Paper:
Hess, S, Bolduc, D and Polak, JW (2010) *Random covariance heterogeneity in discrete choice models.* Transportation, 37 (3). 391 - 411

http://dx.doi.org/10.1007/s11116-009-9255-3

Random covariance heterogeneity in discrete choice models

Stephane Hess* Denis Bolduc† John W. Polak‡

January 26, 2008

Abstract

In this paper, we extend the standard discrete choice modelling framework by allowing for random variations 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 of 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. An application on Stated Preference data for the choice of departure time and travel mode shows that important gains in model performance can be obtained by allowing for random covariance heterogeneity. Furthermore, the approach leads to significant differences in the implied willingness to pay measures, and the substitution patterns between alternatives.

1 Introduction

Discrete choice models have been used extensively in various areas of behavioural research, notably transport studies, for over thirty years. Initially, virtually all applications were based on the Multinomial Logit (MNL) model (cf. McFadden, 1974). Although easy to specify, estimate and apply, the MNL model has significant disadvantages in terms of 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

most notably in the form of very restrictive substitution patterns across alternatives. Initial gains in flexibility were made by the development of structures belonging to the family of Generalised Extreme Value (GEV) models (McFadden, 1978), such as the Nested Logit (NL) model (Williams, 1977; Daly and Zachary, 1978)[1]. These models nest together alternatives that are closer substitutes for each other, with heightened correlation (and hence cross-elasticities) between nested alternatives.

Recently, researchers have begun to increasingly exploit the power of an alternative model form, the Mixed Multinomial Logit (MMNL) model. MMNL 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. The MMNL model is primarily used for the representation of random taste heterogeneity, in a Random Coefficients Logit (*RCL*) framework. However, it can also be exploited to allow for inter-alternative correlation and heteroscedasticity, in an Error Components Logit (*ECL*) framework. For more details on the two specifications, see Ben-Akiva and Bolduc (1996), McFadden and Train (2000), Walker (2001) and Train (2003). Both approaches can be combined straightforwardly for the joint representation of random taste heterogeneity and inter-alternative correlation. Here, an alternative approach is to use integration of GEV-style choice probabilities over the distribution of taste coefficients, leading to a more general GEV mixture model, of which MMNL is simply the most basic example (cf. Bhat and Guo, 2004; Hess *et al.*, 2005a). Recently, researchers have also adapted the *standard* formulation of the MMNL model by linking the parameters of the distribution of random terms to characteristics of the respondents; see for example Greene *et al.* (2006).

While the developments in relation to closed-form GEV as well as GEV mixture models have led to gradual gains in modelling flexibility, little effort has gone into the development of model forms allowing for a representation of heterogeneity across respondents in the correlation structure (and hence substitution patterns) 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. Although some of the covariance heterogeneity can conceivably be accommodated through an appropriate segmentation of the population, it is likely that in general, some within-segment heterogeneity remains. The existing literature seems to contain only two examples of a model allowing for such heterogeneity. The first

[1] See also Ortúzar (2001) for a history of the NL structure.

of these comes in the form of the Covariance Nested Logit (COVNL) model discussed by Bhat (1997), which allows for a parameterisation of the nesting parameters, and hence the substitution patterns in a simple Nested Logit (NL) model. Koppelman and Sethi (2005) later expand this approach by incorporating covariance heterogeneity in the more advanced Generalised Nested Logit (GNL) model, where they additionally allow for heteroscedasticity across respondents through a parameterisation of the scale factor, describing the resulting model as the Heterogeneous Generalised Nested Logit (HGNL) model.

While it is highly desirable to explain any covariance heterogeneity in a deterministic way, this is clearly not always possible. 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; the development of a corresponding framework based on an ECL structure 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 the findings of an application comparing our Mixed Covariance (MCOV) model to structures assuming covariance homogeneity. Finally, Section 4 presents the conclusions of the research.

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 several parts. We first look at the general model form, in Section 2.1, before moving on to the cases of purely random variation in Section 2.2, and combined deterministic and random variation in Section 2.3. Finally, Section 2.4 briefly looks at the case where a discrete mixture approach is used instead of a continuous mixture.

2.1 General model form

In a two-level NL model, the choice probability of alternative i (belonging to nest m) for individual n is given by:

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

with logsum term

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

where $V_{j,n}$ corresponds to the systematic part of the 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 COVNL model of Bhat (1997) expands on the standard NL model, by parameterising the structural parameters λ as:

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

where α_m is a constant, $\boldsymbol{z_n}$ is a vector of attributes of decision maker n, and where $\boldsymbol{\gamma_m}$ is a vector of coefficients. In this notation, $\lambda_{m,n}$ is the structural parameter for nest m and decision maker n. Both α_m and $\boldsymbol{\gamma_m}$ are to be estimated. The subscript on the parameters makes them specific to a given nesting parameter.

To ensure consistency with utility maximisation, $F()$ needs to be specified so as to produce values in the $[0-1]$ interval[2]. Furthermore, Bhat (1997) states that increases in $\boldsymbol{z_n}$ should have a monotonic effect on $\lambda_{m,n}$ (where this ensures consistency in the case of multi-level structures, cf. equation (7)). This double requirement can be satisfied by using a function $F()$ with:

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

[2]It should be noted that this constraint can cause complications as it can prevent problems with the data or model specification from manifesting themselves.

