulation are discussed in the next section.*

Substituting Equations (6) and (7) into Equation (5), finally gives:

$$U_n = X_n\beta + F_n\zeta_n + \nu_n, \tag{8}$$

where, under the assumptions made, $E(U_n) = X_n\beta$ and

$$cov(U_n) = F_n F_n' + g\, I_{J_n}, \tag{9}$$

where $g = \pi^2/6$ is the variance of a standard Gumbel random variable. As it should be expected, the estimation of such model brings some identification issues that need to be discussed. This is especially true and critical when the error process is decomposed as in Equation (8). Our strategy will be to discuss this issue in each one of the special cases that we consider.

3 Response Probabilities

Define $P(i \mid C_n)$ as the probability of drawing a latent vector U_n with $U_{in} > U_{jn}$ for $j \in C_n$ given X_n. Conditional on ζ_n, Equation (8) defines a standard MNL model since ν_n is i.i.d. Gumbel. Therefore, the unconditional choice probability can be expressed as:

$$P(i \mid C_n) = \int_\zeta \Lambda(i \mid C_n, \zeta)\, n(\zeta, I_M)\, d\zeta \tag{10}$$

where

$$\Lambda(i \mid C_n, \zeta) = \frac{e^{X_{in}\beta + F_{in}\zeta}}{\sum\limits_{j \in C_n} e^{X_{jn}\beta + F_{jn}\zeta}}, \tag{11}$$

is the MNL choice probability of alternative i conditional on ζ. Also $n(a, A)$ denotes the value at a multivariate of normal density centered at zero with covariance matrix A.

Remark 4 *For computational reasons, we will replace $P(i \mid C_n)$ with an unbiased simulator that we compute as:*

$$g(i \mid C_n) = \frac{1}{R} \sum_{r=1}^{R} \Lambda(i \mid C_n, \zeta_r), \tag{12}$$

where ζ_r denotes draw r of ζ from a $N(0, I_M)$.

4 Special Cases

Many interesting cases can be embedded in the general specification introduced above.

4.1 Heteroscedastic

This particular case is obtained when the F_n matrix in equation (7) is diagonal. For later reference, we use the notation T_n to designate this matrix. Therefore, in the heteroscedastic case, the model is written as:

$$U_n = X_n\beta + T_n\zeta_n + \nu_n, \tag{13}$$

where T_n is a $(J_n \times J_n)$ diagonal matrix with the standard deviations associated with the J_n alternatives available to individual n. Using a scalar notation, we would write:

$$U_{in} = X_{in}\beta + \sigma_i\zeta_{in} + \nu_{in}, \qquad i \in C_n. \tag{14}$$

Identification

To address the identification issue, we consider the following trinomial example with universal choice set:

$$\begin{aligned} U_{1n} &= \alpha_1 + x_{1n}\gamma + \sigma_1\zeta_{1n} + \nu_{1n} \\ U_{2n} &= \alpha_2 + x_{2n}\gamma + \sigma_2\zeta_{2n} + \nu_{2n} \\ U_{3n} &= \quad + x_{3n}\gamma + \sigma_3\zeta_{3n} + \nu_{3n}, \end{aligned} \tag{15}$$

where the notation should be rather obvious. To identify the model parameters, a given normalization has to be made. In order for the MNL to be a case nested within the general formulation, we set one of the σ_j's to zero.

Problem: There is, in practice, no prior knowledge of $\arg\min_j \sigma_j$, i.e. the minimum variance alternative. With the help of empirical examples, we will come out with intuitively acceptable procedures to identify the good candidates for this normalization.[1]

[1] It is important to mention that the problem that is raised here has been completely ignored in the previous applications that we could review. State a few of them here.

4.2 Factor Analytic

To keep a uniform notation for all sub-cases considered, we write this model as:

$$\xi_n = Q_n T \zeta_n, \tag{16}$$

where Q_n is $(J_n \times M)$ is a matrix of factor loadings and T is the diagonal matrix that contains the factor specific standard deviations. In scalar notation, this gives

$$\xi_{in} = \sum_{m=1}^{M} q_{imn} \sigma_m \zeta_{mn}, \qquad i \in C_n. \tag{17}$$

In order to fix the scale, since the q's are estimated, we make the following normalizations: $\sum_i q_{imn}^2 = 1, \forall i$. We consider two possible types of factor decomposition:

- **exploratory:** $M < J$ and no exclusion restriction on Q_n.

