Random Covariance Heterogeneity in Discrete Choice Models

Stephane Hess* Denis Bolduc† John W. Polak‡

June 17, 2005

Abstract

In this paper, we extend the standard discrete choice modelling framework by allowing for random variation in the substitution patterns between alternatives across respondents, leading to increased model flexibility. The paper shows how such a Mixed Covariance model can be specified either with purely random variation or with a mixture between random and deterministic variation. Additionally, the model can be based on an underlying GEV or ECL structure. Finally, the model can be specified as a continuous mixture or as a discrete mixture. A brief application on simulated data shows that our proposed model structure is able to retrieve variations in the error-structure across respondents, hence avoiding a source of bias in forecasting applications.

1 Introduction

Discrete choice models have been used extensively in various areas of economic research, notably transport studies, for over thirty years. Initially, virtually all applications were based on the basic Multinomial Logit (MNL) model (Luce 1959, Marschak 1960, McFadden 1974), which, although easy to specify, estimate and apply, has significant disadvantages in terms of flexibility, most notably in the form of very restrictive substitution patterns across alternatives (governed by the *IIA* assumption). Initial gains in flexibility

*Centre for Transport Studies, Imperial College London, stephane.hess@imperial.ac.uk, Tel: +44(0)20 7594 6105, Fax: +44(0)20 7594 6102

†Département d'Économique, Université Laval, denis.bolduc@ecn.ulaval.ca, Tel: +1 (418) 656-5427, Fax: +1 (418) 656-2707

‡Centre for Transport Studies, Imperial College London, j.polak@imperial.ac.uk, Tel: +44(0)20 7594 6089, Fax: +44(0)20 7594 6102

were made by the development of structures belonging to the family of Generalised Extreme Value (GEV) models, allowing for heightened correlation between alternatives that are closer substitutes for each other, thus relaxing the assumptions imposed on MNL cross-elasticities. The best known example of a GEV model is the Nested Logit (NL) model (Williams 1977, McFadden 1978, Daly & Zachary 1978), which has been used extensively by researchers and practitioners. Further increases in flexibility can be obtained with the use of models allowing for multi-nest membership. The general form of these models is given by the Cross-Nested Logit (CNL) model (Vovsha 1997, Vovsha & Bekhor 1998, Papola 2004, Bierlaire 2005); other well-known models such as the Ordered Generalised Extreme Value (OGEV) model (Small 1987) or the Paired Combinatorial Logit (PCL) model (Chu 1989) can be seen as special cases of this CNL structure. Two even more general model forms have recently been proposed; the Generalised Nested Logit (GNL) model of Wen & Koppelman (2001) generalises all two-level GEV structures, while Daly & Bierlaire (2005) propose a model form that is also able to generalise existing multi-level GEV models[1]. For an extensive review of GEV models, see Koppelman & Sethi (2000) and Train (2003).

While GEV models avoid the often unrealistic substitution patterns of the MNL model, they are, like the MNL model, restricted to expressing taste heterogeneity across respondents in a deterministic fashion. Given the usual limitations of the data, along with inherent randomness involved in decision-making, there is however generally some remaining non-quantifiable (random) variation in tastes, which, if left untreated, can lead to biased model estimates. Researchers have recently begun to increasingly exploit the power of an alternative model form, the Mixed Multinomial Logit (MMNL) model, in which choice probabilities are expressed as integrals of MNL choice probabilities over the (assumed) distribution of the error terms present in the model, in addition to the usual *IID* extreme-value terms. Due to the absence of a closed-form solution for the MMNL choice-probabilities, numerical techniques, typically simulation, are required in the estimation and application of this model. The computational cost of these numerical techniques meant that the MMNL model remained largely confined to theoretical discussions for years following its development (the MMNL model was first discussed by Boyd & Mellman 1980 and Cardell & Dunbar 1980). However, with the development of ever more powerful computers and simulation techniques (c.f.

[1]The work by Daly & Bierlaire (2005) combines earlier work by Daly (2001) on the *Recursive Nested Extreme Value* model and by Bierlaire (2002) on the *Network GEV* model.

Hess, Train & Polak 2005), the number of applications using the MMNL model has increased significantly over recent years.

Two distinct interpretations of the MMNL model have been discussed in the literature; the Random Coefficients Logit (*RCL*) formulation exploits the error structure of the MMNL model to accommodate a random distribution of tastes across decision-makers, while the Error Components Logit (*ECL*) formulation allows the model to approximate any GEV correlation structure arbitrarily closely. The two approaches can also be combined to allow for the joint modelling of random taste heterogeneity and flexible substitution patterns. The *RCL* formulation is used more regularly than its *ECL* counterpart, with some recent examples in the field of transport studies being given by Bhat (2000), Bhat & Castelar (2002), and Hess & Polak (2004, 2005*b*). The *ECL* formulation has been used amongst others by Bhat (1998), Brownstone & Train (1999) and Hess, Polak, Daly & Hyman (2005). For more details on the two specifications, see McFadden & Train (2000), Walker (2001) and Train (2003). Finally, an alternative to the *ECL* formulation for the joint representation of random taste heterogeneity and flexible correlation structures is to use integration of GEV-style choice probabilities over the distribution of taste coefficients, leading to a Mixed GEV model (c.f. Bhat & Guo 2004, Hess, Bierlaire & Polak 2005*a*). This model form not only reduces the number of random terms in the models to the number of random taste coefficients, hence easing the computational burden, but also avoids important issues of identification that need to be faced when using the *ECL* formulation (c.f. Walker 2001).

While the developments in relation to closed-form GEV as well as GEV mixture models have led to gradual gains in modelling flexibility, by allowing modellers to accommodate correlation across alternatives as well as deterministic and random taste heterogeneity across respondents, little effort has gone into the development of model forms allowing for a representation of heterogeneity across respondents in the correlation structure in place between the different alternatives. Such correlation heterogeneity is however potentially a crucial factor in the variation of choice-making behaviour across decision-makers. As an example, in an airline-choice scenario, travellers' behaviour can be strongly affected by their membership in a given airline's frequent flier programme, to the point that, in the case where seats on their desired flight are not available, they are more likely to switch to a different flight on the same airline than to choose a flight by an alternative airline. In many cases, it may not be possible to accommodate the effects of airline allegiance directly, mainly for data reasons (c.f. Hess & Polak 2005*a*). In these circumstances, the greater substitution between flights on the same airline can be accommodated through a nesting structure that allows for correlation between flights on the same airline. It is clearly possible that the effects of airline allegiance, and hence the level of correlation, vary across travellers, meaning that the use of an approach which imposes covariance homogeneity potentially leads to biased model results.

While some of the covariance heterogeneity can conceivably be accommodated through an appropriate segmentation of the population (using separate models), it is likely that some within-segment heterogeneity remains. The existing literature seems to contain only two examples of a model allowing for such heterogeneity. The first of these comes in the form of the Covariance Nested Logit (COVNL) model discussed by Bhat (1997). In the COVNL model, the structural parameters themselves (and hence the pattern of substitution between alternatives) are a function of socio-demographic attributes of the decision-makers, such that the correlation heterogeneity is explained with the help of these attributes. Koppelman & Sethi (2005) later expand this approach by incorporating covariance heterogeneity in a GNL model[2], where they additionally allow for heteroscedasticity across respondents through a parameterisation of the scale factor (c.f. Swait & Adamowicz 1996, 2001), describing the resulting model as the Heterogeneous Generalized Nested Logit (HGNL) model.

While it is highly desirable to explain any covariance heterogeneity in a deterministic way, this is clearly not always possible, and even where it is possible, there is potentially some remaining random heterogeneity that cannot be explained in a deterministic fashion. The aim of this paper is therefore to develop a model structure that can accommodate random covariance heterogeneity in addition to deterministic covariance heterogeneity. The discussion presented in this paper is based on an underlying GEV model for representing the correlation between alternatives; it is similarly possible to do this with the help of an ECL structure, and the development of such a framework is described in Appendix A.