These conditions are met by the use of a continuous cumulative probability distribution function, where Bhat (1997) suggests the use of the logistic distribution. Here, it is important to note that this does not equate to a random formulation in which the nesting parameters follow a logistic distribution. Here, the logistic distribution is simply used to translate from an infinite space to a $[0-1]$ interval.

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 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 $\boldsymbol{\lambda}$.

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 behaviour of the resulting 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 (cf. Train, 2003). This issue is discussed in more detail in the description of the two special cases in Sections 2.2 and 2.3. However, it can already be mentioned at this point that special care is required in the imposition of a priori bounds on the distribution as this can prevent data or misspecification issues from manifesting themselves[4].

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 model, this is made possible by

[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.

[4]This issue is similar to that discussed by Hess *et al.* (2005b) in the context of the specification of random taste heterogeneity.

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:

$$\lambda_{m,n} = F\left[\left(\alpha_l + \boldsymbol{\gamma_l}'\boldsymbol{z_n}\right) + G\left(\delta_m + \boldsymbol{\eta_m}'\boldsymbol{w_n}\right)\right], \tag{7}$$

where $\boldsymbol{w_n}$ is an additional vector of individual characteristics, which can be the same as $\boldsymbol{z_n}$, and where δ_m and $\boldsymbol{\eta_m}$ are a constant and vector respectively that need to be estimated. Finally, $G\left(\right)$ 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 (as in the approach taken by Bhat 1997), 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

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

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. Extension of this theory to models with more than three levels is straightforward. Extensions to models allowing for cross-nesting are also possible, although slightly more complex. 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. Again, special care is required when imposing a priori bounds on parameters or constraints on the relationship between parameters, as it is important to still allow the effects of structural misspecifications to manifest themselves in the results. In the present context, misspecification would for example become apparent when the majority of the mass for $\widehat{\lambda_{li}}$ is close to 1, suggesting that a reversal of the order of nesting may be appropriate.

The final step in the theoretical development of our proposed model form is the representation of taste (as opposed to covariance) heterogeneity across individuals. 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 $\boldsymbol{\beta}$ over the assumed distribution of the taste coefficients. This comes in addition to the integration over the distribution of the structural parameters.

Let $P_n\left(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}\right)$ 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:

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

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

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

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.

Before proceeding with a detailed discussion of various specific formulations of our proposed model structure, a word should be said about parameter identification. With the formulation using a specific GEV model as the integrand, the only difference between our model and a typical GEV mixture model is that the nesting parameters are given a random treatment as opposed to the taste parameters. Similarly, the model can be seen as an extension of the COVNL model of Bhat (1997) with the only difference being a random as opposed to deterministic treatment of covariance heterogeneity. From either basis, there seem to be no reasons for additional identifiability conditions[5]. In estimation, the model was also found to be well behaved. It is possible that issues with confounding may arise between random taste heterogeneity and random covariance heterogeneity (as in Equation 12), but such issues can also arise when allowing jointly for inter-alternative correlation and random taste heterogeneity (cf. Hess *et al.*, 2005a).

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:

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

where $T\left(\boldsymbol{x}\right)$ 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 $f\left(\boldsymbol{x} \mid \Omega\right)$, and an appropriate transform can be used for $T\left(\boldsymbol{x}\right)$. The issue with this approach is that it imposes a priori bounds on the distribution of λ, such that data or model misspecification issues can be difficult to detect. As such, special care is required in the interpretation of the results when using this approach[6].

The second approach avoids the use of the additional transform $T\left(\boldsymbol{x}\right)$, and draws for the structural parameters are produced directly from the function $f\left(\boldsymbol{\lambda} \mid \Omega\right)$, as shown in equation (5). In this case, the condition

[5] The same is not necessarily the case when basing the model on a ECL structure, as discussed in Appendix A.

[6] The same issue arises in the COVNL model discussed by Bhat (1997) through the use of a transform such as the logistic distribution.

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. However, the direct estimation of the bounds on the distribution allows modellers to establish consistency with utility maximisation directly, by testing whether the range for λ falls within the $0-1$ interval. 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 more advanced options such as 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 can potentially offer more flexibility in the shape of the distribution of the structural parameters. The issue of a priori bounds also disappears. However, it should also be noted that models using flexible distributions with estimated bounds, such as the Johnson S_B, are often very difficult to estimate.

2.3 Model with deterministic and random covariance heterogeneity

It is clearly desirable to explain as much covariance heterogeneity as possible in a deterministic manner, as in the COVNL model of Bhat (1997). However, even the most comprehensive deterministic treatment will leave residual heterogeneity, which can be accounted for using the Mixed Covariance models presented earlier in this paper. Clearly, the most attractive general approach is to formulate a model that can account for both deterministic and random effects, such that for example $\lambda = F(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n} + \epsilon)$, where ϵ is a random component. Two approaches are possible in this case, one is to use a mixed version of a COVNL-style formulation (but within a top-down approach), while the other is to use a functional form for the parameters of the distribution employed 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:

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

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. It can be seen that, by specifying $T\left(\right)$ to be the logistic transform, $H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)$ to yield $\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n}$, and setting $f\left(\boldsymbol{\theta} = \left(\alpha, \boldsymbol{\gamma}\right) \mid \Omega\right) = 1$, 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 or cross-nesting structures (see Section 2.1).