  Identification

  In that case, there should be no identification problem. If $M = J$ then this situation is similar to the heteroscedastic case considered above.

- **confirmatory:** Restrictions are imposed on Q_n by *a priori* setting some q_{imn}'s to zero.

  Identification

  In that case, there should be no identification problem. If $M = J$ then this situation is similar to the heteroscedastic case considered above. An important candidate within the class of confirmatory factor analysis models is the following error component formulation.

The Error Component Formulation

In the error component version, the ξ_n error vector in Equation (7) is assumed to be generated as:

$$\xi_n = G_n T \zeta_n \tag{18}$$

where G_n is $(J_n \times M)$ and T is a $(M \times M)$ diagonal matrix. The difference with the confirmatory factor analytic structure is that the G_n matrix is known while the elements of Q_n have to be estimated. In scalar notation, we write

$$\xi_{in} = \sum_{m=1}^{M} g_{imn}\sigma_m\zeta_{mn}, \qquad i \in C_n, \tag{19}$$

where:

$$g_{imn} = \begin{cases} 1 & \text{if the } m^{th} \text{ element of } \zeta_n \text{ applies to altenative } i \text{ for individual } n \\ 0 & \text{otherwise} \end{cases}$$

Identification

> If $M < J$ the problem does not exist. If $M = J$ then this situation is similar to the heteroscedastic case considered above.

4.3 General Autoregressive Process

In this case, the ξ_n error vector in Equation (7) has the following form:

$$\xi_n = P_n T_n \zeta_n, \tag{20}$$

where $P_n = (I - \rho W_n)^{-1}$. This structure results from assuming that ξ_n is generated from the following first-order autoregressive process:

$$\xi_n = \rho W_n \xi_n + T_n \zeta_n, \qquad \zeta_n \sim N(0, I_{J_n}),$$

where W_n is a $(J_n \times J_n)$ matrix of weights $w_{ij,n}$ describing the influence of each ξ_{jn} error upon the others. Using a general notation, we write it as

$$\begin{aligned} w_{ij,n} &= \frac{w^*_{ij,n}}{\sum_{k=1}^{J_n} w^*_{ik,n}}, \ \forall j \neq i \text{ and } w_{ij,n} = 0 \quad \forall i, \\ w^*_{ij,n} &= g(R^1_{ij,n}, R^2_{ij,n}, ..., R^L_{ij,n}, \theta_1, \theta_2, ..., \theta_H), \ \forall j \neq i \text{ and } w^*_{ij,n} = 0 \ \forall i, \end{aligned} \tag{21}$$

where there are H parameters θ_h and L variables $R^l_{ij,n}$ used to describe the correlation structure in effect. Spatial studies that use Spatial AutoRegressive of order 1(SAR(1)) error processes often define the contiguity structure through a Boolean contiguity matrix. In that case, $w^*_{ij} = 1$ if i and j are contiguous and $w^*_{ij} = 0$ otherwise. See, for instance, Anselin (1989), Cliff and Ord (1981), Blommestein (1983) or Bivand (1984), for more details. For a recent application of SAR(1) processes in economics, see Case (1991).

The normalization used in Equation (21) ensures that the process is stable for values of ρ in the $(-1, 1)$ interval. The interpretation and the sign of ρ, usually referred to as the correlation coefficient, depend on the definition of proximity embodied in w^*_{ij}. For the specification just discussed, a $\rho > 0$ implies that errors of the same sign are grouped together. In practice, the parameters in $w^*_{ij,n}$ could be estimated. However, in the applications considered below, a Boolean structure will be postulated.

Identification

It is technically related to the error component case.