The remainder of this paper is organised as follows. The methodology for the Mixed Covariance GEV model is introduced in Section 2. Section 3 presents an application showing how one specific example of a Mixed Covariance GEV model works in practice. Finally, Section 4 presents the conclusions of the research.

[2]Thus also allowing for cross-nesting.

2 Methodology

We will now develop the structure for our Mixed Covariance GEV model, where the derivation described here looks mainly at the case of a simple two-level NL model; the extension to multi-level as well as cross-nesting structures is possible, and several notes to that extent are made in the text. The exposition of the theory is divided into three parts. We first look at the general model form, in Section 2.1, before moving on to the cases of purely random variation (Section 2.2) and combined deterministic and random variation (Section 2.3).

2.1 General model form

The choice probabilities in a nested model are represented through a product of successive choice probabilities that represent a chain from the root of the tree (upper-most node) to the elementary alternative for which the probability is calculated. In a two-level NL model, the choice probability of alternative i (belonging to nest m) for individual n is then given by:

$$\begin{aligned} P_n(i) &= P_n(S_m) \cdot P_n(i \mid S_m) \\ &= \frac{e^{\lambda_m I_{m,n}}}{\sum_{l=1}^{M} e^{\lambda_l I_{l,n}}} \cdot \frac{e^{\frac{V_{i,n}}{\lambda_m}}}{\sum_{j \in S_m} e^{\frac{V_{j,n}}{\lambda_m}}} \end{aligned} \tag{1}$$

with logsum term

$$I_{m,n} = \ln \sum_{j \in S_m} e^{\frac{V_{j,n}}{\lambda_m}}, \tag{2}$$

where $V_{j,n}$ gives the observed utility for alternative j and individual n, λ_m is the structural parameter associated with nest m, S_m defines the set of alternatives contained in nest m, and M gives the total number of nests. The extension of the choice-probability from equation (1) to the multi-level case is straightforward, with details given for example by Koppelman & Sethi (2000).

The COVNL model of Bhat (1997) expands on the standard NL model, by parameterising the structural parameters λ as:

$$\lambda_{m,n} = F\left(\alpha + \boldsymbol{\gamma}' \boldsymbol{z_n}\right), \tag{3}$$

where α is a constant, $\boldsymbol{z_n}$ is a vector of attributes of decision-maker n, and where $\boldsymbol{\gamma}$ is a vector of coefficients. In this notation, $\lambda_{m,n}$ is the structural parameter for nest m and decision-maker n. Both α and $\boldsymbol{\gamma}$ are to be estimated.

To ensure consistency with utility maximisation, $F()$ needs to be specified so as to produce values in the $0-1$ interval. Furthermore, Bhat (1997) states that increases in $\boldsymbol{z_n}$ should have a monotonic effect on λ_n (where this ensures consistency in the case of multi-level structures, c.f. equation (7)). This double requirement can be satisfied by using a function $F()$ with:

$$
\begin{aligned}
F(-\infty) &= 0 \\
F(+\infty) &= 1 \\
f(x) &= \frac{\partial F()}{\partial x} > 0
\end{aligned}
\tag{4}
$$

These conditions are met by the use of a continuous cumulative probability distribution function, where Bhat (1997) suggests the use of the logistic distribution.

We now extend this approach to the case where λ_m follows a random distribution across individuals. Conditional on a given set of values for the vector (of length M) of structural parameters $\boldsymbol{\lambda}$, the NL choice probabilities are given by equation (1). We now assume that the vector $\boldsymbol{\lambda}$ is distributed according to $f(\boldsymbol{\lambda} \mid \Omega)$, where Ω is a vector of parameters of the distribution of the different elements of $\boldsymbol{\lambda}$. This specification is general, and can be adapted for the special cases presented in Sections 2.2 and 2.3.

The conditional choice probability in equation (1) is now replaced by the unconditional choice probability:

$$
P_n(i) = \int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\lambda}) \, f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda} \tag{5}
$$

$$
= \int_{\boldsymbol{\lambda}} \frac{e^{\lambda_m I_{m,n}}}{\sum_{l=1}^{M} e^{\lambda_l I_{l,n}}} \cdot \frac{e^{\frac{V_{i,n}}{\lambda_m}}}{\sum_{j \in S_m} e^{\frac{V_{j,n}}{\lambda_m}}} f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda}, \tag{6}
$$

where $\boldsymbol{\lambda} = \{\lambda_1, ..., \lambda_M\}$. Here, equation (6) is specific to the two-level NL model given in equation (1), while equation (5) shows the general form, where $P_n(i \mid \boldsymbol{\lambda})$ can represent the conditional choice probability for any GEV model[3]. The logsum term I_m is defined as in equation (2), and it should be noted that this logsum term is conditional on a given value of λ_m, and

[3] In the case of cross-nesting structures, there is an additional dependency on a vector of allocation parameters, which is not explicitly stated in equation (5). There is in that case also a possibility of allowing for deterministic as well as random variations across agents in the allocation parameters.

hence $\boldsymbol{\lambda}$, by being inside the integral. The behaviour of the model depends crucially on the specification used for $f(\boldsymbol{\lambda} \mid \Omega)$, where the requirements on the range of the structural parameters need to be borne in mind. This issue is discussed in more detail in the description of the two special cases in Sections 2.2 and 2.3.

The approach becomes more complicated in the case of multi-level structures, due to the condition that the structural parameters need to decrease as we move down the tree. In the COVNL, this is made possible by specifying the structural parameter of a lower-level nest, λ_l, as in equation (3), and by adapting the specification of the upper-level nesting parameter as:

$$\lambda_{m,n} = F\left[\left(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n}\right) + G\left(\delta + \boldsymbol{\eta}'\boldsymbol{w_n}\right)\right], \tag{7}$$

where $\boldsymbol{w_n}$ is an additional vector of individual characteristics, which can be the same as $\boldsymbol{z_n}$, and where δ and $\boldsymbol{\eta}$ are a constant and vector respectively that need to be estimated. Finally, $G()$ is a monotonically increasing function mapping real numbers onto the space of positive real numbers, such as for example with the exponential distribution.

In the case of the Mixed Covariance NL model, the issue becomes more complicated, as the different structural parameters are now random variables. To ensure consistency with utility maximisation, the distribution of the structural parameters must be specified such that structural parameters belonging to the same link in a tree are no longer distributed independently. As it is desirable not to have to impose a constraint of equality of the structural parameters on a given level[4], it is preferable to use a top-down approach in the notation for the Mixed Covariance NL model, given that a specific node may have multiple *descendants*, while, in a model without cross-nesting, each node has only one direct *ancestor*.

One possible way of ensuring decreasing structural parameters is to specify the values as follows. With an upper-level structural parameter being given by:

$$\lambda_u \sim f\left(\lambda_u \mid \Omega_u\right), \tag{8}$$

the structural parameter of one of its *descendants*, λ_{li}, is given by:

$$\lambda_{li} = \lambda_u \cdot \widehat{\lambda_{li}}, \tag{9}$$

with

$$\widehat{\lambda_{li}} \sim f\left(\widehat{\lambda_{li}} \mid \Omega_{\widehat{\lambda_{li}}}\right), \tag{10}$$

[4]This approach is taken by Bhat (1997).

where, in either case, the subscript imposed on Ω refers to the subelements linked to the structural parameter in question. This approach avoids the need to specify a complete joint density for the structural parameters.

The structural parameter at a lower level is thus given by multiplying the structural parameter at the level above it with a draw from the distribution used for the structural parameter at the lower level. As this draw is contained between 0 and 1, the resulting product is necessarily constrained between 0 and λ_u, giving $0 \leq \lambda_{li} \leq \lambda_u \leq 1$. If, at a given level, the draw from the distribution approaches 1, such that the resulting structural parameter takes the same value as its *ancestor*, this level of the tree becomes obsolete in that link, and the nests below it can be attached directly to the *ancestor* node. Extension of this theory to models with more than three levels is straightforward.