By removing the assumption that $f\left(\boldsymbol{\theta} = \left(\alpha, \boldsymbol{\gamma}\right) \mid \Omega\right) = 1$, we obtain a model with random variation in the structural parameters across individuals. Depending on the specification of $f\left(\boldsymbol{\theta} \mid \Omega\right)$, 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 $H\left(\right)$ and $T\left(\right)$, 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 standard deviation term $\sigma_m \in \Omega$ for 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. 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[7]. In the case where no such parameterisation 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.

2.4 Discrete mixture approach

Thus far, we have looked solely at the case where the covariance heterogeneity is represented through a continuous mixture, i.e., by using integration over the assumed continuous distribution of the structural parameters. However, it is similarly possible to work on the basis of a discrete mixture approach, such that the structural parameters take on only a finite number of values across the population, meaning that the integration is replaced by a weighted summation. This relates to the case of discrete mixture models for random taste heterogeneity, as discussed for example by Hess *et al.* (2007a). In some cases, the justification for using a discrete mixture approach will be primarily a pragmatic one. Indeed, while it can simply be seen as an approximation to a continuous mixture, it has the clear advantage of not requiring simulation in estimation. However, the use of a discrete mixture approach can potentially also give greater freedom in terms of the shape of the distribution. Finally, the discrete approach also has some advantages in terms of interpretation, by providing modellers with some guidelines for a deterministic segmentation of the population.

[7] The interaction between $\boldsymbol{\gamma_{\omega_m}}$ and $\boldsymbol{z_n}$ could also be non-linear.

In a discrete mixture model for covariance heterogeneity, the choice probability for alternative i and individual n in a model with K nests would be given by:

$$P_n\left(i \mid \boldsymbol{\beta}\right) = \sum_{m_1=1}^{M_1} \dots \sum_{m_K=1}^{M_K} P_n\left(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda} =< \lambda_1^{m_1}, \dots, \lambda_K^{m_K} >\right) \cdot \pi_1^{m_1} \cdot \dots \cdot \pi_1^{m_K}, \tag{16}$$

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. For an application of a discrete mixture model for covariance heterogeneity, see Hess (2005).

3 Application

This section summarises the findings of an empirical study conducted to compare our MCOV model to model structures assuming covariance homogeneity. We first present the estimation data in Section 3.1. We then discuss model specification in Section 3.2, before presenting the main estimation results in Section 3.3. Finally, Section 3.4 presents two forecasting applications.

3.1 Data

The analysis makes use of time of day (TOD) data collected in 2000 for the development of the Dutch National Model System (cf. de Jong *et al.*, 2003). An initial Revealed Preference (RP) survey was conducted to select a sample of respondents for the follow up Stated Preference (SP) survey. The RP survey included rail travellers as well as car travellers, contacted at a selection of sites across The Netherlands, concentrating on areas where road and rail congestion was encountered in peak period journeys. In the ensuing SP survey, respondents were presented with four alternatives in each choice situation, three alternatives using their current mode of travel, and one alternative involving a change of mode. The choice thus involves the retiming of an existing tour or the switch of this tour to an alternative mode. The three alternatives on the current mode are described as "retimed earlier", "base", and "retimed later", with the "base" alternative being close in departure time to the actual observed tour. Respondents were presented with 16 SP replications, spread over two games. The travel purposes distinguished were commuting, business, education, and other. This dataset has recently been used in the estimation of ECL and NL models by Hess *et al.* (2007c) and Hess *et al.* (2007b) respectively, while Hess *et al.* (2005c) make use of the data in a time period choice model looking at different ways of specifying the time period constants. In the present analysis, we make use of $2,720$ observations collected from 170 car commuters, who were presented with three car alternatives (base, early and late) and one public transport (rail) alternative.

3.2 Model specification

Four different models were estimated on the TOD data, a simple MNL model, a NL model with a homogeneous correlation structure, and two different MCOV models. In the models allowing for a nesting structure, the three car alternatives were nested against the public transport alternative. Attempts to retrieve covariance heterogeneity in a deterministic fashion (such as in the model of Bhat 1997) were unsuccessful, while accounting for random taste heterogeneity was beyond the scope of the analysis. A linear in parameters specification was used for all models, with a utility specification based largely on the work of Hess *et al.* (2005c) and Hess *et al.* (2007c).