4.4 MNL with Normally Distributed Random Coefficients

The MNL formulation with normally distributed random taste coefficients can be written as:

$$U_n = X_n \beta_n + v_n, \tag{22}$$

where β_n is a K-dimensional random normal vector with mean vector β and covariance matrix Ω. Given the notation and assumptions previously used and replacing β_n with the equivalent relationship: $\beta_n = \beta + \Gamma\zeta_n$, where Γ is the lower triangular Cholesky matrix such that $\Gamma\Gamma' = \Omega$, one obtains Equation (8) with $F_n = X_n\Gamma$. The parameters that need to be estimated in this model are those present in β and in Γ. We now consider the identification issue.

Identification

> If there are alternative specific constants included in the X_n matrix, the identification problem raised above is going to occur.

The other special cases that we have considered assume that the randomness in the coefficients only arise from the alternative specific constants. This assumption was made so that we could better focus on the identification matters.

5 Parameter Estimation

We now describe the method that we use to estimate the joint vector of parameters δ which contains the β vector and those parameters to be estimated in the error structure. We include the last set of parameters into ψ, a $(p \times 1)$ vector. Therefore, the parameter estimation concerns the vector $\delta = (\beta', \psi')'$. In the heteroscedastic model, only the alternative specific standard errors are included in ψ. In the GAR(1) version based on a Boolean contiguity matrix, the same standard error are estimated in addition to the ρ correlation coefficient. The factor analytic and the MNL version with random coefficients are the specifications that can potentially have a very large number of unknown parameters.

5.1 Maximum likelihood

The log likelihood of the sample is

$$L(\delta) = \sum_{n=1}^{N} \ln P(i_n \mid \delta, C_n), \tag{23}$$

where $P(i_n \mid \delta, C_n)$ is the probability associated with the choice made by individual n. The score vector is:

$$\frac{\partial L(\delta)}{\partial \delta} = \sum_{n=1}^{N} \frac{1}{P(i_n \mid \delta, C_n)} \frac{\partial P(i_n \mid \delta, C_n)}{\partial \delta}. \tag{24}$$

Since we focus on the model formulated with the MNL kernel given in Equations (10) and (11), the score can therefore be computed as follows:

$$\frac{\partial L(\delta)}{\partial \delta} = \sum_{n=1}^{N} \frac{1}{P(i_n \mid \delta, C_n)} \int_a \Lambda(i_n \mid \delta, C_n, \zeta) \frac{\partial \ln \Lambda(i_n \mid \delta, C_n, \zeta)}{\partial \delta} n(\zeta, I_M)\, d\zeta, \tag{25}$$

where

$$\ln \Lambda(i_n \mid \delta, C_n, \zeta) = X_{in}\beta + F_{in}\zeta - \ln \sum_{j \in C_n} e^{X_{jn}\beta + F_{jn}\zeta},$$

and where F_n and ζ were defined in Equation (7).

- Iterative estimation method

To simplify the calculations, we suggest to use the BHHH technique which makes use of the first derivatives only. The direction matrix that we call $\mathbf{R}$ is therefore computed as follows:

$$\mathbf{R} = \sum_{n=1}^{N} Z_{in} Z'_{in}, \tag{26}$$

where we define $Z_{in} = P(i_n \mid \delta, C_n)^{-1} \partial P(i_n \mid \delta, C_n)/\partial \delta$. For practical purposes, after convergence, we use $\mathbf{R}_{ML}^{-1}$($\mathbf{R}^{-1}$ evaluated at δ_{ML}, the ML estimator of δ) as the asymptotic variance-covariance matrix of δ_{ML}. Under the usual regularity conditions which are satisfied with identified parametric structures, we can conclude that asymptotically $\delta_{ML} \sim N(\delta, \mathbf{R}_{ML}^{-1})$.

Unless the dimension of ζ is small (≤ 3), the Maximum Likelihood (ML) estimator δ_{ML} just described, cannot be computed in a reasonable amount of time. For models with ζ of larger dimension, we suggest to use the Simulated Maximum Likelihood (SML) methodology.

5.2 Simulated Maximum Likelihood

When high dimensional integrals are involved, one solution to the computational limitations is to replace the hard to compute probability related functions with efficiently simulated ones. As seen before, the log-likelihood and the score vector are expressed as normal mixtures of MNL probabilities and MNL derivatives.

