Department for Transport

SUPPLEMENTARY GUIDANCE

Mixed Logit Models

January 2014

Department for Transport

Transport Analysis Guidance (TAG)

https://www.gov.uk/transport-analysis-guidance-tag

This TAG Unit is part of the section **SUPPLEMENTARY GUIDANCE**

Technical queries and comments should be referred to:

Transport Appraisal and Strategic Modelling (TASM) Division
Department for Transport
Zone 2/25 Great Minster House
33 Horseferry Road
London
SW1P 4DR
tasm@dft.gov.uk

Contents

1 Introduction

1.1 Scope of the guidance

1.1.1 TAG Unit M2 – Variable Demand Modelling offers guidance on mode choice model development, with particular reference to the multinomial logit (MNL) and nested logit (NL) models. Although MNL and NL are perfectly adequate for many applications, the properties of these models may in some contexts become overly restrictive. Three such contexts are of particular relevance to current practice: taste variation, repeated choices and patterns of substitution. Mixed logit (MXL) brings an opportunity to model such contexts, and this unit offers guidance on its usage.

1.1.2 This TAG Unit provides guidance on the procedures and documentation that should be required of those charged with applying MXL. This unit should be considered a supplement to the guidance provided in Supplementary Guidance – Bespoke Mode Choice Models.

1.1.3 More specifically, this TAG Unit:

- Introduces the basic motivation behind MXL, and assesses its present status in travel demand analysis;
- Discusses the strategic factors that should govern its usage;
- Identifies important analytical interests that may be pursued by means of MXL;
- Provides technical guidance on the how these interests should be developed;
- Discusses the training and conduct that can reasonably be expected of the MXL analyst; and
- Lays down a minimal set of activities that should be included in the development of a MXL model, including an associated audit trail.

2 Consideration of the Use of MXL

2.1 The motivation for MXL

2.1.1 In common with MNL and NL, MXL is derived from the paradigm of utility maximisation, and many aspects of model development, data collection and model application coincide.

2.1.2 MXL, as usually applied in practice, might be referred to more precisely as mixed **multinomial** logit (see Appendix A). Moreover, MXL may be thought of as a generalised model within which MNL is an **exact** restricted case, and other models with more complex substitution patterns (e.g. NL) are **approximate** restricted cases. Although this approximation can be as close as the analyst desires, it can never be an exact one; this is because MXL, unlike MNL and NL, does not yield a 'closed form' and must therefore be estimated by simulation.

2.1.3 As well as permitting considerable flexibility in the representation of substitution patterns, this generalised framework offered by MXL provides a forum for the developing the interests of taste variation and repeated choices.

2.2 Strategic considerations

2.2.1 Whilst its versatility may carry considerable attraction, the decision to actually proceed with MXL in practice may not be a trivial one; not least because it incurs a considerable extension in complexity relative to MNL or NL. Indeed, the pitfalls for the unwary are - relative to MNL and NL - perhaps more potent, and MXL modelling should be considered the domain of highly-trained analysts. In particular, issues of model identification are highly complex and it should not be presumed that a convergent MXL is necessarily a valid MXL.

2.2.2 As well as increased demands on the analyst, MXL imposes greater demands on computing power. Moreover, the implementation of MXL is typically more time-intensive than MNL and NL, and carries less certainty in the outcome of the modelling process; these considerations should be balanced against its conceptual appeal.

2.2.3 It should be acknowledged that understanding of MXL - though advancing rapidly - remains incomplete. When one contrasts the plethora of MNL and NL applications with the relatively small number of MXL applications, it becomes clear that MXL practice is presently in its early stages. Given this context, it would seem prudent, at this point in its development, to employ MXL as an addition to, rather than as a replacement for, MNL and NL. Indeed, such advice should not be seen as particularly onerous, since it is good practice to precede MXL modelling with the estimation of relevant MNL and NL specifications. The latter may not only be instructive to the specification and estimation of MXL, but function as a benchmark against which the results of MXL estimation may be usefully judged.

2.2.4 Applications of MXL are more complex than conventional choice model methodologies. Therefore, it is recommended that analysts should commence with the estimation of a standard MNL as a benchmark model.

2.3 Alternative interpretations of MXL

2.3.1 Several alternative interpretations of MXL have been proposed in the literature. Two such interpretations - **Random Parameters Logit (RPL)** and **Error Components Logit (ECL)** - have acquired particular prominence. Whilst RPL and ECL are in mathematical terms entirely equivalent, their respective interpretations are motivated by distinct analytical interests. More specifically, RPL appeals readily to the analysis of taste variation and repeated observations, whilst the ECL interpretation is more amenable to the analysis of complex substitution patterns.

2.3.2 Appendix A provides a formal mathematical definition of the general MXL model. Appendix B and Appendix C then formalise the derivation of the RPL and ECL interpretations, respectively, from the general model.

3 Applications of RPL

3.1 Taste variation

3.1.1 MNL and NL offer only limited facility for analysing the prevalence of taste variation across a population of decision-makers; this involves the estimation of separate taste parameters for each segment of the population (e.g. income group, journey purpose, etc).

