

Luisa Scaccia*

**RANDOM PARAMETERS LOGIT MODELS
APPLIED TO PUBLIC TRANSPORT DEMAND**

Abstract

The multinomial logit model has been used widely as a fundamental tool for the analysis of discrete choices and has found large application in transport studies. However, its restrictive assumptions, such as independence from irrelevant alternatives (IIA) and preference homogeneity across respondents, have motivated the development of more flexible model structures that allow for an increasingly realistic representation of travel behaviour. Among these, a primary role is played by random parameter models. This paper proposes a comparison between two different specifications of random parameter models, namely the mixed logit and the discrete mixture model. An application to public transport demand is illustrated.

JEL CLASSIFICATION: C25; C23; C51

KEYWORDS: DISCRETE CHOICE; DISCRETE AND CONTINUOUS DISTRIBUTIONS; MIXTURE MODELS; RANDOM PARAMETERS.

1. Introduction

Over the past thirty years, the area of travel behaviour research has made vast usage of discrete choice models belonging to the family of Random Utility Models (RUM). For a long time, the high cost of estimating advanced models meant that most applications were limited to the use of the most basic model structures, such as multinomial

* Dip. di Istituzioni Economiche e Finanziarie, Università di Macerata, via Crescimbeni 20, 62100 Macerata, Italy, Tel. +39 0733 258 3242, e-mail: scaccia@unime.it.

and nested Logit. In particular, the Multinomial Logit (MNL) model, or conditional logit model (Daniel McFadden, 1974), possesses many advantages in terms of closed-form solution and simplicity of interpretation and use.

However, the MNL model makes some restrictive assumptions such as independence from irrelevant alternatives (IIA) and preference homogeneity across respondents. Such drawbacks have led to increasing dissatisfaction with the MNL approach. Partly as a response to the perceived weaknesses of the MNL model, partly as a result of the gains in computing power and estimation techniques, Random Parameters Logit (RPL) models have grown in popularity with discrete choice modellers (Kenneth Train, 1998, 2003; McFadden and Train, 2000) over the past ten years. In this approach, the utility of each individual is a function of transport attributes (and, eventually, of individual socioeconomic characteristics) with attribute coefficients that are random and reflect individual preferences. The distribution of the coefficients is generally supposed to be continuous, leading to the so called Mixed Multinomial Logit (MMNL) model. Less frequently, a discrete distribution is used to approximate the real underlying distribution of the random parameters. The resulting Discrete Mixture (DM) model, or mass point mixed logit model, is a particular case of the Latent Class (LC) model (Wagner Kamakura and Gary Russell, 1989; Peter Boxall and Wiktor Adamowicz, 2002). LC models, which have been used in the context of transport studies by, for example, William Greene and David Hensher (2003) and Back Jin Lee et al. (2003), capture taste heterogeneity by assuming that the underlying distribution of tastes can be represented by a discrete distribution, with a small number of mass points that can be interpreted as different classes or segments of individuals. The probability of an individual being assigned to a specific class is modelled as a function of attributes of the respondent and possibly of the alternatives. In DM models, instead, these allocation probabilities are independent of explanatory variables and are simply given by constants that are to be estimated.

Compared to the MMNL model specification, the DM model has the advantage of being relatively simple, reasonably plausible, and computationally appealing. However, it is somewhat less flexible than

the MMNL since the attribute parameters in each class are fixed. In contrast, the main disadvantage of the MMNL is that the distribution of parameters should be specified by the analyst. The recognition of the fact that each model has its virtues and limitations has motivated a flourishing literature attempting to compare these two approaches. Greene and Hensher (2003) compare the MMNL model with the LC model in the context of a real application. Stephane Hess et al. (2007) propose a systematic comparison of continuous and discrete mixture models making use of both real and simulated data. Other applications comparing MMNL and DM specifications are those of Sergio Colombo et al. (2008), Junyi Shen et al. (2006), and Luisa Scaccia (2005), while a theoretical discussion can be found in Michel Wedel et al. (1999). Greater efforts are, however, advocated by Greene and Hensher (2003) to further compare and contrast such advanced models.

The aim of this paper, hence, is to re-explore the potential advantages and limits of MMNL and DM models both from a theoretical point of view and through an application to data on public transport demand. The hope is also to encourage a more widespread use of these models, since the vast majority of large-scale real-world applications still rely mainly on the use of MNL and nested Logit.

The paper is organised as follows. Section 2 briefly resumes the MNL, with special emphasis on its limitations. Section 3 describes MMNL and DM models and the way in which these models overcome most of the limits of MNL model. In Section 4, we apply these two specifications of RPL models to the analysis of some stated preference data on public transport demand. Finally, conclusions are given in Section 5.