The SML approach replaces the response probability for alternative i with:

$$\hat{g}\,(i \mid \delta, C_n) = \frac{1}{R} \sum_{r=1}^{R} \hat{\Lambda}(i \mid \delta, C_n, \zeta_{nr}), \tag{27}$$

an unbiased simulator where ζ_{nr} denotes draw r of a random vector from $N(0, I_M)$. Similarly Z_{in} is substituted by

$$\hat{Z}_{in} = \frac{1}{\hat{g}\,(i \mid \delta, C_n)} \frac{1}{R} \sum_{r=1}^{R} \hat{\Lambda}(i \mid \delta, C_n, \zeta_{nr}) \frac{\partial \ln \hat{\Lambda}(i \mid \delta, C_n, \zeta_{nr})}{\partial \delta}, \tag{28}$$

where $\hat{g}(i \mid \delta, C_n)$ is computed using (27). In the iterative estimation process, one replaces the $\mathbf{R}$ matrix defined in (26) with its simulated counterpart that we denote by $\hat{\mathbf{R}}_{SML}$. Also, the asymptotic results for the joint vector of estimated parameters is $\hat{\delta} \sim N(\delta, \hat{\mathbf{R}}_{SML}^{-1})$.

A well-known result previously obtained in Börsch-Supan and Hajivassiliou (1993), among others, indicates that the likelihood function is simulated with a downward bias. As noted in Remark 4, $g(i \mid C_n)$ provides an unbiased simulator for $P(i \mid C_n)$. A second degree Taylor's expansion of $\ln(g(i \mid C_n))$ around $P(i \mid C_n)$, gives:

$$\ln(g(i \mid C_n)) \approx \ln(P(i \mid C_n)) \quad + \frac{1}{P(i \mid C_n)}(g(i \mid C_n) - P(i \mid C_n)) - \frac{1}{2P(i \mid C_n)^2}(g(i \mid C_n) - P(i \mid C_n))^2.$$

Taking the expected value of this relationship therefore implies that :

$$\hat{L}(\delta) - L(\delta) \approx -\frac{var(g(i \mid \delta, C_n))}{(2P(i \mid \delta, C_n)^2)} \leq 0,$$

which suggests that in order to minimize the bias in simulating the log-likelihood function, it is important to simulate the probabilities with good precision. For a fixed number of replications, to use a variance reduction technique such as drawing variates with an antithetic property is advisable. Our experiments indicate that a simulator with 250 standard draws or 100 antithetic draws is large enough to produce a negligible simulation bias.

6 Applications

In this section, we consider two applications, one based on synthetic data and a second one on real data. The first sample concerns an hypothetical choice situation among three alternatives. The focus is on the parameter identification problems that may occur in the different versions of the Logit kernel MNP considered above. The second application is based on a subset of a survey collected to predict residential telephone demand. Five different telephone service options are considered.

6.1 A Trinomial Choice Situation

The first application concerns an hypothetical choice situation among three alternatives. The model specification that we use is as displayed in Equation

- Simulation for looking the minimum variance in the heteroskedastic model

(15) that we repeat here for convenience. The postulated true parameter values are : $\alpha_1 = \alpha_2 = 1, \gamma = -1, \sigma_1 = 0, \sigma_2 = 1, \sigma_3 = 2$.

$$\begin{aligned} U_{1n} &= \alpha_1 + x_{1n}\gamma + \sigma_1\zeta_{1n} + \nu_{1n} \\ U_{2n} &= \alpha_2 + x_{2n}\gamma + \sigma_2\zeta_{2n} + \nu_{2n} \\ U_{3n} &= \quad + x_{3n}\gamma + \sigma_3\zeta_{3n} + \nu_{3n}, \end{aligned} \tag{29}$$

In generating the data for estimation, we consider the following sample sizes: $N = 500, 1000, 4000$.