3.1.2 The RPL interpretation of MXL adopts a rather different approach, allowing the data more freedom to directly reveal the form of any inherent taste variation, without recourse to any particular segmentation. In its most basic form, RPL yields estimates of the first and second moments (e.g. mean and standard deviation) of the distribution of tastes across the population of interest. This procedure is explained in more detail in Appendix B. More complex specifications can, feasibly, estimate a taste parameter for each and every individual within a population; this is discussed further in Appendix F. Distributional forms are discussed in Appendix D. The notion of a distributed value of time (i.e. the ratio of time and cost parameters) raises the question of whether the time or cost (or both) parameters should be distributed, and this is considered further in Appendix G.

3.1.3 The possibilities offered by RPL in the modelling of taste variation may be particularly relevant to contexts where the impact of segmentation of valuation may be pertinent. It is difficult, however, to offer general advice on the prospective benefits of RPL vis-à-vis a segmented MNL or NL, since such benefits will be case specific and dependent on the particular segmentation adopted. In some cases, RPL may yield more detailed information on the distribution of the value of time than MNL or NL, but whether such detail will materially affect any subsequent economic evaluation is an entirely empirical question. Furthermore, and with reference to earlier discussion, the benefits in terms of additional analytical detail must be weighed against the associated research cost. In other cases, a segmented MNL or NL can provide valuable insight into how average tastes vary across market segments. This information is likely to be of use to those involved in policy and investment analysis, product design and marketing. Given the relative advantages of each method, a practical approach to model building is to use a combination of RPL and market segmentation to capture taste variation.

3.2 Repeated choices

3.2.1 Stated Preference (SP) data is often comprised of several observations from each member of a population, i.e. it is characterised by the phenomena of 'repeated choices'. This creates a propensity for the observations of any particular individual to be correlated, and this may in turn induce bias in the *t*-ratios of parameters estimated by MNL or NL.

3.2.2 Whilst re-sampling techniques such as jack-knifing or bootstrapping offer means for ex post correction of such bias, an alternative approach is to account for such correlation *a priori* in the specification of the choice model. RPL provides a facility for implementing the latter, accommodating any correlation between repeated choices in the specification of the covariance matrix of the distributed parameters (see Appendix B for more detail). Again, it is difficult to issue guidance on the relative merits of jack-knife and bootstrap techniques *vis-à-vis* RPL. Experience in comparing the two is relatively limited; even then, the approaches are conceptually distinct, and not necessarily substitutable. An attraction of jack-knife and bootstrap techniques is that they may be useful in diagnosing a range of specification errors, of which repeated choices may be one; in some contexts, this may be a compelling reason for their usage. In other cases, particularly where several functions of MXL are being combined (e.g. repeated choices and patterns of substitution), it may be more convenient to apply RPL instead. Moreover, the method finally adopted by the analyst may once again imply a trade-off between likely analytical benefits and research costs.

4 Applications of ECL

4.1 Complex patterns of substitution

4.1.1 MNL is characterised by the independence from irrelevant alternatives (IIA) property, meaning that for any two alternatives, the ratio of their choice probabilities is unaffected by the presence or absence of any other alternatives in the choice set (i.e. the 'red bus-blue bus' problem). NL permits a partial relaxation of this property, whereby similar alternatives (or in other words, close substitutes) are grouped together in mutually exclusive 'nests'. IIA then applies within nests but not across nests.

4.1.2 The relaxation offered by NL is 'partial' in the sense that each alternative is accorded membership of a single nest only, and this implies restriction on the permitted patterns of substitution between alternatives. To illustrate this point, consider the schematic example of a mode choice between three alternatives: red bus, blue bus and red train (Figure 1).

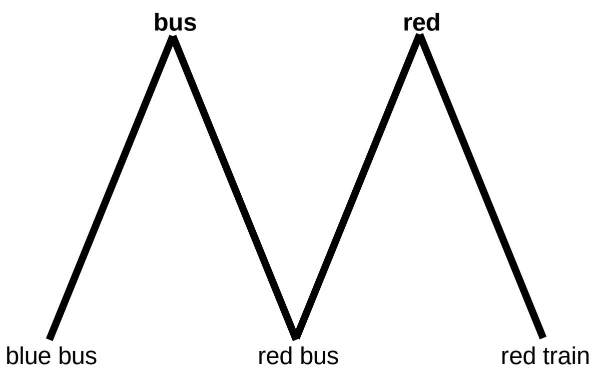

Figure 1 Example of cross-nesting

4.1.3 Two dimensions of similarity are apparent; both i) and ii) are buses, and both i) and iii) are red, thereby giving two notional nests. Whereas blue bus and red train each belong to a single nest, as required by NL, red bus belongs to both nests (i.e. is 'cross-nested'), and this is beyond the scope of NL.

4.1.4 The ECL interpretation of MXL would, by contrast, offer a facility for proceeding with such a representation. Indeed ECL can, in principle, accommodate any pattern of substitution; the more relevant constraint on the complexity of substitution patterns is our ability to compute such models. The procedures required to implement the above structures within ECL are described in greater detail; in Appendix C.