2. The Multinomial logit model

2.1 Model specification

Random utility models assume that the decision maker has a perfect discrimination capability and thus, when faced with J possible alternatives, he will choose the one that maximizes his own utility. Therefore an individual n will choose alternative i if and only if $U_{ni} > U_{nj}, \forall i \neq j$, where

U_{ni} is the utility that individual n is associating with alternative i . While this utility is known to the decision maker, the researcher is, instead, supposed to have incomplete information and he can only observe a portion of the utility, V_{ni} , which is called *representative utility*. Therefore, utility is decomposed as $U_{ni} = V_{ni} + \varepsilon_{ni}$ where ε_{ni} captures the factors that affect utility but are not observable and, thus, are treated as random by the researcher. Assuming a particular density, $f(\varepsilon_n)$ for the vector $\varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nj})$ the researcher can make probabilistic statements about the individual's choice, i.e. the probability that individual n chooses alternative i is $P_n(i) = \Pr(U_{ni} > U_{nj}, \forall i \neq j) = \int I(\varepsilon_{ni} - \varepsilon_{nj} < V_{ni} - V_{nj}, \forall i \neq j) f(\varepsilon_n) d\varepsilon_n$ where $I(\cdot)$ is the indicator function. Different discrete choice models are obtained from different specifications of the distribution of the unobserved portion of utility.

The MNL model assumes that the ε_{ni} are independently, identically distributed (i.i.d.) type I extreme value, i.e. $f(\varepsilon_{ni}) = e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}}$. Under this assumption the probability that individual n will choose alternative i is simply obtained, after some algebraic manipulation, as:

$$P_n(i) = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} \quad (1)$$

When using a linear in parameters specification for the representative utility, equation (1) becomes:

$$P_n(i) = \frac{e^{\beta'x_{ni}}}{\sum_{j=1}^J e^{\beta'x_{nj}}} \quad (2)$$

where x_{ni} is a vector of observed variables relating to alternative i and individual n and β is a vector of unknown parameters.

2.2 Properties of the multinomial logit model

Under the assumptions of the MNL model, the choice probabilities have a simple closed form and are analytically differentiable, and this makes it possible to apply the traditional maximum likelihood procedures for parameter estimation. Moreover, the log-likelihood function

$$L(\beta) = \sum_{n=1}^N \sum_{i=1}^J y_{ni} \ln P_n(i)$$

in which N is the number of individuals in the sample, y_{ni} is a dichotomic variable equal to 1 if person n chooses alternative i and to 0 otherwise, and in which $P_n(i)$ is given in equation (2), is globally concave in parameters β (McFadden, 1974), which helps in the numerical maximization procedures. The simplicity of the MNL model and its computational attractiveness have made it the most widely used method for discrete choice analysis.

However, the assumption that the disturbances are i.i.d. represents an important restriction. First of all, this assumption implies that the utilities associated to different alternatives are uncorrelated, i.e.:

$$\text{Cov}(U_{ni}, U_{nj}) = \text{Cov}(\varepsilon_{nj}, \varepsilon_{nj}) = 0$$

and thus the MNL model is not capable of accounting for unobserved similarities among alternatives. Strictly related to this, is the well known property of independence from irrelevant alternatives (IIA), which states that the ratio between the probabilities of choosing two different alternatives i and k is independent from alternatives other than i and k . In fact, for the MNL logit we have:

$$\frac{P_n(i)}{P_n(k)} = \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} = e^{V_{ni} - V_{nk}}, \quad \forall i, k$$

This ratio is the same no matter what other alternatives are available or what the attributes of the other alternatives are. The same issue can be expressed in terms of cross-elasticity. It can be easily shown that, according to the MNL logit, an improvement in the attributes of an alternative reduces the probabilities for all the other alternatives by the same percentage. This pattern of substitution between alternatives is clearly unrealistic in most situations.