Table 1 Estimation results: Different Structures

		Unidentified Structures			Selected	MNL
	true	R=1000	R=5000	R=2000	R=2000	
Parameter	value	N=500	N=1000	N=4000	N=4000	N=4000
α_1	1	2.76	2.04	1.49	0.98	0.89
		(0.28)	(0.24)	(1.16)	(13.34)	(18.84)
α_2	1	1.87	2.16	1.33	0.88	0.91
		(0.27)	(0.24)	(1.15)	(10.03)	(19.39)
γ	-1	-3.89	-1.72	-1.41	-1.00	-.67
		(-0.28)	(-0.25)	(-1.24)	(-7.15)	(-25.82)
σ_1	0	2.41	2.72	1.52	0	
		(0.21)	(0.20)	(0.65)		
σ_2	1	6.30	3.01	1.95	1.17	
		(0.26)	(0.21)	(0.88)	(2.17)	
σ_3	2	8.33	2.81	3.23	2.10	
		(0.27)	(0.21)	(1.05)	(4.56)	
(Simul.) Log.-Likelihood		-452.22	-934.69	-3764.68	-3765.37	-3778.53
Number of iterations		13	10	9	4	5

The estimation is performed using large number of draws.

6.2 An Empirical Application

In this section, we test our methodology with an application to residential telephone demand analysis. The model involves a choice among five residential telephone service options for local calling. A household survey was conducted in 1984 for a telephone company and was used to develop a comprehensive model system to predict residential telephone demand (Train et al., 1987). Below we use part of the data to estimate a model that explicitly account for inter-dependencies between residential telephone service options using some of the different error structure introduced above. We first describe the data. Then we compare the estimation results obtained.

- The data

Local telephone service typically involves the choice between flat (i.e. a fixed monthly charge for unlimited calls within a specified geographical area) and measured (i.e. a reduced fixed monthly charge for a limited number of calls plus usage charges for additional calls) services. In the current application, five services are involved, two measured and three flat. They can be described as follows:

1. Budget measured - no fixed monthly charge; usage apply to each call made.
2. Standard measures - a fixed monthly charge covers up to a specified dollar amount (greater that the fixed charge) of local calling, after which usage charges apply to each call made.
3. Local flat - a greater monthly charge that may depend upon residential location; unlimited free calling within local calling area; usage charges apply to calls made outside local calling area.
4. Extended area flat - a further increase in the fixed monthly charge to permit unlimited free calling within an extended area.
5. Metro area flat - the greatest fixed monthly charge that permits unlimited free calling within the entire metropolitan area.

The sample concerns 434 households. The availability of the service options of a given household depends on its geographical location. Details are provided in Table 2. In Table 3, we summarize the service option availabilities over the usable sample.

Table 2: Availability of Service Options

Service Options	Geographical Location		
	Metropolitan Areas	Perimeter Exchanges Adjacent to Metro Areas	All Other
Budget Measured	Yes	Yes	Yes
Standard Measured	Yes	Yes	Yes
Local Flat	Yes	Yes	Yes
Extended Flat	No	Yes	No
Metro Flat	Yes	Yes	No

Table 3: Summary Statistics on Availability of Service Options

Service Options	Chosen	Percent	Tot. Avail
Budget Measured	73	0.168	434
Standard Measured	123	0.283	434
Local Flat	178	0.410	434
Extended Flat	3	0.007	13
Metro Flat	57	0.131	280
Total :	434	1.000	1595

The model that we use in the present analysis is intentionally specified to be simple. The explanatory variables used to explain the choice between the five service options are four alternative specific constants which correspond to the first four service options and a generic cost variable which is defined as the natural log of the monthly cost of the different service options expressed in dollars.

The results are displayed in Table 4:

Table 4: Heteroscedastic Case: Telephone Data *

	Unidentified Structures				Selected	MNL
	R=150	R=250	R=1000	R=5000	R=1000	
Parameter	N=434	N=434	N=434	N=434	N=434	N=434
Altern. specific constants						
Budget Measured (1)	-3.44	-3.37	-3.32	-44.2	-3.38	-2.46
	(-4.27)	(-5.30)	(-4.20)	(n.a.)	(-3.62)	(-7.84)
Standard Measured (2)	-2.69	-2.63	-2.56	-34.2	-2.60	-1.74
	(-3.83)	(-4.20)	(-3.77)	(n.a.)	(-3.50)	(-6.28)
Local Flat (3)	-1.49	-1.39	-1.37	-18.4	-1.40	-0.54
	(-2.77)	(-2.83)	(-2.71)	(n.a.)	(-2.60)	(-2.57)
Extended Flat (4)	-1.11	-1.07	-1.09	-13.6	-1.07	-0.74
	(-0.56)	(-0.81)	(-0.58)	(n.a.)	(-0.99)	(-1.02)
Log Cost	-2.78	-2.77	-2.70	-37.9	-2.73	-2.03
	(-4.58)	(-4.56)	(-5.45)	(n.a.)	(-5.09)	(-9.47)
σ_1	0.29	0.05	0.23	17.6	0.38	
	(0.21)	(0.02)	(0.09)	(n.a.)	(0.19)	
σ_2	0.63	0.53	0.35	17.0	0.39	
	(0.57)	(0.41)	(0.16)	(n.a.)	(0.19)	
σ_3	3.06	2.99	2.83	48.3	2.95	
	(2.52)	(2.49)	(2.65)	(n.a.)	(2.59)	
σ_4	0.37	0.39	0.44	16.2	0.08	
	(0.03)	(0.04)	(0.04)	(n.a.)	(0.02)	
σ_5	0.18	0.34	0.01	22.2	0	
	(0.07)	(0.15)	(0.01)	(n.a.)		
Simulated Log Likelihood	-471.39	-469.70	-471.14	-470.49	-470.75	-477.56
Number of iterations	12	9	12	44	10	9

*Asymptotic t-statistics are in parenthesis.

According to the results obtained in estimating unidentified structures, it seems reasonable to conclude that, the base alternative could be either one of alternatives 1,2,4 or 5. The estimation performed based on 5000 replications indicates that alternative 4 should be selected as the base. As it should be expected, with very large number of simulation draws, the hessian should be singular in situations with unidentified error structures. With R=1000, alternative 5 clearly appears to be the base and since to use the last alternative as a reference is quite usual, this is the structure that we select. The T matrix that corresponds with this structure is as follows:

$$T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Table 4 also gives the estimation results obtained for the MNL. The log likelihood of the MNL model is -477.56 and there is an obvious gain to incorporate heteroscedasticity in the model.

In Table 5, the estimation results of three alternative error component specifications are provided. In column 2, a two factor case is considered where factor one applies to alternatives 1 and 2, factor 2 to alternatives 3 and 4. The next two columns refer to a three factor structure. The second column uses the same structure as in the first one except that the extra factor applies to the last alternative. As expected, based on the heteroscedastic estimates previously obtained as presented in Table 4, the estimated standard error associated with this alternative specific factor is found to be not significantly different from zero. The last column corresponds to a structure where alternative three, the one with apparently the highest variability, is associated with a specific factor. Alternatives 1 and 2 share the first factor and alternatives 4 and 5 the third factor. Clearly, this structure is the one to favor.

Table 5: Error Component Case: Telephone Data*

	Two Factors	Three Factors	
	R=1000	R=1000	R=1000
Parameter	N=434	N=434	N=434
Altern. specific constants			
Budget Measured (1)	-3.62	-3.62	-3.76
	(-5.14)	(-4.84)	(-4.80)
Standard Measured (2)	-2.85	-2.84	-2.97
	(-4.38)	(-4.07)	(-4.02)
Local Flat (3)	-1.47	-1.47	-1.61
	(-2.84)	(-2.56)	(-2.52)
Extended Flat (4)	-1.52	-1.51	-1.19
	(-1.62)	(-1.56)	(-1.09)
Log Cost	-3.04	-3.06	-3.20
	(-4.98)	(-4.89)	(-5.02)
σ_1	1.31	1.30	1.46
	(1.31)	(0.99)	(1.05)
σ_2	1.31	1.30	1.46
	(1.31)	(0.99)	(1.05)
σ_3	2.98	3.02	3.36
	(2.62)	(2.56)	(2.57)
σ_4	2.98	3.02	0.57
	(2.62)	(2.56)	(0.21)
σ_5	0	0.35	0.57
		(0.11)	(0.21)
Simulated Log Likelihood	-470.91	-470.87	-470.38
Number of iterations	9	11	14

*Asymptotic t-statistics are in parenthesis.