Extensions to models allowing for cross-nesting is also possible, although slightly more tedious. In this case, a given node can have multiple ancestors, and the condition of decreasing structural parameters needs to apply for each of the links to an ancestor. This means that the structural parameter at a given node needs to be less than or equal to that of the direct *ancestor* with the lowest structural parameter. Hence, in equations (9) and (10), λ_u is accordingly replaced by the structural parameter of this specific *ancestor* node. As it is thus possible to adapt this approach for models allowing for cross-nesting as well as for models allowing for multi-nest membership, it can be seen that the approach should be applicable for all existing GEV structures.

The final step in the theoretical development of our proposed model form is the representation of taste heterogeneity across individuals, where this heterogeneity relates to the coefficients multiplying the attributes of the alternatives, as opposed to the structural parameters. The above framework clearly already allows for deterministic variations in tastes; additional random variation can be accommodated very easily in the present model form, through integration of the choice probabilities that are conditional on β over the assumed distribution of the taste coefficients. This comes in addition to the integration over the distribution of the structural parameters.

Let $P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda})$ give the choice probability of alternative i for individual n, conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. Following the theory described in this section, we then have:

$$P_n(i \mid \boldsymbol{\beta}) = \int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}) f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda}. \tag{11}$$

By assuming that the tastes are distributed randomly across decision-makers according to $g(\boldsymbol{\beta} \mid \Theta)$, with parameter vector Θ, we obtain the unconditional choice probability[5]:

$$P_n(i) = \int_{\boldsymbol{\beta}} P_n(i \mid \boldsymbol{\beta})\, g(\boldsymbol{\beta} \mid \Theta) \mathrm{d}\boldsymbol{\beta}$$
$$= \int_{\boldsymbol{\beta}} \left(\int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda})\, f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda} \right) g(\boldsymbol{\beta} \mid \Theta)\, \mathrm{d}\boldsymbol{\beta}. \tag{12}$$

2.2 Model with purely random covariance heterogeneity

We now look at the case where any variation in the structural parameters (and hence the correlation) across individuals is purely random. Two possible approaches arise in this case.

In the first approach, we rewrite the choice probabilities in equation (5) as:

$$P_n(i) = \int_{\boldsymbol{x}} P_n(i \mid \boldsymbol{\lambda} = T(\boldsymbol{x}))\, f(\boldsymbol{x} \mid \Omega) \mathrm{d}\boldsymbol{x}, \tag{13}$$

where $T(\boldsymbol{x})$ is a transform that maps the elements in $\boldsymbol{x}$ from the real space of numbers into the $0-1$ interval. With this approach, any choice of statistical distribution can be used for $f(\boldsymbol{x} \mid \Omega)$, and a transform such as the logistic distribution can be used for $T(\boldsymbol{x})$.

The second approach avoids the use of the additional transform $T(\boldsymbol{x})$, and draws for the structural parameters are produced directly from the function $f(\boldsymbol{\lambda} \mid \Omega)$, as shown in equation (5). In this case, the condition on the range of the structural parameters applies directly at the level of $f(\boldsymbol{\lambda} \mid \Omega)$, leading to a requirement to use distributions bounded on either side, with the left bound being greater than 0, and the right bound being smaller than 1. The vector Ω now contains the parameters of the actual distribution of the structural parameters, as opposed to the distribution of the random vector $\boldsymbol{x}$ used as the base of the transform described in the first approach. A number of different statistical distributions can be used with this approach, including basic examples such as the *Uniform* or *Triangular*, or indeed the Johnson S_B distribution.

It is not clear a priori which of the two approaches is preferable. The former approach allows for greater freedom in the choice of distribution for $f(\boldsymbol{x} \mid \Omega)$, while the latter approach provides more control over the actual shape of the distribution of the structural parameters. The merits of the two approaches potentially need to be evaluated on a case-by-case basis.

[5]Although beyond the scope of the present discussion, it is possible to expand this approach to the case where $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ follow some form of joint distribution.

2.3 Model with deterministic and random covariance heterogeneity

While the description in Section 2.2 has shown that the framework developed in Section 2.1 can be adapted straightforwardly to allow for a purely random distribution of structural parameters across individuals, the use of this approach leads to similar issues of interpretation as in the case of randomly distributed taste coefficients in a GEV mixture model. Indeed, this approach provides little information about the values of the structural parameters for a given individual or a given population group, although posterior methods can be used to infer some such information. It is thus clearly preferable to as much as possible explain this covariance heterogeneity in a deterministic manner, as in the COVNL model of Bhat (1997). As mentioned in the introduction, this is not always possible, such that the Mixed Covariance models presented in this paper present a useful alternative. However, it is conceivable that there are cases in which it is possible to explain some of the variation in a deterministic way, while some remaining part of covariance heterogeneity can only be explained in a random way, along the lines of $\lambda = F\left(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n} + \epsilon\right)$, where ϵ is a random component. Two approaches are possible in this case, one is to use a mixed version of a formulation analogous to the COVNL formulation (but within a top-down approach), while the other is to parameterise the parameters of the distribution used to represent covariance heterogeneity in the Mixed Covariance GEV model. We will now look at these two approaches in turn.

2.3.1 Extension of COVNL approach

We begin the description of this approach by rewriting the choice probabilities in equation (5) as:

$$P\left(i\right) = \int_{\boldsymbol{\theta}} P\left(i \mid \boldsymbol{\lambda} = T\left(H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)\right)\right) f\left(\boldsymbol{\theta} \mid \Omega\right) \mathrm{d}\boldsymbol{\theta}, \tag{14}$$

In this notation, $T\left(\right)$ is defined as previously as a transform mapping independent elements from the real space of numbers into the $0-1$ interval. The function $H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)$ is used to generate a vector of length m of real numbers, as a function of the parameters contained in the vector $\boldsymbol{\theta}$ and the vector of individual-specific attributes $\boldsymbol{z_n}$, with $\boldsymbol{\theta}$ being distributed according to $f\left(\boldsymbol{\theta} \mid \Omega\right)$. This model can be seen to be an extension of the COVNL model described in Section 2.1 as follows. Let us assume that we have a model with a single structural parameter λ_m. It can be seen that, by specifying $T()$ to be the logistic transform, $H(\boldsymbol{z_n}, \boldsymbol{\theta})$ to yield $\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n}$, and setting $f(\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}) \mid \Omega) = 1$, the model reduces to the COVNL model. In this case, the parameters contained in the vector $\boldsymbol{\theta}$ are fixed across individuals. However, the model uses a top-down approach, which makes for easier adaptation in the case of multi-level structures or cross-nesting structures (see Section 2.1).

By removing the assumption that $f(\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}) \mid \Omega) = 1$, we obtain a model with random variation in the structural parameters across individuals. Depending on the specification of $f(\boldsymbol{\theta} \mid \Omega)$, only some of the elements in θ will be random, allowing for example for a random offset α across individuals, with purely deterministic variation on top of it, or a fixed offset point with random and deterministic variation on top of it, or both. Different choices for $H()$ and $T()$ (with appropriate domain conditions) lead to differences in model behaviour. Finally, it can be seen that by setting all elements in $\boldsymbol{z_n}$ to be zero, we obtain a model with purely random variation as in the first approach described in Section 2.2. This completes the extension of the COVNL framework to the case with random parameters.

2.3.2 Parameterisation of distributional parameters

We will base our derivation of the parameterisation method on the second approach described in Section 2.2, such that draws for λ_m are obtained directly from an appropriate distribution with an acceptable domain, as opposed to requiring the use of a transform (which is also possible). Let us assume that we have $\omega_m \in \Omega$, such that ω_m represents for example the mean used in the distribution function of structural parameter λ_m, with a corresponding variable $\sigma_m \in \Omega$ giving the dispersion parameter of the distribution of structural parameter λ_m. For now, let us assume that σ_m stays constant across individuals; extension to the case where it varies (deterministically across individuals) in addition to ω_m is straightforward. We now look at the case where some of the variation in λ_m is explained by random variation (through using the distribution $f(\lambda_m \mid \omega_m, \sigma_m)$ and some variation is explained by the attributes of the decision-maker, by parameterisation of ω_m. This approach acknowledges the fact that we can get an idea of the structural parameter of a given decision-maker with the help of individual-specific attributes (as in the COVNL model), but that there is some remaining error, or deviation, from this estimate. Specifying $\omega_{m,n}$ to be the mean value of the distribution of λ_m for decision-maker n, we can then simply use:

$$\omega_{m,n} = \alpha_{\omega_m} + \boldsymbol{\gamma_{\omega_m}}' \boldsymbol{z_n}, \tag{15}$$

where $\boldsymbol{z_n}$ represents a vector of attributes of decision-maker n, and α_{ω_m} and $\boldsymbol{\gamma_{\omega_m}}$ represent a constant and a vector of coefficients respectively, both of which are specific to the parameter ω_m.