No further details are required for the MNL and NL models. For the two MCOV models, a specification falling into the group described in Section 2.2 was used, i.e., a purely random approach used in conjunction with a transform mapping to the $0-1$ interval. Specifically, we use a logistic distribution, such that $T(x) = \frac{1}{1+e^{-x}}$, and use two different choices for the distribution of x. In the first model, MCOV_U, we set $f(x) \sim U[x_a, x_a + x_b]$, while, in the second model, MCOV_N, we set $f(x) \sim N(x_\mu, x_\sigma)$. Attempts at using other choices for $f(x)$ led to very similar results, as did attempts at using other formulations for $T(x)$. Finally, some attempts were also made to use a distribution with estimated bounds and without the use of an additional transform $T(x)$, hence not imposing a priori bounds on the distribution of λ, and allowing data issues or model misspecification problems to manifest themselves. However, these attempts were unsuccessful, with the estimated bounds not being significantly different from 0 and 1 respectively, suggesting that the imposition of the $0-1$ bounds is in this case acceptable. Here, it should be noted that the estimation of a model using Johnson's S_B distribution with bounds fixed at 0 and 1 is equivalent to model MCOV_N, i.e., using a logistic transform of a normally distributed variable.

All models were coded in Ox 4.2 (Doornik, 2001), where the two MCOV models were specified so as to acknowledge the repeated choice nature of the data, by carrying out the integration (respectively simulation) at the level of the probabilities for sequences of choices for a given individual, rather than the level of individual choices.

3.3 Estimation results

The estimation results for the four different models are summarised in Table 1, where the specification for the utility function is the same across the four models. The results show significant negative impacts of increases in travel time ($\beta_{\text{TT,car}}$, $\beta_{\text{TT,PT}}$) and travel cost ($\beta_{\text{TC,car}}$, $\beta_{\text{TC,PT}}$), for both car and public transport (PT). Estimates for the effect of increases in frequency for public transport are positive, but not significantly different from zero at reasonable levels of confidence.

In addition to the marginal utility coefficients for travel time, travel cost and frequency, three sets of dummy variables were estimated, associated with shifts away from the base alternative to either of the two retimed car alternatives or the PT alternative. The three generic dummy variables (δ_{early}, δ_{late}, and δ_{PT}) are all negative and significantly different from zero. There are additional effects for respondents who regularly work from home and respondents in part-time employment. The former are less likely than the overall population to choose the early alternative[8] or the PT alternative[9], but are more likely to choose a late departure[10]. For part-time workers, there is an additional negative penalty for both retimed alternatives, along with a positive additional dummy[11] for the PT alternative[12].

Additional parameters were estimated for the three remaining models to account for the correlation between the three car alternatives. In the NL model, we obtain a value of 0.5 for the structural parameter λ, which is significantly different from 0 and 1 at high levels of confidence. This implies a high correlation of 0.75 between the error terms for the three car alternatives. Both MCOV models retrieve significant levels of heterogeneity for the structural parameter λ, where, before transformation using the logistic distribution $\frac{1}{1+e^{-x}}$, MCOV$_{\text{U}}$ has x distributed uniformly between -2.324 and

[8] $\delta_{\text{early,regularly work from home}}$ is not different from zero beyond the 89% level of confidence.

[9] $\delta_{\text{PT,regularly work from home}}$ is not different from zero beyond the 77% level of confidence.

[10] However, the combined effect of δ_{late} and $\delta_{\text{late,regularly work from home}}$ is still negative.

[11] $\delta_{\text{PT,part-time work}}$ is significant only at the 76% level.

[12] Where the combined effect of δ_{PT} and $\delta_{\text{PT,part-time work}}$ is still negative.

	MNL		NL		$MCOV_U$		$MCOV_N$	
Null LL	-3517.56		-3517.56		-3517.56		-3517.56	
Final LL	-2329.07		-2320.23		-2191.42		-2189.8	
Par	14		15		16		16	
Adj.ρ^2	0.3339		0.3361		0.3725		0.3729	
	est.	asy. t-rat.	est.	asy. t-rat.	est.	asy. t-rat.	est.	asy. t-rat.
$\beta_{\text{TT,car}}$	-0.0232	-9.94	-0.0170	-6.97	-0.0146	-7.35	-0.0140	-7.31
$\beta_{\text{TT,PT}}$	-0.0200	-8.03	-0.0173	-7.07	-0.0160	-6.75	-0.0158	-6.69
$\beta_{\text{TC,car}}$	-0.0255	-6.71	-0.0222	-6.28	-0.0206	-5.80	-0.0201	-5.69
$\beta_{\text{TC,PT}}$	-0.0507	-4.73	-0.0382	-3.56	-0.0320	-3.11	-0.0307	-2.98
β_{freq}	0.1318	1.21	0.1185	1.11	0.1048	0.99	0.1041	0.98
δ_{early}	-1.5594	-23.64	-0.8602	-6.62	-0.7013	-8.37	-0.6607	-8.02
$\delta_{\text{early,regularly work from home}}$	-0.2528	-1.60	-0.1483	-1.79	-0.1095	-1.29	-0.0726	-1.04
$\delta_{\text{early,part-time work}}$	-0.5992	-3.91	-0.3016	-3.30	-0.2190	-1.85	-0.2210	-1.97
δ_{late}	-2.3483	-26.61	-1.2628	-6.50	-1.0615	-8.01	-0.9991	-7.73
$\delta_{\text{late,regularly work from home}}$	0.3568	2.01	0.1640	1.73	0.1391	1.19	0.1604	1.44
$\delta_{\text{late,part-time work}}$	-0.4296	-2.13	-0.2134	-1.98	-0.1355	-0.77	-0.1394	-0.87
δ_{PT}	-2.3653	-6.97	-2.2839	-6.95	-2.1766	-6.72	-2.1626	-6.70
$\delta_{\text{PT,regularly work from home}}$	-0.3442	-1.20	-0.3087	-1.10	-0.3345	-1.18	-0.3253	-1.15
$\delta_{\text{PT,part-time work}}$	0.2753	1.17	0.2861	1.25	0.2654	1.16	0.2595	1.14
$\lambda^{(i)}$	-	-	0.5030	6.12	-	-	-	-
x_a	-	-	-	-	-2.3240	-11.38	-	-
x_b	-	-	-	-	4.2919	7.95	-	-
x_μ	-	-	-	-	-	-	-0.4347	-1.65
x_σ	-	-	-	-	-	-	1.3558	7.45
VTTS_{car}	54.64		46.03		42.43		41.96	
VTTS_{PT}	23.64		27.25		30.13		30.94	