Another shortcoming of the MNL model is its lack of flexibility as a representation of behaviour, implying that all individuals have the same tastes. More precisely, the MNL model can accommodate for systematic taste variation: socioeconomic variables can be included as interactions with attributes or as interactions with alternative-specific constants, or different models can be estimated for different subsets of data. Hence, the MNL model can accommodate for variation that relates to observed characteristics of the individuals, but not for random taste variation. Finding better ways to represent heterogeneity in choice modelling is important to improve understanding of the factors underlying consumer behaviour and willingness to pay, and how the benefits and costs of policies are distributed across recipients. Finally, the MNL logit model can be extended to the analysis of panel data only if the unobserved factors that affect choices can be considered independent over the repeated choices. In this case, the MNL model can be used to examine panel data in the same way as purely cross-sectional data, assuming that the error components are independent over individuals, choices and time. Then, the probability that individual n will chose the set of alternatives $\mathbf{i} = \{i_1, K, i_T\}$ is:

$$P_n(\mathbf{i}) = \prod_{t=1}^T \frac{e^{\beta'x_{n,t}}}{\sum_{j=1}^J e^{\beta'x_{nj,t}}}$$

The model can accommodate for dynamics linked to observed factors, such as a person's past choices influencing current choices or lagged response to changes in attributes. However the independence of errors over repeated choices makes it impossible to handle dynamics associated with unobserved factors.

3. Random parameters logit models

3.1 Model specification

The RPL models assume, as well as the MNL model, that the ε_{ni} are i.i.d. type I extreme value. However, the parameters β are no longer

considered as fixed, but are now assumed to vary across the population, according to a certain probability distribution. The result is a mixture of models, where the underlying choice probability, conditional on the value of the parameters β , is simply the logit probability, i.e.:

$$P_n(i | \beta) = \frac{e^{\beta'x_i}}{\sum_{j=1}^J e^{\beta'x_j}}$$

When the parameters β are assumed to vary in a continuous way in the population, according to a probability density function $f(\cdot)$, the mixture of models is defined as:

$$P_n(i) = \int_{\mathbb{R}^K} P_n(i | \beta) f(\beta) d\beta \tag{3}$$

and it is generally referred to as MMNL model.

As an alternative, it can be assumed that the number of possible values for the taste parameters is finite and thus the parameter β_q of a generic q -th attribute, for $q = 1, K, Q$, has a discrete distribution, with m_q mass points labelled β_q^l , for $l = 1, K, m_q$, each of them associated with probability π_q^l satisfying the conditions that $0 \leq \pi_q^l \leq 1, \forall q, l$, and $\sum_{l=1}^{m_q} \pi_q^l = 1, \forall q$. In this case, the mixture of models, generally referred to as DM model, is defined as:

$$P_n(i) = \sum_{l_1=1}^{m_1} \dots \sum_{l_Q=1}^{m_Q} P_n(i | \beta_1^{l_1}, K, \beta_Q^{l_Q}) \pi_1^{l_1} \dots \pi_Q^{l_Q},$$

where the conditional probabilities $P_n(i | \beta_1^u, K, \beta_Q^w)$ are simply the logit probabilities.

2.2 Properties of random parameter logit models

The MNL model is clearly a particular case of both the MMNL and the DM models, which is obtained when the mixing distribution of β is degenerate at fixed values. Apart from this trivial case, RPL models

generalize the MNL model and allow to overcome its limits.

First of all, in the MMNL and the DM models, the parameters β are random and represent the tastes of individual decision makers, thus allowing for heterogeneous tastes in the population. In this framework, unlike the MNL model, both systematic and random taste variations can be accommodated. Variations related to observed attributes of the individual can be captured through specification of explanatory variables, as in MNL models, and/or the mixing distribution. For example, cost of transport may be divided by the individual's income to allow the relative importance of cost to decline as income rises. The random parameter of this variable, then, represents the unobserved variation over people with the same income on the value that they place on cost. This unobserved taste variation cannot be captured under the MNL model.

Secondly, under RPL models, the utilities of different alternatives can be correlated even if the errors are independent over alternatives. For example, for the MMNL model:

$$Cov(U_{ni}, U_{nj}) = Cov(\beta' x_{ni} + \varepsilon_{ni}, \beta' x_{nj} + \varepsilon_{nj}) = x_{ni}' x_{nj} B$$

where B is the variance-covariance matrix of β . Thus various correlation patterns, and hence substitution patterns, can be obtained. For example, a situation in which an improvement in alternative i draws proportionally more from alternative j than from alternative k can be easily represented, simply specifying an element of x that is positively correlated between i and j but negatively correlated or uncorrelated between i and k , and allowing the parameter of this variable to be random. This flexibility in representing various substitution patterns obviously breaks the undesirable IIA property, which characterizes the MNL model.