4.1.5 The notion of cross-nesting described above may be applied to structures of considerably greater complexity, as required. If one is interested in mode choice for airport access, for example, one might specify a nest for airport choice, a nest for carrier choice, and a nest for mode choice, with cross-nesting between the three. If one is interested in mode and departure time choice, one might specify mode choice and departure time nests, with crossing between the two; this could feasibly be extended to include cross-nesting within the departure time nest, thereby accounting for substitution between neighbouring departure times (i.e. an approximation to the Ordered Generalised Extreme Value (OGEV) model).

4.1.6 The latter is illustrated in Figure 2, for a choice set of three departure times (8:05am, 8:10am and 8:15am). It may be seen that:

- Nest 1 is a single alternative nest for 8:05am
- Nest 2 includes two correlated alternatives: 8:05am and 8:10am
- Nest 3 includes three correlated alternative: 8:05am, 8:10am and 8:15am
- Nest 4 includes two correlated alternatives: 8:10am and 8:15am
- Nest 5 is a single alternative nest for 8:15am

Since some departure times belong to more than one nest, this induces covariance between those nests.

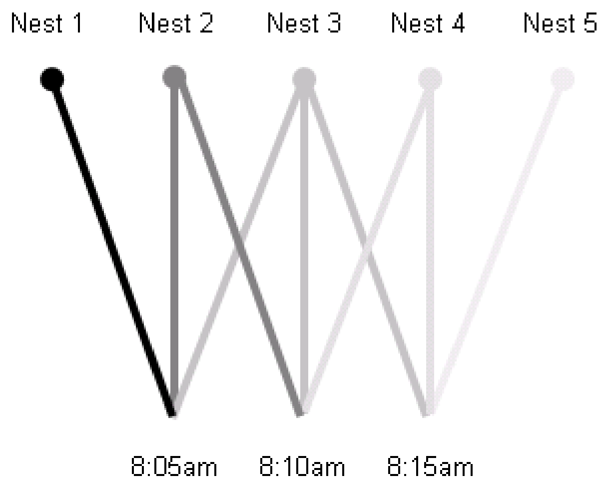

Figure 2 Example of cross-nesting based on ordered alternatives

4.2 Merging data

4.2.1 The ECL interpretation can be further exploited to approximate the 'trick' of merging datasets (e.g. RP and SP data, or two different SP datasets) within the apparatus of NL. Thus if, in Figure 2, the 'bus' and 'red' nests arose from different datasets, the scale of the error variance could be straightforwardly normalised by introducing different log sum parameters (i.e. θ_{bus}, θ_{red}) on the two nests. This procedure is explained in more detail in Appendix C.

5 Applying RPL and ECL together

5.1 Exploiting the equivalence between RPL and ECL to yield a more general model

5.1.1 Whilst any one of the above interests could provide reasonable justification for employing MXL, the real power of the model arises where two or more interests are pursued together. For example, one might wish to specify an ECL interpretation to accommodate cross-nesting; one might then wish to enhance the model to investigate the prevalence of taste variation, via the introduction of an RPL interpretation. Since ECL and RPL are mathematically equivalent, it is perfectly viable to pursue both interests together. In this way, MXL can be seen as a general modelling framework, within which a range of advanced analytical interests may be pursued.

6 Practical considerations of using MXL

6.1 Model design

6.1.1 The analyst should consider the motivation for using MXL. This is likely to derive from one or more of the following interests:

- Taste variation;
- Repeated choices;
- Complex patterns of substitution;
- Merging data.

6.1.2 Having identified the principal analytical interests, this will direct the analyst towards the RPL or ECL (or both) interpretations of MXL. To recap, taste variation and repeated choices are readily amenable to RPL (Appendix B), and the other two interests to ECL (Appendix C).

6.1.3 The project manager should, in consultation with the analyst, consider the prospective analytical benefits of embarking on MXL modelling, and rationalise this against the projected research costs.

6.1.4 MXL modelling should, at this formative stage in its adoption by practitioners, be regarded as a supplement to the more established methods of MNL and NL modelling. This does not necessarily mean that every single procedure must be performed twice; rather that any MXL should, for purposes of benchmarking, be presented against a reasonable MNL or NL comparator.

6.1.5 Analysis should always commence with the estimation of a standard MNL as a benchmark model; indeed some software automatically estimates MNL as a default starting point for estimating MXL.

6.1.6 For any such interest, there may be value in estimating further benchmark models: e.g., estimating a range of NL structures might offer insight into possible patterns of substitution; similarly, estimating MNL with various patterns of segmentation might usefully inform an interest in MXL modelling of taste variation.

6.1.7 The third and fourth interests listed above - complex patterns of substitution and merging data - can be implemented in a reasonably prescriptive manner, exploiting the ECL interpretation.

6.1.8 The first and second interests - taste variation and repeated choices - require a more exploratory approach to the analysis, involving the RPL interpretation. With reference to Appendix D, the analyst should experiment with different functional forms, different combinations of fixed and distributed parameters, different forms for the distributional parameters (e.g. Normal, log Normal, triangular, uniform, etc.), and different truncations of those distributions (where appropriate). The choice of analytical distribution may be usefully informed by an examination of the relevant empirical distribution. Before embarking on RPL modelling, it is sensible practice to first estimate an RPL with all parameters fixed, and then validate the resultant model against MNL.