$^{(i)}$ asymptotic t-ratio calculated with respect to 1 instead of 0

Table 1: Estimation results on Dutch TOD data

1.9679, while MCOV_N has x distributed normally with a mean of -0.4347 and a standard deviation of 1.3558, where the mean x_μ is only significantly different from zero at the 90% level. After transformation using the logistic distribution, this gives a range for λ between 0.0892 and 0.8774 in MCOV_U and between 0 and 1 in in MCOV_N.

In terms of model fit, and in comparison to the base MNL model, we observe an improvement in log-likelihood (LL) by 8.84 units for the NL model, which at the cost of one additional parameter, is significant, but far from dramatic. On the other hand, for the two MCOV models, at the cost of two additional parameters when compared to the MNL model, we observe very significant improvements by 137.65 units for the MCOV_U model and 139.27 units for the MCOV_N model. The difference in these improvements shows the advantage of allowing for covariance heterogeneity. Here, it should however be acknowledged that part of the gains in model fit could also be due to capturing serial correlations across replications for the same respondent, through using a panel approach in estimation. Although not generally discussed, it should be said that the same issue arises when comparing results from a MMNL model estimated with a panel formulation to those from a MNL model[13]. In the present context, it should however be said that the models clearly indicate the presence of covariance heterogeneity, independently of the issue of how much of the gains in model fit the representation of this heterogeneity may be responsible for.

As a final comparison between the models, we look at the implied valuation of travel time savings (VTTS) for the two modes of transport. Here, consistent with the findings of Hess *et al.* (2007c), all models show higher VTTS for car travel than for rail travel. However, there are important differences across the different model structures. Indeed, the difference between the VTTS measures for the two modes is especially marked in the MNL model, where the ratio is of the order of 2.3. This decreases to 1.7 in the NL model, but reduces further to around 1.4 in the two MCOV models. As such, the differences between the models are not solely confined to model fit, where it can be seen that, especially when comparing NL to MNL, the dif-

[13] It is a common observation that MMNL models estimated using a panel formulation produce much better model fit than models estimated using a cross-sectional formulation (see for example **?**. A large part of these additional gains can be explained on the basis that an assumption of constant tastes across choices for the same respondent makes more sense than an assumption of equal variation across respondents and choice situations. However, the panel formulation potentially also captures factors unrelated to taste heterogeneity, such as correlation across choices for the same respondent. The disentangling of the various effects accounted for by the panel formulation (and explaining the gains in model fit) is an important area for further research.

ferences in model fit are relatively small when compared to the differences in the VTTS. Finally, there are also important differences in the implied substitution patterns between alternatives; these are looked at in detail in the following section.

3.4 Forecasting

To give an account of the differences across the four models in terms of the implied substitution patterns between alternatives, two forecasting exercises were conducted, looking at the change in sample shares following increases in travel time and travel cost for the base alternative by 20%. This allows us to gauge respondents' behaviour in terms of changing departure time (i.e., shifting to earlier or later) or mode following changes to the base alternative[14].

An important question arises at this stage. No major complications arise in the forecasting using the two closed form models, MNL and NL, where the applicable formulae are used in conjunction with the parameter estimates from Table 1 to produce probabilities for each of the four alternatives. However, in the two MCOV models, issues arise due to the use of a panel formulation in estimation. With this approach, during model estimation, the integration/simulation over random distributions is carried out over respondents rather than individual choice situations. As such, it is the probability of a sequence of choices that is simulated as opposed to the probability of individual choices. The reasoning behind this method is the assumption that tastes vary across respondents but stay constant across observations for the same individual.

However, in forecasting, we require probabilities for individual choice situations. This means that the simulation over whole sequences is not possible[15]. Similarly however, it is not appropriate to use a cross-sectional approach in forecasting, i.e. using simulation over individual choices. Indeed, this would equate to treating two observations from the same respondent in the same manner as two observations from two separate respondents, an approach that would not be consistent with the method used during estimation. To get around this issue, we rely on individual-specific draws obtained

[14] In actual policy analysis, some calibration of the data would be required when forecasting on the basis of SP data. Here, with the aim being purely methodological, this can be avoided.