In the case where no parameterisation of the parameters of the distributions is (or can be) used, only the constant α_{ω_m} will be estimated. In this case, $\omega_{m,n}$ stays the same across respondents, and the only differences in the value of λ_m across respondents are due to random variation. On the other hand, a model version that is very similar to the COVNL model can be obtained by only using one distributional parameter for each structural parameter, i.e. by setting

$$P\left(\lambda_{m,n} = \omega_{m,n} \mid \Omega\right) = 1 \tag{16}$$

This is equivalent to setting the dispersion term σ_m to be equal to zero. In this case, different structural parameters are still used for different individuals, but they no longer vary randomly across individuals; the variation is entirely deterministic. By further setting $\boldsymbol{\gamma_{\omega_m}} = 0$ for all m, the model reduces to the NL model.

2.3.3 Discussion

It is of interest to briefly discuss the differences between the two approaches. Both approaches attain the goal of jointly introducing deterministic and random covariance heterogeneity. The former approach has the advantage of easier interpretation, and possibly simplifies more easily to models with purely deterministic covariance heterogeneity, as well as models with fixed covariances. The only apparent advantage of the second approach is that it can avoid the need for additional transforms in the case where strictly bounded statistical distributions are used. Although, like the first approach, this variant also allows for an impact by an unlimited number of socio-demographic attributes, their impact needs to be gauged simultaneously for a minimum of two separate values, giving the mean and dispersion of the associated statistical distribution. As such, the former approach is probably preferable, although a detailed empirical comparison would be needed to reach a definitive answer.

3 Application

In this section, we present an application of one specific type of Mixed Covariance GEV model, namely a discrete mixture of a two-level NL model, with two possible levels of correlation in the population, leading to a Discrete Mixture Covariance NL (DM-COVNL) model. As such, the work presented here relates to the discussion on discrete mixture models by Hess, Bierlaire & Polak (2005*b*), where, in the present context, the mixture allows for covariance heterogeneity, as opposed to taste heterogeneity.

The justification for using the DM-COVNL model instead of a continuous mixture in this application is primarily a pragmatic one. Indeed, while it can simply be seen as a special case of a continuous mixture, it has the clear advantage of not requiring simulation in estimation. However, the discrete approach also has some advantages in terms of illustration of the differences with a homogoneuous covariance model, as well as having conceptual advantages in terms of the notion of an unobserved attribute leading to inter-alternative correlation for only part of the population of decision-makers.

Let the choice probability for alternative i and individual n in a model with K nests be given by:

$$P_n\left(i \mid \boldsymbol{\beta}\right) = \sum_{m_1=1}^{M_1} \cdots \sum_{m_K=1}^{M_K} P_n\left(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda} =< \lambda_1^{m_1}, \ldots, \lambda_K^{m_K} >\right) \cdot \pi_1^{m_1} \cdot \ldots \cdot \pi_1^{m_K}, \tag{17}$$

where the structural parameter λ_k, associated with the k^{th} nest, takes on M_k separate values, defined as λ_k^1 to $\lambda_k^{M_k}$, where each has an associated probability (or mass), with $0 \leq \pi_k^{m_k} \leq 1 \, \forall k, m_k$, and where $\sum_{m_k=1}^{M_k} \pi_k^{m_k} = 1 \, \forall k$. Here, in addition to the taste coefficients, estimates need to be produced for the different levels for all the structural parameters, as well as for the associated probabilities.

With the aim of illustrating the ability of the model to recover covariance heterogeneity, and to show the bias resulting from an inappropriate assumption with regards to covariance homogeneity, the application presented here makes use of quasi-simulated data, making use of data from an SP survey conducted to estimate the hypothetical demand for a new high-speed transit system in Switzerland; the Swiss Metro (c.f. Abay 1999, Bierlaire et al. 2001). This dataset provides us with *good* attribute-level data, avoiding the issues caused by the randomness in purely-simulated data. The choice-vectors used in the various case-studies are entirely independent of the original SP survey responses.

Three alternatives were included in the choice-set; car, rail and Swiss Metro (SM). Only three attributes, namely travel-time, travel-cost, and headway (for rail and SM) were used here. In each case-study, separate travel-time coefficients were used for the three modes ($\beta_{TT,car}$, $\beta_{TT,rail}$, and $\beta_{TT,SM}$), in conjunction with a common travel-cost coefficient (β_{TC}), a joint headway coefficient for rail and SM (β_{HW}), and two ASCs, for car and SM (δ_{car} and δ_{SM}). A sample of $9,000$ observations was used, based on an original sample of $3,000$ observations, where the data augmentation was based on small random variations of the original attribute levels. The generation of the data is based on the principle of nine observations per individuals, as opposed to a purely cross-sectional approach. Formally, in the generation of the data, the $1,000$ individuals were split into two groups. In the first group, representing 30% of the population, there is high correlation between the error-terms for the rail and SM alternatives, with a structural parameter equal to 0.3. In the remaining 70% of the population, the structure equates to a MNL model. The allocation to the two groups is performed on a purely random basis (taking into account the $30\% - 70\%$ split), such that a deterministic segmentation of the population cannot be used to account for the differences in correlation structure. This construction represents a situation in which, for example, for some individuals, an unobserved attribute leads to heightened substitution between rail and SM, while, for the remainder of the population it does not[6].

On the basis of the resulting individual-specific structural parameters, and the coefficient values reported for the *true* model in Table 1, the choice-probabilities for the three alternatives were calculated for each individual, on the basis of a two-level NL structure nesting rail with SM, where, for those individuals with $\lambda = 1$, the probabilities correspond to a MNL structure. A Monte-Carlo exercise was then used to determine the chosen alternative. As such, for each individual, the actual structural parameter applying for that respondent was used. This is more correct, and consistent with the underlying *true* model, than an approach which uses simulation over the two values, assigning to each individual the weighted choice probability across the two values for λ. As such, the resulting dataset reflects a real-world situation (in which a single value applies for each individual), rather than a DM-COVNL approximation to such a real-world situation. This in turn

[6]This could for example simply reflect an inherent dislike of car-travel for some respondents.

means that the estimation can show how well the DM-COVNL model, which does use a weighted average across the two values for λ, can replicate the *true* model[7].

Three separate models were estimated on the resulting dataset; a MNL model, a simple NL model nesting together rail and SM, and a DM-COVNL model. All three models were coded in Ox. In the DM-COVNL model, we estimate two distinct structural parameters for the rail-SM nest, specified as λ_a and λ_b. As such, with λ_a, the choice probability of rail in the t^{th} choice situation for individual n is given by:

$$P_{n,t}\left(rail \mid \lambda_a\right) = \frac{e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \cdot \frac{e^{\frac{V_{rail,n,t}}{\lambda_a}}}{e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}}, \tag{18}$$

where $V_{rail,n,t}$, $V_{SM,n,t}$ and $V_{car,n,t}$ give the observed utility for rail, SM and car respectively, for individual n, in choice situation t. The corresponding choice probabilities for SM and car are given by:

$$P_{n,t}\left(SM \mid \lambda_a\right) = \frac{e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \cdot \frac{e^{\frac{V_{SM,n,t}}{\lambda_a}}}{e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}}, \tag{19}$$

and

$$P_{n,t}\left(car \mid \lambda_a\right) = \frac{e^{V_{car,n,t}}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \tag{20}$$

On the basis of equations (18), (19) and (20), the probability of the observed sequence of choices for individual n, conditional on λ_a, is given by:

$$L\left(n \mid \lambda_a\right) = \prod_{t=1}^{T_n}\left[\delta_{n,t,rail} P_{n,t}\left(rail \mid \lambda_a\right) + \delta_{n,t,SM} P_{n,t}\left(SM \mid \lambda_a\right) + \delta_{n,t,car} P_{n,t}\left(car \mid \lambda_a\right)\right],$$

[7] It should be noted that similar issues apply in the case of continuous mixture models, where the use of a single draw from the assumed distribution of β for each individual in the generation of the simulated data is more consistent with the conceptual notion of a model with randomly varying tastes across respondents, than the use an actual mixture model in the simulation of the data (see also Hess, Bierlaire & Polak 2005*c*).