7 Conclusion

This paper developed a MNP formulation that permits a parsimonious specification of the covariance structure among the alternatives and contains MNL as a special case. The methodology is most useful in situations with large choice sets which require the use of probability simulators. The estimation method used is simulated maximum likelihood that utilizes modern simulators with efficient properties. A first example based on synthetic data and also a real empirical case study which involves a choice among five residential telephone service options were used to demonstrate the feasibility and desirability of the proposed approach. Clearly, more testing and experience with other applications are needed to further demonstrate the applicability of the proposed model.

References

[1] Anselin, L. 1989, "Spatial Econometrics: Methods and Models", Kluwer Academic Publishers.

[2] Ben-Akiva, M. and D. Bolduc, 1989, "Multinomial Probit with Taste Variation and Autoregressive Alternatives", *69th Annual Meeting, Transportation Research Board*, Washington.

[3] Ben-Akiva, M. and D. Bolduc,1991, "Multinomial Probit with Autoregressive Error Structure", Département d'économique, Université Laval, Cahier 9123.

[4] Ben-Akiva, M.and S. Lerman, 1985, "Discrete Choice Analysis", Cambridge, MA, MIT Press.

[5] Berkovec, J. and S. Stern, 1991, "Job Exit Behavior of Older Men", *Econometrica*, Vol. 59, No. 1, January 1991, 189-210.

[6] Bivand, R.S., 1984, "Regression Modelling with Spatial Dependence: an Application of Some Class Selection and Estimation Methods", *Geographical Analysis*, 16, 1 (January),25-37.

[7] Blommestein, H.J., 1983, "Specification and Estimation of Spatial Econometric Models. A Discussion of Alternative Strategies for Spatial Economic Modeling", *Regional Science and Urban Economics*, 13, 251-270.

[8] Bolduc, D., 1987, "On the Estimation of Linear Models with Generalized SAR(1) Error Processes", Département d'économique, Université Laval.

[9] Bolduc, D., 1992, "Generalized Autoregressive Errors in the Multinomial Probit Model", *Transportation Research B - Methodological,* Vol. 26B, No. 2, 155-170.

[10] Bolduc, D., 1994, "Estimation of Multinomial Probit Models using Maximum Simulated Likelihood with Analytical Gradient and the GHK Choice Probability Simulator", working paper, CREST, Paris.

[11] Börsch-Supan, A. and V. Hajivassiliou, 1993, "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models", *Journal of Econometrics,* 58, 347-368.

[12] Bunch, D.A.., 1991, "Estimability in the Multinomial Probit Model", *Transportation Research B- Methodological,* Vol. 25B, 1-12.

[13] Case, A., 1991, "Spatial Correlation ini Household Demand", *Econometrica.*

[14] Cliff A.D, and J.K. Ord, 1981, "Spatial Processes, Models and Application", Pion, London.

[15] Daganzo, C., 1979, "Multinomial Probit", New-York: Academic Press.

[16] Hausman, J.A. and D.A. Wise, 1978, "A Conditional Probit Model for Qualitative Choice Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica*, 46, 403-426.

[17] Hordijk, L., 1974, "Spatial Correlation in the Disturbances of a Linear Interregional Model", *Regional and Urban Economics,* 4, 117-140.

[18] Hotz, v.J. and S. Sanders, 1990, "The Estiation of Stochastic Dynamic Discrete Choice Models by the Method of Simulated Moments", D.P. University of Chicago.

[19] Lerman, S., and C. Manki, 1981, "On the Use of Simulated Frequencies to Approximate Choice Probabilities", in C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge Mass.: MIT Press.

[20] Magnus, J.R. and H. Neudecker, 1988, "Matrix Differential Calculus with Applications in Statistics and Econometrics", *Wiley series in probability and mathematical statistics.*

[21] McFadden, D., 1989, "A Method of Simulated Moments for Estimatin of the Multinomial Probit without Numerical Integration", Econometrica, 57.

[22] Pakes, A. and D. Pollard, 1989, "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57.

[23] Train, K., D. McFadden and M. Ben-Akiva, 1987, "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices", *Rand Journal of Economics*, 18,1.