[15] As an example, we might be interested in predicting the change in a specific alternative's probability in a certain setting following changes in one of the attributes of this alternative.

after estimation through conditioning on observed choices (for discussions on this approach, see for example Train 2003, Sillano and Ortúzar 2004). With this approach, we choose for each respondent the value from the distribution of λ that is most likely to have led to the observed sequence of choices. With $L(Y_n \mid \lambda)$ giving the probability of the observed sequence of choices for respondent n at a specific value for λ, the probability of observing the specific value of λ given the choices of respondent n is given by:

$$K(\lambda \mid Y_n) = \frac{L(Y_n \mid \lambda) f(\lambda)}{\int_\lambda L(Y_n \mid \lambda) f(\lambda)\,\mathrm{d}\lambda}, \tag{17}$$

where λ is distributed according to $f(\lambda)$.

Here, we replace the continuous formulation by a discrete approximation, using summation over a very high number of draws. As such, a mean for the conditional distribution for respondent n is obtained as:

$$\widehat{\lambda_n} = \frac{\sum_{r=1}^{R} \left[L(Y_n \mid \lambda_r)\lambda_r\right]}{\sum_{r=1}^{R} L(Y_n \mid \lambda_r)}, \tag{18}$$

where λ_r with $r = 1, \ldots, R$ are independent draws with equal weight from the distribution of λ.

Using the approach from equation (18), draws for λ were produced for each of the 170 respondents for the two different models $\mathrm{MCOV_U}$ and $\mathrm{MCOV_N}$. In this very simplistic forecasting exercise, we only make use of the conditional means, where the use of the distribution around the conditional means would additionally provide us with confidence intervals.

The results of the two forecasting exercises are shown in Table 3 for an increase in travel time for the base alternative by 20% and Table 4 for an increase in travel cost for the base alternative by 20%. For the two MCOV models, the forecasting exercise was carried out using the simulation approach as well as using the means of the conditional distributions, to illustrate the differences between the two approaches. In this context, it is of interest to look briefly at the substitution patterns for one specific observation. The results of this process are shown in Table 2, where the specific individual chosen for this example had all four alternatives available. The means of the conditional distribution of λ for this individual are 0.5934 in $\mathrm{MCOV_U}$ and 0.5353 in $\mathrm{MCOV_N}$, so both relatively close to the NL value of 0.5. Here, it can be seen that the MNL model predicts a decrease in the choice probability for the base alternative by 7.23%, where, due to the independence from irrelevant alternatives (*IIA*) assumption, this is distributed proportionally across the three remaining alternatives, leading to

		MNL	NL	using simulation MCOV$_U$	using simulation MCOV$_N$	using conditional means MCOV$_U$	using conditional means MCOV$_N$
Base	base alt.	70.16%	70.33%	66.83%	62.50%	59.78%	60.64%
	early dep.	18.61%	17.84%	18.80%	21.22%	23.45%	22.94%
	late dep.	8.45%	8.01%	10.21%	12.24%	12.78%	12.19%
	PT	2.78%	3.82%	4.16%	4.04%	4.00%	4.23%
Predicted	base alt.	65.09%	63.16%	61.06%	56.99%	53.89%	54.42%
	early dep.	21.77%	22.47%	22.58%	24.71%	27.03%	26.77%
	late dep.	9.89%	10.09%	11.77%	13.89%	14.73%	14.23%
	PT	3.25%	4.28%	4.58%	4.41%	4.34%	4.58%
Change	base alt.	-7.23%	-10.19%	-8.63%	-8.83%	-9.85%	-10.25%
	early dep.	+17.00%	+25.95%	+20.10%	+16.46%	+15.30%	+16.67%
	late dep.	+17.00%	+25.95%	+15.30%	+13.54%	+15.30%	+16.67%
	PT	+17.00%	+12.04%	+10.23%	+9.14%	+8.65%	+8.42%

Table 2: Results for first forecasting exercise for a single choice situation

increases by 17%. In the NL model, the decrease in probability for the base alternative is slightly larger, but, more importantly, the *IIA* assumption now only holds inside the car nest, but not across nests, there is a bigger shift towards the remaining two car alternatives than towards the rail alternative. The same principle applies when using the means of the conditional distributions with the two MCOV models, where the percentage increases are slightly less pronounced, due to the lower initial probability for the base alternative. The difference in substitution towards the car alternatives and the PT alternative remains very similar (roughly two to one). The most interesting observation from Table 2 can however be made when looking at the predicted changes in the MCOV models when using simulation rather than the means of the conditional distributions. Here, the *IIA* assumption within nests is violated, where there is now a bigger shift towards the early departure than towards the late departure. This is not consistent with the underlying structure of the MCOV model, reinforcing the idea that it is important to use the conditional draws rather than simulation in application and forecasting.