(21)

where T_n gives the number of choice-situations for respondent n (equal to 9 in this application), and where the dummy variable $\delta_{n,t,rail}$ is equal to 1 if respondent n chooses rail in the t^{th} choice-situation, and zero otherwise, with a corresponding definition for $\delta_{n,t,SM}$ and $\delta_{n,t,car}$.

With an equivalent notation in the case of the second structural parameter, λ_b, the contribution by individual n to the likelihood function is given by:

$$L(n) = \pi_{\lambda_a} L(n \mid \lambda_a) + \pi_{\lambda_b} L(n \mid \lambda_b), \tag{22}$$

where π_{λ_a} and π_{λ_b} give the mass for λ_a and λ_b respectively, with $0 \leq \pi_{\lambda_a} \leq 1$, $0 \leq \pi_{\lambda_b} \leq 1$, and $\pi_{\lambda_a} + \pi_{\lambda_b} = 1$. The fact that the weighting over the two support points occurs at the level of $L(n \mid \lambda_a)$ and $L(n \mid \lambda_b)$, rather than at the level of individual choice probabilities, reflects the notion that tastes stay constant across replications for the same individual.

Finally, on the basis of equation (22), the log-likelihood function for the DM-COVNL model used in this example is given by:

$$\begin{aligned} LL &= \ln\left(\prod_{n=1}^{N} L(n)\right) \\ &= \sum_{n=1}^{N} \ln\left[\pi_{\lambda_a} L(n \mid \lambda_a) + \pi_{\lambda_b} L(n \mid \lambda_b)\right], \end{aligned} \tag{23}$$

where N gives the total number of individuals, with, in the present application, $N = 1,000$.

The estimation results for the three models are summarised in Table 1, together with the coefficient values used in the generation of the data. The results show that the use of the NL model leads to statistically significant improvements in model fit over the MNL model, by 20.89 units, at the cost of one additional parameter. The DM-COVNL model leads to the best model fit overall, offering an improvement by 29.62 units over the NL model, with two additional estimated parameters (λ_b and π_{λ_a}). It should be said that, although statistically significant, these improvements are not dramatic, suggesting that the likelihood function is relatively unaffected by the treatment of correlation. Additionally, it can be seen that the results in terms of willingness-to-pay indicators are very similar across the three models. Indeed, the recovery of the *true* values is very good, and the dif-

	True model	MNL		NL		DM-COVNL	
Final LL	-	-7136.16		-7115.27		-7085.65	
Parameters	-	7		8		10	
adj. rho^2	-	0.2776		0.2796		0.2824	
		est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4	-4.3977	-34.31	-4.0194	-30.96	-3.9547	-30.77
δ_{SM}	-3	-3.5073	-33.80	-3.0582	-28.01	-3.0376	-28.51
β_{TC}	-0.1	-0.1082	-51.03	-0.0999	-42.17	-0.0994	-41.82
β_{HW}	-0.02	-0.0233	-32.06	-0.0205	-27.62	-0.0202	-28.27
$\beta_{TT,car}$	-0.03	-0.0331	-39.36	-0.0302	-33.17	-0.0300	-32.97
$\beta_{TT,rail}$	-0.04	-0.0446	-47.61	-0.0402	-37.32	-0.0399	-37.31
$\beta_{TT,SM}$	-0.035	-0.0382	-37.14	-0.0350	-33.32	-0.0347	-33.42
λ_a	1	-		0.78	7.52	1.00	0.00
λ_b	0.3	-		-		0.32	12.50
π_{λ_a}	0.7	-		-		0.71	9.96
Monetary value	CHF/hour	CHF/hour		CHF/hour		CHF/hour	
TT_{car}	18.00	18.36		18.13		18.11	
TT_{rail}	24.00	24.74		24.17		24.11	
TT_{SM}	21.00	21.21		21.05		20.96	
HW	12.00	12.95		12.33		12.22	

T-statistics for structural parameters calculated wrt 1

Table 1: Estimation results on mixed covariance data

ferences in bias are very small, where the lowest bias is obtained with the DM-COVNL model[8].

More significant differences however arise when looking at the implications in terms of correlation between the unobserved utility components for the rail and SM alternatives. The MNL model, by definition, offers no treatment of the correlation, and as such fails to allow for the heightened substitution between rail and SM. The simple two-nest NL model is based on the assumption of a homogeneous correlation structure. Here, the estimate produced for the unique nesting parameter in this model, at 0.78, is virtually indistinguishable from the weighted average of the two structural parameters present in the *true* population $(0.3 \cdot 0.3 + 0.7 \cdot 1.0 = 0.79)$. This result is consistent with a similar observation made in the case of discrete

[8]The fact that the bias decreases as we move from the MNL model to the NL model and on to the DM-COVNL model does suggest some interaction between observed and unobserved utility components, where a proper treatment of the unobserved utility components in the DM-COVNL model results in lower impact on the observed utility components.

	Original probabilities			Forecasted probabilities		
	Rail	SM	Car	Rail	SM	Car
True model	17.49%	33.01%	49.49%	4.07%	38.38%	57.55%
MNL	15.65%	30.76%	53.59%	3.15%	35.32%	61.53%
NL ($\lambda = 0.78$)	14.41%	32.41%	53.18%	2.17%	38.67%	59.16%
DM-COVNL ($\lambda = 1$)	17.38%	32.23%	50.39%	4.07%	37.42%	58.50%
DM-COVNL ($\lambda = 0.32$)	4.94%	35.06%	60.00%	0.03%	38.98%	60.98%
DM-COVNL (total)	13.73%	33.06%	53.21%	2.89%	37.88%	59.23%

	Relative change			Bias in predicted change		
	Rail	SM	Car	Rail	SM	Car
True model	-76.76%	+16.28%	+16.28%	-	-	-
MNL	-79.87%	+14.82%	+14.82%	+4.06%	-8.96%	-8.96%
NL ($\lambda = 0.78$)	-84.93%	+19.31%	+11.26%	+10.65%	+18.62%	-30.84%
DM-COVNL ($\lambda = 1$)	-76.55%	+16.10%	+16.10%	-0.27%	-1.07%	-1.07%
DM-COVNL ($\lambda = 0.32$)	-99.30%	+11.18%	+1.65%	+29.37%	-31.29%	-89.89%
DM-COVNL (total)	-78.96%	+14.57%	+11.32%	+2.87%	-10.48%	-30.46%

Table 2: Forecasting on mixed covariance data: representative individual with $\lambda_a = 1.0$ (observation 2,044)

mixture models for taste heterogeneity by Hess, Bierlaire & Polak (2005*b*), reflecting the fact that single parameter models yield estimates that are weighted averages of the actual values present in the population. It should be noted that, in the current example, with only two parameters, this approximation is made relatively easy, and more bias could be expected in the presence of more than two values for a parameter. Finally, the DM-COVNL is able to essentially perfectly recover the nature of the correlation structure in place in the *true* data; λ_a obtains a value equal to 1.0, while, for λ_b, the estimated value is very close to the true value of 0.3, with the difference being significant only at the 28% level. Similarly, the estimated shares for the two structural parameters, at $\pi_{\lambda_a} = 0.71$ and $\pi_{\lambda_b} = 0.29$ are indistinguishable from the true $70\% - 30\%$ split. In an actual application, where the true number of structural parameters present in the population is not known, and where it can in any case not be assumed to be finite, it would, after model estimation, be of interest to proceed with a posterior analysis, to produce the most likely structural parameter for each of the individuals. The same approach would be used in the case of a continuous mixture model. On the basis of the results from such an analysis, attempts could then be made to relate the correlation to socio-demographic attributes, and to use an appropriate segmentation in later forecasting applications.