6.1.9 Whatever the motivation, MXL is estimated by means of a maximum simulated likelihood procedure (Appendix E). Experience reveals that this may not be a trivial exercise; it may require considerable effort and experimentation before successful convergence is achieved and, in particular, the robustness of the estimated results is established.

6.1.10 A frequent problem is that MXL converges to a local optimum, and it is thus essential that the analyst experiments with a range of starting values for the parameters (informed perhaps by those estimated in the benchmark models).

6.1.11 The estimation may also demonstrate considerable sensitivity to the type and number of random draws adopted in the estimation procedure. It is difficult to offer precise advice on such matters, although experience would point towards a recommendation of at least 200 Halton sequences. That said, it remains the responsibility of the analyst to demonstrate the stability of his or her model against the type and number of draws adopted.

6.1.12 The interpretation and diagnostics of the estimated MXL should largely follow the usual practices applied to MNL and NL; TAG Unit M2 offers appropriate guidance in this regard. It is however important that this exercise is conducted comparatively - with MXL formulations considered alongside benchmark models - in order to illuminate any substantive distinctions between MXL and more basic models.

6.1.13 RPL yields, in its most basic form, measures of the first and second moments of the distribution of any random parameter. This approach may be usefully extended to consider separate taste parameters for different segments of the population. Indeed, if repeated choices of an individual are exploited, then one can potentially estimate a separate taste parameters for each and every individual within the sample (Appendix F).

6.1.14 Measurements of willingness to pay should be derived from MXL in accordance with the procedure outlined in Appendix G. Since experience reveals that such measurements may differ significantly from those of MNL or NL, it is again important that these are reported comparatively. The decision of which results to adopt and which to discard must be clearly justified; this may be influenced by other properties (e.g. the robustness) of the estimated MXL.

6.1.15 A message that hopefully emerges is that many of the above activities may be contingent on one another; some specifications may be more easily estimated than others; some specifications may be robust to changes in the estimation procedure whilst others may not.

7 References

Ben-Akiva, M. and Lerman, S. R. (1985) Discrete Choice Analysis, Theory and Application to Travel Demand. The MIT Press, Cambridge, Massachusetts.

Hensher, D. and Greene, W. (2003) 'The Mixed Logit model: the state of practice' in Transportation, 30, pp133-176.

McFadden, D. (1978) Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J. (eds) Spatial Interaction Theory and Residential Location. North-Holland, Amsterdam.

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

8 Document Provenance

8.1.1 This TAG Unit is for supplementary information alongside the main TAG guidance. It was formerly TAG Unit 3.11.5

Appendix A Fundamentals of MXL

A.1 Definition of MXL

A.1.1 In order to proceed with the interests outlined in section 1, it is useful to define MXL more formally, as follows. Before proceeding, it might be remarked that the literature has not, to date, united behind a definitive notation, and different contributors have employed different conventions. The following adheres reasonably closely to Train's (2003) notation.

A.1.2 MXL expresses choice probability as a weighted average or *mixture* of logit probabilities estimated at different values of β with $f(\beta)$ the *mixing* distribution. More formally, choice probability is given by the integral of MNL choice probabilities over a density distribution of parameters, that is:

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta \qquad (1)$$

where P_{ni} is the probability that individual *n* chooses alternative *i*, and $L_{ni}(\beta)$ is the MNL probability at parameter *β*, i.e.

$$L_{ni}(\beta) = e^{V_{ni}(\beta)} \Big/ \sum_{j=1}^{J} e^{V_{nj}(\beta)} \qquad (2)$$

and $f(\beta)$ is a density function; the interest is in estimating the parameters of this density.

A.1.3 In relating this definition to the analytical interests of taste variation, repeated choices, patterns of substitution, and data merger, we shall exploit two alternative but entirely equivalent interpretations of MXL; random parameters logit (RPL) and error components logit (ECL).

Appendix B Applications of RPL

B.1 Random Parameters Logit (RPL)

B.1.1 According to RPL, the utility derived by individual *n* from alternative *j* is given by:

$$U_{nj} = V_{nj}(\beta) + \varepsilon_{nj} \qquad (3)$$

where $V_{nj}(\beta) = \beta_n' x_{nj}$

and x_{nj} are observable variables that characterise the individual n and alternative j, β_n is a vector of parameters relating to these variables for individual n, and ε_{ni} is a random error term that is IID extreme value over individuals and alternatives. The β parameters vary over individuals with density $f(\beta)$.