We next move on to the sample level forecasts in Table 3 and Table 4. Due to the averaging across respondents, the *IIA* assumption clearly does not hold at the level of sample shares in any of the models. However, some interesting differences still arise. Across both forecasting scenarios, the MNL model tends to underestimate the changes in sample shares following the increases in travel time or travel cost for the base alternative. The results for the NL model are roughly similar to those for the two MCOV models,

		MNL	NL	MCOV_U	MCOV_N
Base	base alt.	69.34%	69.34%	69.86%	69.95%
	early dep.	16.91%	16.91%	16.19%	16.11%
	late dep.	8.75%	8.75%	8.93%	8.92%
	PT	5.00%	5.00%	5.02%	5.02%
Predicted	base alt.	60.09%	56.48%	57.31%	57.00%
	early dep.	22.11%	24.72%	24.45%	24.70%
	late dep.	11.49%	12.91%	12.44%	12.53%
	PT	6.31%	5.89%	5.81%	5.77%
Change	base alt.	-13.33%	-18.54%	-17.96%	-18.51%
	early dep.	30.72%	46.17%	51.01%	53.36%
	late dep.	31.27%	47.51%	39.22%	40.39%
	PT	26.27%	17.81%	15.62%	14.93%

Table 3: Results for first forecasting exercise

		MNL	NL	MCOV_U	MCOV_N
Base	base alt.	69.34%	69.34%	69.86%	69.95%
	early dep.	16.91%	16.91%	16.19%	16.11%
	late dep.	8.75%	8.75%	8.93%	8.92%
	PT	5.00%	5.00%	5.02%	5.02%
Predicted	base alt.	65.61%	63.15%	63.57%	63.42%
	early dep.	18.99%	20.63%	20.27%	20.39%
	late dep.	9.88%	10.79%	10.73%	10.78%
	PT	5.52%	5.44%	5.43%	5.41%
Change	base alt.	-5.38%	-8.93%	-9.00%	-9.34%
	early dep.	12.30%	22.02%	25.22%	26.63%
	late dep.	12.91%	23.19%	20.14%	20.76%
	PT	10.40%	8.74%	8.06%	7.78%

Table 4: Results for second forecasting exercise

where, in the MCOV models, there is however a more pronounced shift towards the early car alternative. Here, the small level of differences between the NL and MCOV models can be partly put down to the averaging across respondents, where more significant differences arise at the individual level, as highlighted in Table 2. Finally, from a more policy oriented perspective, the results across models show a bigger effect of changes in travel time than travel cost, characterised by a bigger shift away from the base alternative (by a factor of around two to one), where the actual substitution patterns clearly remain unaffected.

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 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. Finally, the model can be specified as a continuous mixture or as a discrete mixture.

The application presented in Section 3 has illustrated the advantages of allowing for random covariance heterogeneity in an example making use of stated preference data for the choice of departure time and travel mode in the Netherlands. The results have shown that, while moving from a simple MNL model to a NL structure leads to minor gains in model performance, far more significant gains can be obtained through additionally allowing for a variation across respondents in the level of correlation between alternatives that are nested together. Here, the application shows great variation across respondents, ranging from low levels of correlation in the error terms (to the order of 30%), to almost perfectly correlated error terms for nested alternatives for some individuals. This leads to very different patterns of substitution between the alternatives across respondents, a phenomenon that cannot be represented with models assuming a homogeneous correlation structure. The insights gained by allowing for such covariance heterogeneity can be of great value in forecasting applications, as illustrated in Section 3.4, most notably at the observation specific level (cf. Table 2). However, another important difference arises between the models in terms of significant differences in the values of willingness to pay indicators, such as the VTTS. This suggests that by making an assumption of covariance homogeneity, modellers potentially are at risk of obtaining biased estimates for marginal utility coefficients, leading to unreliable policy indicators.

Several avenues for further research can be identified, including the development of more sophisticated mixed covariance structures, and the testing of mixed covariance structures on other datasets. 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. Finally, it also remains of interest to estimate models with different distributional assumptions for the structural parameters, partly with a view to avoiding a priori range assumptions as is the case when using a transform mapping onto the $0 - 1$ interval.

In closing, it should be said 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, and with the help of posterior analyses, 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, 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.

Acknowledgements

Part of the work described in this paper was carried out during a guest stay by the first author in the Institute of Transport and Logistics Studies at the University of Sydney. The authors would like to thank Frank Hofman for making the Dutch National model data available for the present research. The authors would also like to thank John Rose for extensive feedback and discussions on an earlier version of this paper.

References

Ben-Akiva, M. and D. Bolduc (1996) *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*, Working paper, Department of Civil and Environmental Engineering, MIT, Cambridge, MA.

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

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

de Jong, G., A. Daly, M. Pieters, C. Vellay, M. A. Bradley and F. Hofman (2003) A model for time of day and mode choice using error components logit, *Transportation Research Part E: Logistics and Transportation Review*, **39** (3) 245–268.

Doornik, J. A. (2001) *Ox: An Object-Oriented Matrix Language*, Timberlake Consultants Press, London.

Greene, W. H., D. A. Hensher and J. M. Rose (2006) Accounting for heterogeneity in the variance of unobserved effects in mixed logit models, *Transportation Research Part B: Methodological*, **40** (1) 75–92.