	Original probabilities			Forecasted probabilities		
	Rail	SM	Car	Rail	SM	Car
True model	37.96%	25.03%	37.01%	20.86%	38.64%	40.50%
MNL	36.60%	33.01%	30.38%	29.22%	36.86%	33.92%
NL ($\lambda = 0.78$)	36.92%	32.33%	30.75%	28.63%	37.34%	34.03%
DM-COVNL ($\lambda = 1$)	38.13%	33.83%	28.05%	31.17%	37.63%	31.20%
DM-COVNL ($\lambda = 0.32$)	36.56%	25.01%	38.43%	20.61%	37.46%	41.93%
DM-COVNL (total)	37.67%	31.24%	31.09%	28.07%	37.58%	34.35%

	Relative change			Bias in predicted change		
	Rail	SM	Car	Rail	SM	Car
True model	-45.07%	+54.39%	+9.45%	-	-	-
MNL	-20.16%	+11.64%	+11.64%	-55.26%	-78.60%	+23.17%
NL ($\lambda = 0.78$)	-22.45%	+15.51%	+10.65%	-50.18%	-71.49%	+12.70%
DM-COVNL ($\lambda = 1$)	-18.25%	+11.24%	+11.24%	-59.51%	-79.33%	+18.98%
DM-COVNL ($\lambda = 0.32$)	-43.62%	+49.75%	+9.11%	-3.21%	-8.53%	-3.56%
DM-COVNL (total)	-25.47%	+20.29%	+10.47%	-43.47%	-62.69%	+10.80%

Table 3: Forecasting on mixed covariance data: representative individual with $\lambda_b = 0.3$ (observation 7,301)

In practice, posterior analyses of this nature are used very sparsely; in the absence of the resulting insight into the *actual* structural parameters, the mixture model, in this case the DM-COVNL model, would thus potentially be used directly in forecasting[9]. As such, it is of interest to compare the forecasting performance of the three models. To illustrate the differences in performance depending on the correlation structure in place in the *true* data, two representative individuals were selected, one belonging to the group with λ_a (respondent 228), and one belonging to the group with λ_b (respondent 812). In each case, a single observation was selected, where, for respondent 228, the first observation was used (observation $2,044$), while, for respondent 812, the second observation was used (observation $7,301$). The forecasting analysis looks at the changes in the choice probabilities for the three alternatives following an increase in the cost of rail-travel by 10%. The results of the forecasting exercise are summarised in Table 2 for observation $2,044$ and Table 3 for observation $7,301$. The bias measure used as an indicator of the correct recovery of the behaviour implied by the *true* model is defined as $\frac{\Delta - \Delta_T}{\Delta_T}$, where Δ gives the proportional change in choice-probability in the target model, and Δ_T gives the proportional change in choice-probability in the *true* model. In each case, the results for the DM-

[9] As opposed to using a posterior segmentation.

COVNL are split into three parts, showing the results for the part of the model that uses λ_a, the part of the model that uses λ_b, and the results for the combined model. This gives an idea of the gains in performance that could be expected if the forecasting exercise was preceded by a posterior analysis that was able to yield an appropriate segmentation, while also giving an account of the bias introduced by using the actual DM-COVNL, instead of its sub-parts.

The results for observation $2,044$ (Table 2) show a decrease in the choice probability of rail from 17.49% to 4.07%, following an increase in rail-fares by 10%[10]. The fact that λ_a is used for this individual implies an equal relative shift of probability towards SM and car. The same applies in the MNL model and the DM-COVNL sub-model with λ_a, resulting in the lowest bias for these two models, where the fact that the bias in the DM-COVNL sub-model is lower than in the MNL model (with the same treatment of correlation) can potentially be explained on the basis of more accurate estimates for the marginal utility coefficients. This is a result of the fact that the overall DM-COVNL model accounts for the correlation in the second subgroup, which the MNL model does not, where interaction between the observed and unobserved utility components leads to the bias in the estimates. The effects of the correlation structure become most visible when looking at the forecasts produced by the NL model, with $\lambda = 0.78$, and the DM-COVNL sub-model with $\lambda_b = 0.32$. Here, either approach leads to biased forecasts, by falsely indicating heightened substitution from rail to SM, where, due to the higher implied correlation, the bias is bigger in the DM-COVNL sub-model with λ_b than in the NL model. Here, it should also be noted that the DM-COVNL sub-model with λ_b significantly underestimates the original choice-probability for rail. Finally, the combined DM-COVNL model leads to lower bias than the NL model, where it should also be said that the DM-COVNL model performs quite well overall for the changes in the probability for rail and SM, with the only major bias, when compared to the MNL model[11], arising for the change in the probability of the car alternative.

The results for observation $7,301$ (Table 3) show a decrease in the choice probability of rail from 37.96% to 20.86%, following an increase in rail-fares by 10%. With individual 812 belonging to the 30% of the population with heightened correlation between rail and SM, the true model shows a much bigger relative shift from rail to SM than to car, a situation that is

[10] Lower decreases were observed at the population level (-35.89%), but the individual-observation results are used here, as they provide more insight into substitution patterns.

[11] Which has the clear advantage in this case in terms of the correct correlation structure.

recovered almost perfectly in the DM-COVNL sub-model using $\lambda_b = 0.32$. The MNL model wrongly predicts equal relative shifts in probability from rail to SM and from rail to car, where the same applies for the DM-COVNL sub-model using $\lambda_a = 1.0$. While the NL model correctly recovers the fact that there is a bigger than proportional shift towards SM than towards car, it underestimates the extent of the differences, through underestimating the correlation between the unobserved utility terms for rail and SM. The same occurs in the overall DM-COVNL model, where the underestimation is however less severe than in the NL model[12]. It should also be said that all models, except the DM-COVNL sub-model with $\lambda_b = 0.32$, significantly underestimate the decrease in the probability of the rail alternative, where this bias is however smallest in the overall DM-COVNL model, which also obtains the lowest overall bias out of the three full models.

In summary, this application has shown that the DM-COVNL model is able to recover the distribution of the covariance in the simulated dataset arbitrarily closely, while the simple NL model produces a weighted mean of the *true* values, on the basis of an assumption of covariance homogeneity. The forecasting application has also shown that the DM-COVNL model leads to lower bias than the NL model. Here, it should be noted that, in the special case described here, the MNL model performs well for the part of the population with no correlation between rail and SM, whereas it leads to significant bias in the remaining part of the population[13]. The fact that, in each case, the lowest bias is obtained by the appropriate DM-COVNL sub-model again illustrates the potential gains that could be obtained by conducting a posterior analysis to attempt to relate the difference in correlation structure to socio-demographic attributes with the aim of obtaining an appropriate segmentation for use in the actual forecasting exercise.

4 Summary and Conclusions

The aim of this paper was to extend the standard discrete choice modelling framework so as to allow for random variations in the covariance structure across respondents. The discussion in this paper has centred on the case of an underlying GEV model, and specifically, a two level NL model. The extension to other underlying GEV structures poses no major difficulties, as

[12] The shift from rail to SM is close to twice as big as the shift from rail to car, while, in the NL model, the ratio is below 1.5. In the true model, the ratio is close to 6.

[13] Much poorer overall performance would be obtained in the case where, in the *true* model, both structural parameters are inferior to 1, or if the share for $\lambda_a = 1$ was smaller.

described in the text, while the use of an alternative approach, based on an underlying ECL structure, is described in more detail in Appendix A.