B.1.2 As we shall see subsequently (Appendix C), there is considerable flexibility in specifying the form of $f(\beta)$. If, for the sake of simplicity, we assume that $\beta \sim N(\mu, \sigma)$, then (3) can be re-written:

$$U_{nj} = \mu' x_{nj} + \sigma_n' x_{nj} + \varepsilon_{nj} \qquad (4)$$

B.2 Taste variation

B.2.1 Interpreting (4), μ can be seen to be the mean of the distribution of tastes for x_{nj} across a sample of individuals, and σ_n the standard deviation of tastes for x_{nj} pertaining specifically to individual n. In this way, we have a means of representing taste variation across the sample. The estimation process then yields estimates of both μ and σ. If σ is significantly different from zero then one can infer that the sample exhibits taste variation; otherwise RPL collapses to MNL.

B.3 How many and which parameters should be analysed for taste variation?

B.3.1 Although some parameters may exhibit taste variation, the properties of these parameters - in particular derived valuations - may not necessarily accord with previous or indeed 'acceptable' limits. In such cases, it is left to the analyst to make a judgement as to whether to retain the distributed parameters, to adjust the specification of the RPL in the hope of achieving a more appealing result (for example, fixing the cost parameter whilst allowing the numerator of the valuation to vary, or introducing a truncated distribution), or to dispense with RPL altogether. Whilst it is difficult to offer definitive guidance on such matters, the analyst should at all times be able to justify his or her course of action; such a justification will always be more persuasive if it is underwritten by a comprehensive analysis investigating a range of possible responses to such problems. That notwithstanding, it is always good practice when embarking on an intensive examination of taste variation, to first estimate an RPL with all parameters fixed and then confirm the results against MNL.

B.3.2 Quite aside from the fundamental question of whether a particular parameter demonstrates taste variation - which is an entirely empirical matter - issues of model identification may imply restrictions on how many parameters can feasibly be specified as random (and how many must remain fixed). The analyst may therefore need to prioritise some parameters over others in the representation of taste variation.

B.4 Repeated choices

B.4.1 A further application of RPL is in the modelling of repeated choices by individuals in SP experiments, although it requires a slight extension on the preceding analysis.

B.4.2 Consider a sample of individuals *n*, each responding to replications *t* of the choice experiment. Making the common simplifying assumption that the taste parameters vary over individuals but are constant over replications for each individual (effectively implying that there are no habit or learning effects), the utility derived by individual *n* from alternative *j* in replication *t* is given by:

$$U_{njt} = \beta'_n x_{njt} + \varepsilon_{njt} \qquad (5)$$

where ε_{njt} is IID extreme value over individuals, alternatives and replications. MNL assumes independence between the stochastic error of observations from different replications. Where dependence exists but is not accounted for, parameter estimates are unbiased but the variance estimates are downward biased.

B.4.3 RPL is able to accommodate such dependence, thus:

Let $\beta \sim N(\mu,\sigma)$; (5) can then be re-written:

$$U_{njt} = \mu' x_{njt} + \sigma'_n x_{njt} + \varepsilon_{njt}$$

and $\sigma'_n x_{njt} + \varepsilon_{njt}$ will be correlated over alternatives *and* replications due to the common influence of σ_n.

B.4.4 Finally, we need to adjust the expressions (1) and (2) to accommodate the additional dimension *t*. The probability that the individual makes a particular sequence of choices is, conditional on *β*, given by the product of MNL choice probabilities:

$$\mathbf{L}_{n\mathbf{i}}(\beta)=\prod_{t=1}^{T}\left[e^{V_{nit}(\beta)} \Big/ \sum_{j=1}^{J} e^{V_{njt}(\beta)}\right] \qquad (6)$$

In other words (6) provides a panel data analogy of (2), which relates to cross-sectional data. Choice probability is then, analogously to (1), given by the integral of (6) over all values of *β*:

$$P_{n\mathbf{i}}=\int \mathbf{L}_{n\mathbf{i}}(\beta)f(\beta)d\beta$$

B.4.5 It is important to note that, in general:

$$\int\left(\prod L_{ni}(\beta)\right)f(\beta)d\beta \neq \prod\left(\int L_{ni}(\beta)f(\beta)d(\beta)\right)$$

That is to say, the integral and multiplicative sum should be taken in the correct order.

Appendix C Applications of ECL

C.1 Error components logit (ECL)

C.1.1 The error components interpretation takes account of correlations between the utility of choice alternatives. According to this interpretation, the utility derived by individual *n* from alternative *j* is given by:

$$U_{nj}=V_{nj}(\beta)+\varepsilon_{nj} \qquad (7)$$

where $V_{nj}(\beta)=\alpha' x_{nj}+\gamma'_n z_{nj}$

and x_{nj} and z_{ni} are vectors of observed variables relating to individual *n* and alternative *j*, α is a vector of fixed parameters, γ_n is a vector of random terms pertaining to individual *n* with zero mean, and ε_{nj} is distributed IID extreme value. The γ_n parameters, which are interpreted as *error components*, are combined with ε_{nj} to define the stochastic portion of utility:

$$\eta_{nj}=\gamma'_n z_{nj}+\varepsilon_{nj}$$

C.2 Patterns of substitution

C.2.1 By specifying z_{ni} in particular ways, one can induce particular patterns of correlation between the stochastic utility of different alternatives. In particular:

- If z_{ni} is identically zero, then this is representative of zero correlation in utility over alternatives. In this case, RPL gives rise to the IIA property and ECL offers an approximation to MNL.
- If on the other hand z_{ni} is non-zero, then this induces a non-zero covariance between the stochastic utility of alternatives *i* and *j*:

$$Cov(\eta_{ni},\eta_{nj})=E(\gamma'_n z_{ni}+\varepsilon_{ni})(\gamma'_n z_{nj}+\varepsilon_{nj})=z'_{ni}Wz_{nj}$$

where W is the covariance of γ_n. Indeed if the error components are independent, such that the off-diagonal elements of W are zero, then the stochastic utility of alternatives *i* and *j* will still have non-zero covariance.