Finally, RPL models can be easily generalized to allow for repeated choices by each sampled individual. A simple way to do it consists in treating the random parameters as varying over people but being constant over choice situations for each person. Utility from alternative i in choice situation t by person n can then be written as $U_{nit} = \beta_n' x_{nit} + \varepsilon_{nit}$, with ε_{nit} being i.i.d. extreme value over time, individuals and alterna-

tives, and β_n being the parameter vector specific to subject n . Hence, the conditional probability of subject n choosing the set of alternatives $\mathbf{i} = \{i_1, \dots, i_T\}$ is:

$$P_n(\mathbf{i} | \beta_n) = \prod_{t=1}^T \frac{e^{\beta_n' x_{n,t}}}{\sum_{j=1}^J e^{\beta_n' x_{n,t}}} \quad (4)$$

and the unconditional probability under the MMNL model is then obtained integrating the conditional probabilities with respect to the mixing distribution of β . Lagged dependent variables can also be added without changing the conditional probabilities in equation (4), since lagged variables entering U_{nit} will be uncorrelated with the error terms for period t . A similar generalization can be easily obtained also for the DM model. In both cases, the randomness of the parameters allows the inclusion in the model of dynamics associated with unobserved factors.

The flexibility in representing taste heterogeneity, substitution patterns among alternatives and temporal dynamics in panel data, offered by RPL models, comes at the cost of complex specification, estimation and application issues related to these models (see Hensher and Greene, 2003). In terms of model specification, the first relevant issue is that of the choice of the parameters that are to be random. This choice is particularly important since the random parameters are the basis for accommodating correlation across alternatives and defining the degree of unobserved heterogeneity. McFadden and Train (2000) suggest a Lagrange multiplier test for testing the presence of random components against the null hypothesis of fixed-value parameters. This test statistic has the advantage that its asymptotic distribution under the null hypothesis does not depend on the parameterization of the mixing distribution under the alternative. However, it is recognized to have a low power in most situations and hence further research in this field is advisable in order to develop more powerful procedures.

Once the parameters that are to be random have been chosen, under the MMNL the problem arises of selecting an appropriate distribution for these parameters. Most popular specifications have been the normal, triangular, uniform and lognormal distributions. However, in practical

applications, any of them has shown its deficiencies, generally related to sign and length of tails – see Hess et al. (2005) for an interesting discussion of this issue. Clearly, the assumptions made during model specification have a direct influence on model results, and an inappropriate choice of mixture distribution for a given taste coefficient can lead to problems in interpretation and potentially misguided policy-decisions (Cinzia Cirillo and Kay Axhausen, 2006; Mogens Fosgerau, 2006; Hensher, 2006; Hess and Axhausen, 2004). Hensher and Greene (2003) suggest an empirical procedure based on a kernel density estimator to parameter estimates after applying a jackknife procedure to a multinomial logit model. This method allows one to visually inspect the distribution of parameters. Mogens Fosgerau and Michel Bierlaire (2007) proposed a method based on a semiparametric specification to test if a random parameter of a discrete choice model indeed follows a given distribution. Fosgerau (2008) also describes a nonparametric test procedure which uses a combination of smoothed residual plots and a test statistic able to detect general misspecification.

The problem of selecting an appropriate distribution does not arise in the DM model. The use of a discrete distribution may be seen as a nonparametric estimator of the random distribution and the researcher is not required to make any prior assumption on the shape of this distribution. However, in this case, the issue of selecting the number of support points arises. The likelihood ratio test cannot be used to choose between models with different numbers of support points, even if models are nested, and it is necessary to resort to information criteria like the AIC (Hirotugu Akaike, 1974) or the BIC (Gideon Schwarz, 1978).

The issue of parameter estimates is also more complicated in RPL models than in the MNL model. For the MMNL, the choice probabilities in equation (3) do not have a closed form and simulation methods are required for parameters estimation. In practice, M values $\beta_{i1}, \dots, \beta_{iM}$ are drawn from $f(\hat{a})$ and used to calculate the simulated probabilities

$$\tilde{P}_n(i) = \frac{1}{M} \sum_{m=1}^M \frac{e^{\beta_{(m)}' x_n}}{\sum_j e^{\beta_{(m)}' x_n}}$$

which are then used to evaluate the simulated log-likelihood

$$SLL(\beta) = \sum_{n=1}^N \sum_{i=1}^J y_n \ln \tilde{P}_n(i)$$

The number of draws required to secure a stable set of parameter estimates varies enormously, according to the complexity of the model specified, and estimation can be particularly time consuming. The use of Halton draws (John Halton, 1960) can sensibly reduce computational time, however some authors stress the need for further investigation of their properties in simulation-based estimation (see Zsolt Sándor and Kenneth Train, 2004).