The development of the Mixed Covariance GEV structure in this paper has shown how it is possible to allow jointly for random as well deterministic variations in the covariance structure across respondents. Additionally, it is possible, by adding an extra layer of integration, to allow for random taste heterogeneity, in addition to covariance heterogeneity. Here, it should also be noted that additional random terms can be added to allow for heteroscedasticity across alternatives, leading to additional dimensions of integration.

The application presented in Section 3 has described one special case of a Mixed Covariance GEV model, in which the mixture is discrete rather than continuous. The results have shown that the DM-COVNL structure is able to recover the covariance structure in place in the data very closely, and leads to lower bias in forecasting than the simple NL model, which is based on the assumption of a homogeneous covariance structure.

Much work remains to be done, including the development of more sophisticated mixed covariance structures, the testing of continuous mixture structures on simulated data, and the use of discrete and continuous mixture structures with real data. Here, it should be noted that the discussion in this paper has focussed primarily on variations in the extent of correlation across respondents, rather than variations in the actual correlation structure. The latter applies for example in the case where, for individual A, there is correlation between alternatives 1 and 2, while, for individual B, there is correlation between alternatives 2 and 3. Such variations in the actual structure can, in the absence of an appropriate segmentation, be accommodated in a cross-nesting framework, with the variation in structure accounted for primarily through variations in the allocation parameters.

In closing, it should be said again that mixed covariance models should in part be seen as an explanatory tool, which, unlike other models, have the power to highlight the presence of variations in the covariance across respondents. On the basis of such results, the modeller can then attempt to refine the model to accommodate some covariance heterogeneity in a deterministic fashion, either through a segmentation of the data, or by parameterising the covariance structure, as described by Bhat (1997), potentially with additional random covariance heterogeneity, as described in Section 2.3. If such attempts at a deterministic approach fail, it is still desirable, for interpretation as well as forecasting reasons, to try to link the variations to socio-demographic information through a posterior analysis[14]. However, if this is not possible, then it is clearly preferable to account for the variation in a random way (in interpretation as well as forecasting), as opposed to maintaining the assumption of covariance homogeneity. Either way, the modelling approach described in this paper is thus a valuable tool for the analysis of choice behaviour.

References

Abay, G. (1999), *Nachfrageabschaetzung Swissmetro: Eine stated-preference Analyse (in German)*, Berichte des Nationalen Forschungsprogrammes 41: Verkehr und Umwelt, Bericht F1, EDMZ, Bern.

Bhat, C. R. (1997), 'Covariance Heterogeneity in Nested Logit models: economic structure and application to inter-city travel', *Transportation Research Part B: Methodological* **31**(1), 11–21.

Bhat, C. R. (1998), 'Accommodating flexible substitution patterns in multidimensional choice modeling: formulation and application to travel mode and departure time choice', *Transportation Research Part B: Methodological* **32**(7), 425–440.

Bhat, C. R. (2000), 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation Science* **34**(2), 228–238.

Bhat, C. R. & Castelar, S. (2002), 'A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area', *Transportation Research Part B: Methodological* **36**(7), 593–616.

Bhat, C. R. & Guo, J. (2004), 'A mixed spatially correlated Logit model: formulation and application to residential choice modeling', *Transportation Research Part B: Methodological* **38**(2), 147–168.

Bierlaire, M. (2002), *The Network GEV model*, Proceedings of the 1^{st} Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.

[14]Here, it should be said that the same reasoning applies in the case of mixture models looking for taste heterogeneity; again, a deterministic treatment is clearly preferable for interpretation as well as forecasting reasons, and the mixture model can thus be seen as an explanatory tool.

Bierlaire, M. (2005), *A theoretical analysis of the cross-nested Logit model*, Annals of Operations Research, forthcoming.

Bierlaire, M., Axhausen, K. W. & Abay, G. (2001), *The acceptance of modal innovation : The case of Swissmetro*, Proceedings of the 1^{st} Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.

Boyd, J. & Mellman, J. (1980), 'The effect of fuel economy standards on the u.s. automotive market: a hedonic demand analysis', *Transportation Research Part A: General* **14**(5-6), 367–378.

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

Cardell, S. & Dunbar, F. (1980), 'Measuring the societal impacts of automobile downsizing', *Transportation Research Part A: General* **14**(5-6), 423–434.

Chu, C. A. (1989), *Paired Combinatorial Logit Model for Travel Demand Analysis*, Proceedings of the fifth World Conference on Transportation Research, Yokahama, Japan.

Daly, A. (2001), *Recursive Nested Extreme Value Model*, Working Paper 559, ITS, University of Leeds, Leeds, U.K.

Daly, A. & Bierlaire, M. (2005), *A general and operational representation of GEV models*, Transportation Research Part B: Methodological, forthcoming.

Daly, A. & Zachary, S. (1978), Improved multiple choice models, *in* D. Hensher & Q. Dalvi, eds, 'Identifying and Measuring the Determinants of Mode Choice', Teakfields, London.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*a*), Capturing taste heterogeneity and correlation structure with Mixed GEV models, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, pp. 55–76.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*b*), *Discrete mixture models with applications to the estimation of value of travel-time savings*, paper presented at the 5^{th} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*c*), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236.

Hess, S. & Polak, J. W. (2004), *Mixed Logit estimation of parking type choice*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. & Polak, J. W. (2005*a*), *Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area*, CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. & Polak, J. W. (2005*b*), 'Mixed logit modelling of airport choice in multi-airport regions', *Journal of Air Transport Management* **11**(2), 59–68.

Hess, S., Polak, J. W., Daly, A. & Hyman, G. (2005), *Flexible Substitution Patterns in Models of Mode and Time of Day Choice: New evidence from the UK and the Netherlands*, Transportation, forthcoming.

Hess, S., Train, K. & Polak, J. W. (2005), *On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice*, Transportation Research Part B: Methodological, forthcoming.

Koppelman, F. S. & Sethi, V. (2000), Closed form discrete choice models, *in* D. Hensher & K. Button, eds, 'Handbook of Transport Modeling', Pergamon Press, Oxford.

Koppelman, F. S. & Sethi, V. (2005), 'Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior', *Transportation Research Part B: Methodological* **39**(9), 825–853.

Luce, R. (1959), *Individual choice behavior: a theoretical analysis*, J.Wiley and Sons, New York.

Marschak, J. (1960), Binary choice constraints on random utility indications, *in* K. Arrow, ed., 'Stanford Symposium on Mathematical Methods in the Social Science', Stanford University Press, Stanford, CA, pp. 312–329.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York.

McFadden, D. (1978), Modelling the choice of residential location, *in* A. Karlquist, ed., 'Spatial Interaction Theory and Planning Models', North Holland, Amsterdam, chapter 25, pp. 75–96.

McFadden, D. & Train, K. (2000), 'Mixed MNL Models for discrete response', *Journal of Applied Econometrics* **15**, 447–470.

Papola, A. (2004), 'Some developments on the cross-nested Logit model', *Transportation Research Part B: Methodological* **38**(9), 833–854.

Small, K. (1987), 'A discrete choice model for ordered alternatives', *Econometrica* **55**(2), 409–424.

Swait, J. & Adamowicz, W. (1996), *The Effect of Choice Environment and Task Demands on Consumer Behavior: Discriminating between Contribution and Confusion*, Working Paper, Department of Rural Economy, University of Alberta, Edmonton, Canada.

Swait, J. & Adamowicz, W. (2001), 'Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice', *Organizational Behavior and Human Decision Processes* **86**(2), 141–167.

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

Vovsha, P. (1997), 'Application of a Cross-Nested Logit model to mode choice in Tel Aviv, Israel, Metropolitan Area', *Transportation Research Record* **1607**, 6–15.

Vovsha, P. & Bekhor, S. (1998), 'The Link-Nested Logit model of route choice: overcoming the route overlapping problem', *Transportation Research Record* **1645**, 133–142.

Walker, J. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, PhD thesis, MIT, Cambridge, MA.

Wen, C.-H. & Koppelman, F. S. (2001), 'The Generalized Nested Logit Model', *Transportation Research Part B: Methodological* **35**(7), 627–641.