C.2.2 This framework permits representation of any desired pattern of correlation between alternatives. Indeed, ECL can approximate any random utility model; this includes any model derived from McFadden's (1978) Generalised Extreme Value (GEV) family (see section C.5).

C.3 An approximation to NL

C.3.1 Consider for example an approximation to NL. Let the vector z_{ni} consist of a dummy variable for each nest *k* set to one if alternative *j* is a member of nest *k*, and to zero otherwise. Formally:

$$\gamma'_n z_{nj} = \sum_{k=1}^{K} \gamma_{nk} d_{jk}$$

where d_{jk} = 1. if *j* is in nest *k* and 0 otherwise. Since γ_{nk} enters the utility of each alternative within nest *k* this induces correlation between those alternatives.

C.3.2 Let the error components be distributed independently Normally, with the error variance of each nest k common across individuals n, i.e.: γ_{nk} IID $N\left(0, {\sigma_k}^2\right)$. Then σ_k^2 represents not only the variance of the error component γ_{nk}, but the covariance between the stochastic utility of any two alternatives in nest *k*. The variance of the error component can be added to the variance of the IID extreme value term to give the variance of the stochastic portion of utility:

$$Var(\eta_{ni}) = E(\gamma_k + \varepsilon_{ni})^2 = \sigma_k^2 + \pi^2/6$$

C.3.3 The correlation between any two alternatives *i* and *j* in nest *k* is given by:

$$Corr(\eta_{ni}\eta_{nj}) = \frac{\sigma_k^2}{\sigma_k^2 + \pi^2/6}$$

and, with reference to Ben-Akiva and Lerman (1985), the log-sum parameter θ in NL, can be approximated by the expression:

$$\theta = \sqrt{1 - Corr(\eta_{ni}, \eta_{nj})}$$

C.3.4 Although the above defines a general representation, one would typically wish to normalise the overall variance of the random error, i.e. estimate a homoscedastic NL. This can be achieved by: with reference to C.3.1, including dummy variables for single-alternative nests as well as for binary and multinomial nests; and with reference to C.3.2, specifying the error components for each and every nest as independent but of common variance, i.e. γ_{nk} IID $N\left(0, \sigma^2\right)$.

C.4 An approximation to CNL

C.4.1 Exploiting the potential of ECL more fully, consider also an application to the example problem in section 5.2; i.e. an approximation to cross-nested logit (CNL). If again we let the vector z_{nj} consist of a dummy variable for each nest *k* (in this case $k = 1, 2$), set to one if alternative *j* is a member of nest *k*, and to zero otherwise, then we would have:

- For the bus nest ($k = 1$):
  - For the red bus alternative ($j = 1$): $d_{11} = 1$
  - For the blue bus alternative ($j = 2$): $d_{21} = 1$
  - For the red train alternative ($j = 3$): $d_{31} = 0$

- For the red nest ($k = 2$):
  - For the red bus alternative ($j = 1$): $d_{12} = 1$
  - For the blue bus alternative ($j = 2$): $d_{22} = 0$
  - For the red train alternative ($j = 3$): $d_{32} = 1$

There would thus be two error components, one for each nest; their specification then proceeds along the lines already outlined for NL.

C.5 Other random utility models

C.5.1 Let us now return to the earlier assertion that ECL can approximate any random utility model. Figure 3 illustrates some of the inherent relations between ECL and the most prominent models in travel demand analysis, distinguishing between members and non-members of the GEV family. Bold arrows are indicative of 'a generalisation of' (e.g. 2-level NL is a generalisation of MNL, or in other words MNL is a restricted case of 2-level NL), whereas dotted arrows are indicative of 'may be approximated by'. It may thus be seen that ECL can approximate all of the models considered by the figure, with the exception of C-logit, which is not a random utility model.