An obvious advantage of the DM approach compared to MMNL models is that, if the model for the conditional choice probabilities used inside the mixture has a closed form, as the MNL model does, then the DM has itself a closed form. However, the non-concavity of the log-likelihood function does not allow the identification of a global maximum, even for discrete mixtures of MNL. Given the potential presence of a high number of local maxima, performing several estimations from various starting points is advisable. Moreover, constrained maximum likelihood must be used to account for constraints on the weights π_q^i . Finally, parameter estimation in DM models frequently suffers from clustering of mass points, which causes models with more support points to collapse back to more parsimonious specifications.

All these specification, estimation and application issues of the RPL models, as well as the advantages they offer over the MNL model, will be highlighted in the next section through an application to real data.

4. Application

4.1 Data set description

The data set refers to a study carried out between January and Mars 2004 on the bus service which links the centre of Urbino (Italy), where the University is located, to Sogesta, a residential location with more than 100 students (Edoardo Marcucci and Luisa Scaccia, 2005). The distance between the two locations is about 2 km and the bus took about

9 minutes to cover it. The aim was to analyse the attributes of the local public transport and to investigate possible interventions to improve the service. The quality of the service was, in fact, considered as unsatisfactory and many students were used to hitchhike along the road.

In order to identify the attributes that characterize the quality of the service, a focus group of 30 students was interviewed. The group resulted particularly sensitive to five attributes of the service: cost of monthly ticket, headway, first and last run, real time information displays, bus shelters. Each attribute was further described by five levels. The attributes and corresponding levels, as used in the study, are shown in Tab. 1.

Tab. 1 - Attributes and levels characterizing the quality of the bus service

ATTRIBUTES	LEVELS
Cost of monthly ticket	12.80 €
	15.40 €
	16.00 € (real situation)
	18.20 €
	21.80 €
Headway	28 minutes
	31 minutes
	45 minutes (actual situation)
	59 minutes
	64 minutes
First and last run	05:15 -- 03:00
	06:15 -- 02:00
	07:15 -- 01:00 (actual situation)
	08:15 -- 24:00
	09:15 -- 23:30
Information displays	Real time information displays
	Without real time information displays (actual situation)
Bus shelters	At Meccale and Segesat
	Only at Meccale (actual situation)

Using the software CBC (Choice - Based Conjoint) of Sawtooth Software (<http://www.sawtoothsoftware.com>), questionnaires were created, each containing 15 choice exercises, 11 of which were random, 2 aimed at testing the quality of the answers, and 2 aimed at testing preference stability. Each choice exercise contained four hypothetical

alternatives as shown in Tab. 2. Respondents were approached randomly and face-to-face interviews were carried out at the bus stop. A total number of 50 respondents took part in the study, providing a data set of 750 observations.

Tab. 2 - An example of a stated choice experiment

ALTERNATIVE A	ALTERNATIVE B	ALTERNATIVE C	ALTERNATIVE D
Use of mobile gadget 101F	Use of mobile gadget 201F	Use of mobile gadget 301F	None of previous alternative
Travelers 3 minutes	Travelers 25 minutes	Travelers 30 minutes	
First and last rail (07:15 - 01:00)	First and last rail (08:15 - 2:00)	First and last rail (07:15 - 01:00)	
With real-time information displays	Real time information displays	Real time information displays	
In shelters of Venezia and Suzzola	In shelters of Venezia and Suzzola	In shelters of Venezia	

4.2 Model specification and results

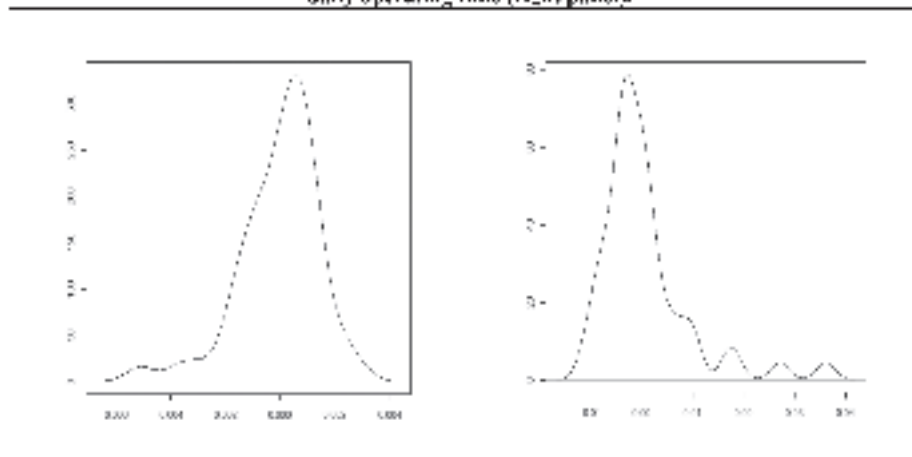
Three different models were specified and estimated on the data: a simple MNL model, a MMNL model and a DM model. In the MMNL and DM models, the repeated choice nature of the data was taken into account by specifying the likelihood function with the integration and summation, respectively, outside the product over replications for the same respondent. The models were estimated in Biogeme (Bierlaire 2003, 2008). The utility of MNL model was specified as a linear function of all the attributes and also of alternative specific constants (ASC). The significance of an ASC related to an unlabelled alternative would imply that, after controlling for the effects of the modelled attributes, this alternative has been chosen more or less frequently than the others, revealing alternative ordering effects. However this was not the case with the present dataset and all the ASCs turned out to be not significantly different from zero and were thus removed from further analysis. Also the coefficient of the dummy for the presence of real time information displays turned out to be not significantly different from zero (p-value of 0.57) and the dummy variable was then removed from the specification