Williams, H. C. W. L. (1977), 'On the Formulation of Travel Demand Models and Economic Evaluation Measures of User Benefit', *Environment & Planning A* **9**(3), 285–344.

A Appendix: Development of ECL approach

We now describe how the ECL formulation of the MMNL model can be adapted to allow for covariance heterogeneity. We first review the basic theory behind the ECL model (Section A.1) and show how it can be used to approximate the COVNL model (Section A.2). We then proceed to the case where the covariance heterogeneity is purely random (Section A.3), and to the case where part of the variation is deterministic with a remaining random part (Section A.4).

A.1 General ECL formulation

In the ECL model, correlation across alternatives is introduced through the use of error-components that are shared between alternatives that are closer substitutes for each other. The error-components take on the form of Normally-distributed random variables with a mean of zero, and a standard deviation of σ, where the estimate for σ is related to the correlation between the alternatives.

Ignoring for the moment the issues of identification discussed by Walker (2001), and the question of homoscedasticity[15], the utilities of two alternatives that have some correlation in the unobserved part of utility would be written as:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n} + \zeta_1 \tag{24}$$

and

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \zeta_1, \tag{25}$$

where $V_{i,n}$ and $V_{j,n}$ give the observed part of utility for alternatives i and j and respondent n, and $\varepsilon_{i,n}$ and $\varepsilon_{j,n}$ are *iid* type I extreme-value terms.

[15] Basic ECL approximations to GEV models are heteroscedastic, while GEV models are homoscedastic, an issue that can be addressed by *cancelling* out the heteroscedasticity in ECL models through the use of additional error-components.

The additional error-term ζ_1 is distributed $N(0, \sigma_1)$. With this, the covariance between the two alternatives is given by σ_1^2, while the variance for the individual utilities is given by $\sigma_1^2 + \frac{\pi^2}{6}$, leading to a correlation of:

$$\rho^2 = \frac{\sigma_1^2}{\sigma_1^2 + \frac{\pi^2}{6}}. \tag{26}$$

It is easy to see that it is possible to rewrite the utility of alternative j as:

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sigma_1 \xi_1, \tag{27}$$

where $\xi_1 \sim N(0,1)$, and where the subscript on ξ remains in use to guarantee that individual draws are taken for each error-component (with the same draws taken for the same error-component across alternatives).

For the choice-probabilities, integration over the $N(0,1)$ draws for the error-components is required. Let Ψ_j define the set of error-components included in the utility function of alternative j, such that:

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sum_{k \in \Psi_j} \sigma_k \xi_k \tag{28}$$

This notation allows for any structure for the error-components, including homoscedastic as well as heteroscedastic ones. The choice probability for alternative i and individual n is now given by:

$$P_n(i \mid \boldsymbol{\sigma}) = \int_{\xi_1} \ldots \int_{\xi_K} \left[\frac{exp\left(V_{i,n} + \sum_{k \in \Psi_i} \sigma_k \cdot \xi_k\right)}{\sum_{j \in C_n} exp\left(V_{j,n} + \sum_{l \in \Psi_j} \sigma_l \cdot \xi_l\right)} \cdot \prod_{k=1}^{K} \phi(\xi_k) \right] \mathrm{d}\xi_K \ldots \mathrm{d}\xi_1, \tag{29}$$

where K gives the total number of error-components used, and $\phi()$ is the standard Normal density function.

A.2 Adapting the ECL formulation for deterministic covariance heterogeneity

The ECL formulation can be extended straightforwardly to allow for deterministic covariance heterogeneity by parameterising σ_k, for example by setting $\sigma_k = f(\boldsymbol{\theta}, \boldsymbol{z_n})$, where $\boldsymbol{\theta}$ is a vector of parameters, and where $\boldsymbol{z_n}$ is defined as before. The only condition applying to $f()$ is that it yields positive values for the standard deviations[16]; equation (26) guarantees that the resulting correlation falls between 0 and 1.

[16]This merits some clarification. Estimation code can deal with negative values for standard deviation parameters in the case where they are only used in the form of variances

A.3 Adapting the ECL formulation for purely random covariance heterogeneity

In the standard ECL formulation of the MMNL model, the choice probabilities are obtained by integration over the distribution of the error-components, with additional integration over the distribution of random taste-coefficients in the case of added random taste heterogeneity. Focussing for now on the case of error-components for correlation only (as opposed to additional taste heterogeneity), random covariance heterogeneity can be introduced by additional integration over the distribution of the variances of the error-components.

The choice probability is in this case given by:

$$P_n\left(i\right) = \int_{\sigma_1} \dots \int_{\sigma_K} \left[P_n\left(i \mid \boldsymbol{\sigma}\right) \cdot \prod_{k=1}^{K} g\left(\sigma_k \mid \boldsymbol{\theta_k}\right) \right] \mathrm{d}\sigma_K \dots \mathrm{d}\sigma_1, \tag{30}$$

where $P_n\left(i \mid \boldsymbol{\sigma}\right)$ is the choice probability for alternative i, conditional on the vector of standard deviations $\boldsymbol{\sigma}$, as in equation (29), and where $g\left(\sigma_1 \mid \boldsymbol{\theta_1}\right)$ is the density function for σ_1, with parameters given by the vector $\boldsymbol{\theta_1}$. Here an appropriate choice of distribution for the standard deviations is of crucial importance, given that they need to take on positive values[17]. An alternative to the use of bounded distributions comes is to use a transform mapping monotonically from the real domain to the space of positive numbers. The adaptation of equation (30) to this case is straightforward.

A.4 Adapting the ECL formulation for joint deterministic and random covariance heterogeneity

The extension of the approach described in Section A.3 to the case allowing jointly for deterministic and random covariance heterogeneity is straightforward. We reuse the formulation from Section A.2, where $\sigma = f\left(\boldsymbol{\theta}, \boldsymbol{z_n}\right)$. This time however, we allow some of the elements of $\boldsymbol{\theta}$ to be randomly distributed across individuals. The choice probability for alternative i and

as opposed to standard deviations; in fact, in unconstrained estimation, it can often be observed that estimation packages produce *negative* estimates for the standard deviations. The problems arise in the case where $f\left(\right)$ allows for positive as well as negative values for σ depending on the values of $\boldsymbol{z_n}$, leading to an underestimated mean level of correlation.

[17] Again, this requirement is used solely to avoid an underestimation of the mean level of correlation in the case where the distribution yields positive as well as negative estimates for σ.

decision-maker n is now rewritten as:

$$P_n(i) = \int_{\boldsymbol{\theta_1}} \cdots \int_{\boldsymbol{\theta_K}} \left[P_n\left(i \mid \sigma_k = f\left(\boldsymbol{\theta_k}, \boldsymbol{z_n}\right) \forall k\right) \cdot \prod_{k=1}^{K} g\left(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k}\right) \right] \mathrm{d}\boldsymbol{\theta_K} \ldots \mathrm{d}\boldsymbol{\theta_1}, \tag{31}$$

where $\boldsymbol{\theta_k}$ is distributed according to $g\left(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k}\right)$, where the notation allows for correlation between individual elements in $\boldsymbol{\theta_k}$. It can easily be seen that this approach reduces to the purely random formulation in Section A.3, if those parameters associated with z_n are zero[18], and the purely deterministic formulation in Section A.2, in the case where $g\left(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k}\right)$ produces only a single (fixed) value for the vector $\boldsymbol{\theta_k}$.

A.5 Discussion

The discussion presented here has shown how the ECL framework can be adapted to allow for deterministic as well as random covariance heterogeneity. In practice, it should be said that, due to the additional dimensions of integration, the mixed covariance ECL approach is generally more expensive in estimation and application than its GEV based counterparts described in the main part of this paper, albeit that it has the advantage of a simpler form for the integrand (MNL vs more general GEV). An additional issue however arises with regards to identification, where appropriate conditions on identifiability need to be worked out on a case-by-case basis.

[18] I.e., only a constant is estimated, which is distributed randomly across respondents.