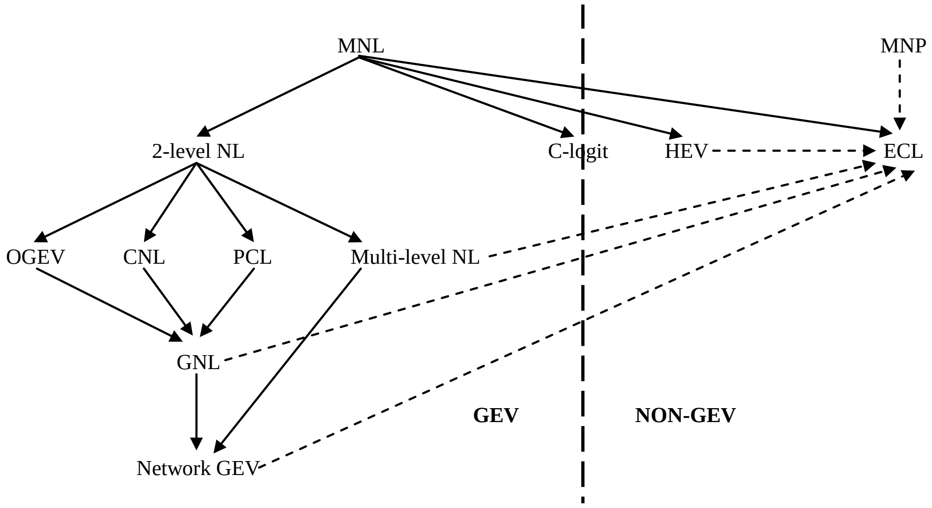

Figure 3 The relation between ECL and other random utility models

C.6 Merging different sources of data

C.6.1 Since ECL can approximate NL, it perhaps comes as little surprise that ECL can be further exploited to approximate the 'trick' of using NL to merge two datasets with different IID extreme value error variances.

C.6.2 In order to implement this, it is necessary to adopt the dataset with the lower error variance as the 'base', and to specify single alternative nests (i.e. d_j = 1) for each of the alternatives *j* within the dataset of greater error variance. A practical obstacle is that it may not be obvious *a priori* which dataset has the lower error variance, in which case it would be necessary to try both permutations and then compare their respective results (it may be that one permutation fails to converge, or yields an inference of zero scale difference). The latter problem does not apply to the NL trick, such that the ECL equivalent may at times appear relatively onerous.

C.6.3 Consider in particular the popular interest in merging RP and SP datasets, where RP would typically have the greater error variance. In this case one would specify a single alternative nest for each of the RP alternatives, with the SP alternatives as the base. An error component is then assigned to each of the RP alternatives, assuming the error components to be distributed independently Normally, with zero mean and standard deviation common across both RP alternatives and individuals. This would yield the following utility functions for the RP and SP alternatives, respectively:

$$U_{nj}^{RP} = \alpha' x_{nj} + \gamma' z_{nj} + \varepsilon_{nj}$$

$$U_{nj}^{SP} = \alpha' x_{nj} + \varepsilon_{nj}$$

C.6.4 Thus the RP alternatives would have stochastic error $\eta_{nj} = \gamma' z_{nj} + \varepsilon_{nj}$, as compared with the stochastic error of the SP alternatives ε_{nj}. In this way, ECL accommodates the greater error variance of the RP data, with γ' the 'scale' parameter. It might be noted that this is not directly comparable to the scale parameter emanating from the trick (which is represented in terms of the θ parameter of NL). In order to carry out such a comparison one must first apply the earlier equations for approximating θ by means of η_{nj}.

C.7 Equivalence between ECL and RPL

C.7.1 Having now formalised both the ECL and RPL interpretations, one can confirm the assertion made in section 2.4.1 that the two models are mathematically equivalent. With reference to (4) and (7), (4) could be interpreted as an error components model with $z_{nj} = x_{nj}$ while (7) could be interpreted as a random parameters model with fixed parameters on x_{nj} and random parameters with zero mean on z_{nj}.

Appendix D Distributional Forms

D.1 Distributional form of the random parameters

D.1.1 In estimating RPL, there exists considerable flexibility in specifying the distributional form of the density function $f(\beta)$, and this brings a need for careful analysis. The most common distributional forms are Normal, log Normal, uniform and triangular. The choice of distributional form reflects a balance between a priori theoretical restrictions (e.g. some parameters - such as cost - should have a particular sign) and empirical properties of the model in estimation (i.e. data may support some forms better than others).

D.2 Normal

D.2.1 Normal is often a popular starting point, not least because its properties are flexible, and widely understood. Following from these properties, the Normal is a valid distributional form where, *a priori*, there is no apparent restriction on the sign of the parameter. An example of such a parameter would be an alternative-specific constant (ASC); in a mode choice between train and car, some individuals might ceteris paribus prefer train to car, and others might prefer car to train. Against the attractions of the Normal, one must weight its disadvantages, particularly the fact that the distribution is unbounded. Some observations may be extreme when compared to the mean, and this can have a considerable influence on measurements (e.g. valuations) derived from the model.

D.2.2 Practical solutions to the problems of extreme and 'wrong' sign observations include: specifying the standard deviation of the distribution as some function of the mean, thereby limiting the proportion of such observations, and/or truncating the distribution (e.g. at zero).

D.3 log Normal

D.3.1 An alternative means of restricting the sign of a parameter is to defer to the log Normal distribution. The log Normal considers only the non-negative domain, and this property may be readily exploited to achieve the necessary restriction. Whilst, in this respect, the log Normal may offer a compelling attraction over the Normal, a further problem of the Normal - that of extreme observations – may however be exacerbated. This is because the log Normal transforms the symmetric mesokurtic distribution of the Normal to one that is positively skewed and leptokurtic (e.g. contrast the left-handed panels of Figure 1). Valuations derived from a log Normal parameterisation of RPL may, as of consequence, appear extreme. With reference to earlier discussion, an appropriate response to this problem would be to consider the potential for truncation.