of the utility function. Parameter estimates for the remaining attributes are given in Tab. 3. Notice that the first and last run times were specified in the utility function in terms of daily operating time (i.e. subtracting the first run time from the last run time).

To specify the RPL models, the Lagrange multiplier test suggested by McFadden and Train (2000) was used to decide which parameters are to be random. The null hypothesis of no mixing was rejected for the parameters of the attributes headway and daily operating time. The cost parameter was instead treated as non random because, in this way, the estimation of marginal willingness to pay (WTP) for an improvement in a certain attribute is simplified and its distribution is simply the distribution of that attribute's coefficient. Moreover, treating the parameter of the cost as fixed allows to restrict the cost variable to be non positive for all individuals.

Once the parameters of headway and daily operating time were chosen to be random, the jackknife procedure proposed by Hensher and Greene (2003) was used to obtain a kernel density estimator to parameters distribution. The results are shown in Fig. 1. Both of the densities seem to be unimodal and, even if some skweness can be noticed, the normal density was chosen to approximate the distribution of parameters of headway and daily operating time in the MMNL model. The model was then estimated using 1,000 random draws.

Fig. 1 Kernel density estimate for the parameters of headway (left panel) and daily operating time (right panel).



In the specification of the DM model, 2 support points were chosen for the distribution of both random parameters. When models with a larger number of support points were tried, the smallest probability happened to be further split, becoming not significantly different from zero, causing these models to collapse back to the one with 2 support points. Notice that many different starting values for parameter estimates were used to deal with the local maxima problem.

Parameter estimates for the three models specified are given in Tab. 3. It is observed that under any specification, the signs of the parameter estimates are as expected. The negative signs of the cost and headway parameters indicate that the utility of the trip maker decreases with an increase in the magnitude of the respective attributes. The utility increases instead as the daily running time does. For the qualitative attribute, bus shelters, the positive sign indicates that the presence of this attribute at both bus stations is considered utility.

Tab. 3 - Estimation results for the three models specified.

	MNL			SOSNL			DM2		
Ln Lik	-125.309			-191.296			-280.082		
N	4			6			10		
Adj. R ²	0.456			0.557			0.495		
Est. time	30:50			15:51			00:02		
	Est.	Std. Err.	p-value	Est.	Std. Err.	p-value	Est.	Std. Err.	p-value
β_{cost}	0.106	0.0267	0.00	0.253	0.0248	0.00	0.241	0.0221	0.00
$\beta_{headway}$	0.327	0.149	0.05	0.322	0.17e	0.05	0.333	0.156	0.04
$\beta_{running}$	-0.06247	0.006	0.00	-	-	-	-	-	-
$\beta_{shelters}$	-	-	-	-0.7941	0.0749	0.00	-	-	-
$\beta_{shelters^2}$	-	-	-	0.06e6	0.0136	0.00	-	-	-
$\beta_{shelters^3}$	-	-	-	-	-	-	-0.0287	0.00882	0.00
$\beta_{shelters^4}$	-	-	-	-	-	-	0.0296	0.0121	0.00
$\beta_{shelters^5}$	-	-	-	-	-	-	-0.0162	0.0028	0.00
$\pi_{shelters}$	-	-	-	-	-	-	0.674	0.120	0.00
$\beta_{running^2}$	0.115	0.0128	0.00	-	-	-	-	-	-
$\beta_{running^3}$	-	-	-	0.620	0.0815	0.00	-	-	-
$\beta_{running^4}$	-	-	-	0.354	0.0792	0.00	-	-	-
$\beta_{running^5}$	-	-	-	-	-	-	0.067	0.0214	0.00
$\beta_{running^6}$	-	-	-	-	-	-	0.075	0.0275	0.00
$\beta_{running^7}$	-	-	-	-	-	-	0.0531	0.0207	0.00
$\pi_{running}$	-	-	-	-	-	-	0.185	0.0225	0.00