Hess, S. (2005) Advanced discrete choice models with applications to transport demand, Ph.D. Thesis, Centre for Transport Studies, Imperial College London.

Hess, S., M. Bierlaire and J. W. Polak (2005a) Capturing taste heterogeneity and correlation structure with Mixed GEV models, in R. Scarpa and A. Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, chap. 4, 55–76, Springer Publisher, Dordrecht, The Netherlands.

Hess, S., M. Bierlaire and J. W. Polak (2005b) 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., M. Bierlaire and J. W. Polak (2007a) A systematic comparison of continuous and discrete mixture models, *European Transport*, **36**.

Hess, S., A. Daly, C. Rohr and G. Hyman (2007b) On the development of time period and mode choice models for use in large scale modelling forecasting systems, *Transportation Research Part A: Policy and Practice*, **41** (9) 802–826.

Hess, S., J. W. Polak and M. Bierlaire (2005c) Functional approximations to alternative-specific constants in time-period choice-modelling, in H. Mahmassani (ed.), *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, chap. 28, 545–564, Elsevier, Amsterdam.

Hess, S., J. W. Polak, A. Daly and G. Hyman (2007c) Flexible substitution patterns in models of mode and time of day choice: New evidence from the uk and the netherlands, *Transportation*, **34** (2) 213–238.

Hess, S. and J. M. Rose (2007) *Intra-respondent taste heterogeneity in instantaneous panel surveys*, paper to be presented at the 11th triennial WCTR conference, July 2007, Berkeley, CA.

Koppelman, F. S. and V. Sethi (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.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour, in P. Zarembka (ed.), *Frontiers in Econometrics*, 105–142, Academic Press, New York.

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

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

Ortúzar, J. de D. (2001) On the development of the nested logit model, *Transportation Research Part B: Methodological*, **35** (2) 213–216.

Sillano, M. and J. de D. Ortúzar (2004) Willingness-to-pay estimation from mixed logit models: some new evidence, *Environment & Planning A*, **37** (3) 525–550.

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

Walker, J. (2001) Extended discrete choice models: Integrated framework, flexible error structures, and latent variables, Ph.D. Thesis, MIT, Cambridge, MA.

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[16], the utilities of two alternatives that have some correlation in the unobserved part of utility would be written as:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n} + \sigma_1 \xi_1 \tag{19}$$

and

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sigma_1 \xi_1, \tag{20}$$

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. The additional error term ξ_1 is distributed $N(0,1)$. 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:

$$\rho^2 = \frac{\sigma_1^2}{\sigma_1^2 + \frac{\pi^2}{6}}. \tag{21}$$

[16]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.

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:

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sum_{k \in \Psi_j} \sigma_k \xi_k \tag{22}$$

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:

$$P_n(i \mid \boldsymbol{\sigma}) = \int_{\xi_1} \ldots \int_{\xi_K} \left[\frac{exp\left(V_{i,n} + \sum_{k \in \Psi_i} \sigma_k \cdot \xi_k\right)}{\sum_{j \in C_n} exp\left(V_{j,n} + \sum_{l \in \Psi_j} \sigma_l \cdot \xi_l\right)} \cdot \prod_{k=1}^{K} \phi(\xi_k) \right] \mathrm{d}\xi_K \ldots \mathrm{d}\xi_1, \tag{23}$$

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[17]; equation (21) guarantees that the resulting correlation falls between 0 and 1.

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

[17] 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 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 $f()$ 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.

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:

$$P_n(i) = \int_{\sigma_1} \dots \int_{\sigma_K} \left[P_n(i \mid \boldsymbol{\sigma}) \cdot \prod_{k=1}^{K} g(\sigma_k \mid \boldsymbol{\theta_k}) \right] \mathrm{d}\sigma_K \dots \mathrm{d}\sigma_1, \tag{24}$$

where $P_n(i \mid \boldsymbol{\sigma})$ is the choice probability for alternative i, conditional on the vector of standard deviations $\boldsymbol{\sigma}$, as in equation (23), and where $g(\sigma_1 \mid \boldsymbol{\theta_1})$ 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[18]. An alternative to the use of bounded distributions comes in the use of a transform mapping monotonically from the real domain to the space of positive numbers. The adaptation of equation (24) 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 $\sigma = f(\boldsymbol{\theta}, \boldsymbol{z_n})$. 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 decision maker n is now rewritten as:

$$P_n(i) = \int_{\boldsymbol{\theta_1}} \dots \int_{\boldsymbol{\theta_K}} \left[P_n(i \mid \sigma_k = f(\boldsymbol{\theta_k}, \boldsymbol{z_n}) \, \forall k) \cdot \prod_{k=1}^{K} g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k}) \right] \mathrm{d}\boldsymbol{\theta_K} \dots \mathrm{d}\boldsymbol{\theta_1}, \tag{25}$$

where $\boldsymbol{\theta_k}$ is distributed according to $g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k})$, 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

[18] 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.

those parameters associated with z_n are zero[19], and the purely deterministic formulation in Section A.2 in the case where $g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k})$ 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.

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