The interpretation of the coefficients is not meaningful except for significance and sign. Therefore, the marginal WTPs for the different attributes are calculated by taking minus the ratios between the coef-

ficients of the attributes and the coefficient of the cost attribute. These values represent the marginal rates of substitution between the attributes and the cost and provide an idea of how much the travellers, on average, are willing to pay for a positive unit change in each quantitative attribute under consideration. The WTP for the bus shelter attribute indicates the willingness to pay to have bus shelters also at Sogesta and not only at Mercatale. In the MMNL model, with a fixed cost coefficient and normally distributed attributes, marginal WTPs are also normally distributed. Tab. 4 summarizes the marginal WTP estimates from the different models under consideration. For example, the MMNL shows an average WTP of 0.36 Euros more on the monthly ticket to obtain 1 minute less of headway. This value is very closed to the one obtained under the MNL model. Under the DM model, the population is divided into two groups of almost equal size with respect to their WTP for shorter headway. The first group shows a very small propensity to pay for a shorter headway, while the second group is much more sensitive to a long headway and hence is willing to pay more to shorten it. Looking at the WTP for operating time, it can be noticed that the DM model identifies a segment of the population, the 18.5% of it, which is not sensitive to operating time (the parameter $\beta_{run.time,2}$ in Tab. 3 is not significantly different from zero) and hence is not willing to pay for an improvement of this attribute, i.e. for an extension of the operating time.

Tab. 4 - Estimation of the WTP for the different attributes under the three models specified.

WTP	MNL	MMNL	DM(2)	
Headw.	1.54	1.37	1.46	
Shelter	-	-	-	
Headw.	-0.34	-0.36	-0.16	-0.71
Shelter	-	0.26	-	-
Run.time	2.15	2.49	2.89	0.24
Shelter	-	1.19	-	-

Moving on to the comparison of the three different models, from Tab. 3 it can be noticed that both the RPL models offer a significant improvement in model fit over the MNL model. The MMNL offers the best performance, with an adjusted ρ^2 equal to 0.507. The DM model, even if improving over the MNL model, does not fit the data so well as

the MMNL model. This probably depends on the fact that the inspection of the parameter distributions through the jackknife procedure seems to reveal unimodal distributions, for which the DM with 2 mass points might offer a not very good approximation. Probably a larger number of support points would be required, if the real parameter distributions were effectively unimodal and continuous, but this would determine difficulties in the estimation procedures, as alluded to in Section 3.2.

Another observation relates to the much lower estimation cost of the DM model, with an estimation time of 2 seconds, compared to the 10 minutes and 21 seconds required by the MMNL model. This much lower estimation time would give a significant advantage to the DM model in the case of larger data sets. However, the estimation time for the MMNL model is particularly high, since we decided to use random rather than Halton draws in the estimation.

5. Summary and conclusions

RPL models allow to overcome all the limits of the simple MNL model. As a drawback, several issues arise in both model specification and estimation. In this paper, we summarized the theoretical results concerning RPL models, considering both continuous distributions, leading to the MMNL model, and discrete distributions, leading to the DM model, for the parameters. We outlined the advantages and the limits of both these RPL models, also making use of an application to a real data set concerning the public transport demand.

The results from the application clearly show the major advantage of the DM approach in terms of estimation cost, due to the fact that parameter estimates have a closed form solution and does not rely on simulation processes. Moreover, this approach avoids the problem of choosing an adequate distribution for the random parameters. However, even if the DM model provides a considerable improvement over the MNL model, for the example at hand it does not seem to be competitive with the MMNL model in terms of fit to the data. The reason is probably that, in this case, the underlying distribution of the random parameters seems to be continuous and unimodal, and, thus, 2 support points are not enough to provide a reasonable approximation to it. Hess et al.

(2007), in a simulation study, find that DM models, with a sufficiently large number of support points, can offer a very good approximation to the normal distribution. However, with our data set, we experienced a clustering of mass points which did not allow us to estimate DM models with a number of support points larger than 2. This is probably due, also, to the fact that the data set is a relatively small one.

In this regard, the paper provides a different perspective to that of Hess et al. (2007), Shen et al. (2006) and Colombo et al. (2008) in the ongoing discussion about the comparison between MMNL and DM models. In small data sets, with continuous and unimodal underlying distribution of the random parameters, MMNL models, generally requiring a smaller number of parameters to account for heterogeneity in tastes, could perform better than DM models.

References

Akaike H. (1974), *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, 19, 716-723.

Bierlaire M. (2003). *BIOGEME: A free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire M. (2008). *An introduction to BIOGEME Version 1.7*, biogeme.epfl.ch.

Colombo S., Hanley N., Louviere J. (2008), *Modelling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture*, Stirling Economics Discussion Papers 2008-28, University of Stirling, Department of Economics.

Boxall P.C., Adamowicz W. (2002), *Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach*, Environmental and Resource Economics, 23, 421-446.

Cirillo C., Axhausen K.W. (2006), *Evidence on the distribution of values of travel-time savings from a six-week travel diary*, Transportation Research Part A, 40, 444-457.

Fosgerau M. (2006), *Investigating the distribution of the value of travel time savings*, Transportation Research Part B, 40, 688-707.

Fosgerau M. (2008), *Specification testing of discrete choice models: a note*

on the use of a nonparametric test, *Journal of Choice Modelling*, 1, 26-39.

Fosgerau M., Bierlaire M. (2007), *A practical test for the choice of mixing distribution in discrete choice models*, *Transportation Research Part B*, 41, 784-794.

Greene W.H., Hensher D.A. (2003), *A latent class model for discrete choice analysis: contrasts with mixed logit*, *Transportation Research Part B: Methodological*, 37, 681-698.

Halton J. (1960), *On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals*, *Numerische Mathematik*, 2, 84-90.

Hensher D.A. (2006), *The signs of the times: imposing a globally signed condition on willingness to pay distributions*, *Transportation*, 33, 205-222.

Hensher D.A., Greene W.H. (2003), *Mixed logit models: state of practice*, *Transportation*, 30, 133-176.

Hess S., Axhausen K.W. (2004), *Checking our assumptions in value-of-travel-time modelling: Recovering taste distributions*, CTS Working Paper, Centre for Transport Studies, Imperial College London.

Hess S., Bierlaire M., Polak J.W. (2005), *Estimation of value of travel-time savings using mixed logit models*, *Transportation Research Part A: Policy and Practice*, 39, 221-236.

Hess S., Bierlaire M., Polak J.W. (2007), *A systematic comparison of continuous and discrete mixture models*, *European Transport \ Trasporti Europei*, 37, 35-61.

Kamakura W., Russell G. (1989), *A probabilistic choice model for market segmentation and elasticity structure*, *Journal of Marketing Research*, 26, 379-390.

Lee B.J., Fujiwara A., Zhang J., Sugie Y. (2003), *Analysis of Mode Choice Behaviours based on Latent Class Models*, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne.

Marcucci E., Scaccia L. (2005), *Alcune applicazioni dei modelli a scelta discreta al settore dei trasporti*, In: E. Marcucci (Ed.), *I modelli a scelta discreta nel settore dei trasporti. Teoria, metodologia e applicazioni*, Carocci editore, Roma.

McFadden D. (1974), *Conditional logit analysis of qualitative choice behaviour*, In: P.C. Zarembka (Ed.), *Frontiers in Econometrics*, New York, Academic Press, 105-142.

McFadden D., Train K. (2000), *Mixed MNL models for discrete response*, *Journal of Applied Econometrics*, 15, 447-470.

Sándor Z., Train K. (2004), *Quasi-random simulation of discrete choice models*, *Transportation Research Part B: Methodological*, 38, 313-327.

Scaccia L. (2005), *L'eterogeneità delle preferenze nel trasporto merci: un confronto tra diversi metodi per catturarla*, In: G. Borruso and G. Polidori (Ed.), *Riequilibrio ed integrazione modale nel trasporto delle merci. Gli attori ed i casi italiani*, Franco Angeli, Milano, 83-95.

Schwarz G.E. (1978), *Estimating the dimension of a model*, *Annals of Statistics*, 6, 461-464.

Shen J., Sakata Y., Hashimoto Y. (2006), *A Comparison between Latent Class Model and Mixed Logit Model for Transport Mode Choice: Evidences from Two Datasets of Japan*, Discussion Papers in Economics and Business 06-05, Osaka University, Graduate School of Economics and Osaka School of International Public Policy (OSIPP).

Train K. (1998), *Recreation demand models with taste differences over people*, *Land Economics*, 74, 230-239.

Train K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Wedel M., Kamakura W., Arora N., Bemmaor A., Chiang J., Elrod T., Johnson R., Lenk P., Neslin S., Poulsen C.S. (1999), *Discrete and continuous representations of unobserved heterogeneity in choice modelling*, *Marketing Letters*, 10, 219-232.