D.4 Triangular

D.4.1 In contrast to the Normal and log Normal, the triangular distribution is bounded, and this may lessen the sensitivity of the model to extreme data points. It does not however impose restriction on sign, at least not without appropriate truncation. Moreover, the triangular may in some circumstances offer an appropriate alternative to the Normal.

D.5 Uniform

D.5.1 The uniform distribution is similarly bounded - specifically in the 0-1 interval.

D.6 Discrete

D.6.1 Although the above commentary has focussed particularly on continuous distributions, MXL imposes no such restriction on the form of the distribution. Whilst continuous distributions will be adequate (and indeed most appropriate) for standard practice, more advanced users may on occasion exploit discrete distributions, particularly where there is interest in 'latent' segmentation of the population.

Appendix E Maximum Simulated Likelihood

E.1 Random Draws

E.1.1 Unlike MNL and NL, MXL is not a 'closed form' model and therefore requires estimation by maximum simulated likelihood. In implementing this method of estimation, the analyst must make judgements with respect to the type and number of draws. With regards to the former, the analyst must select between pseudo-random and quasi-random draws (see Chapter 9 of Train, 2003). An attraction of quasi-random draws is that they employ 'intelligent' sampling, and this achieves some efficiency over pseudo-random draws in the number of draws required to achieve a given level of precision. Further efficiency may be achieved by employing advanced sampling procedures such 'shuffled' or 'scrambled' quasi-random draws.

E.1.2 Since an increase in draws may impose a significant increase in estimation run time, this may be an attractive reason for choosing quasi-random draws over pseudo-random draws. A range of quasi-random sampling methods exist, although most experience thus far has been gained with Halton sequences. It is difficult to offer precise advice on the number of draws, since this will vary on a case-by-case basis. Whilst experience would point towards a recommendation of **at least** 200 Halton sequences, it remains the responsibility of the analyst to demonstrate the stability of his or her model against the adopted number of draws.

E.1.3 Recent research has seen the advent of various algorithms designed to achieve efficiency in the number of draws - and therefore run time - for a given level of precision in the estimated parameters. Examples include so-called 'scrambled' and 'shuffled' Halton sequences. Although these show considerable promise, and are beginning to permeate practice, consensus is yet to be reached on the respective merits of the various candidate methods. If the analyst wishes to exploit such methods, then it is important once again that he or she demonstrate the stability of the model to alternative specification options.

E.2 Run Times

E.2.1 As was alluded to in the previous discussion, maximum simulated likelihood estimation may place significant demands on computer time, and project planning should therefore anticipate and accommodate such activities, both in terms of hardware and the scheduling and duration of tasks. As a rough indication, it is worth acknowledging that if *n* draws were taken, then MXL would take slightly more than *n* times as long as MNL to estimate, presuming that the iteration process for MXL were as efficient as that for MNL (which it is not).

Appendix F Individual-Level Parameters

F.1 Taste variation across segments

F.1.1 In its most basic form, RPL yields estimates of the first and second moments (e.g. mean and standard deviation) of the distribution of tastes across the population of interest. This can be straightforwardly extended to estimate first and second moments for each of several defined population segments.

F.2 Taste variation across individuals

F.2.1 Where individuals supply several repeated choices, this brings an opportunity for the analyst to estimate a taste parameter specific to each individual; the relevant procedure is set out in Chapter 11 of Train, 2003. The approach involves the use of Bayes Rule and Maximum Simulated Likelihood to identify where on a given distribution an individual's tastes lie. The individual's tastes are identified conditional on the distribution of tastes across the population and the individual's choices.

Appendix G Willingness To Pay

G.1 Deriving willingness to pay from choice models

G.1.1 An important inference from travel choice models is the monetary value of travel attributes (e.g. the value of time). This measurement can be derived reasonably straightforwardly from MNL and NL as the ratio of the marginal utility of the attribute to be valued and the marginal utility of cost. For example, where time and cost are specified to be linearly additive within the utility function:

$$V_{ni} = \alpha T_{ni} + \beta C_{ni}$$

and the value of time is given by the ratio of time and cost parameters:

$$VOT = \frac{\partial V / \partial T}{\partial V / \partial C} = \frac{\alpha}{\beta}$$

G.2 Deriving willingness to pay from RPL

G.2.1 The above contrasts with RPL, where parameter ratios such as α/β are postulated to follow a distribution, and the procedure required to derive attribute values becomes more involved.

G.2.2 If only the numerator of α/β is distributed then the value of time can be derived as the mean of the time coefficient divided by the cost coefficient.

G.2.3 If both time and cost are distributed then the procedure is more complicated, involving the use Monte Carlo simulation.

G.3 Implications for distributional form

G.3.1 In specifying parameters as distributed, it is important to acknowledge any requirements on their sign; e.g. value of time should be positive. Such requirements may be imposed through the use of bounded distributions or the truncation of unbounded distributions; see the discussion in Appendix D.

G.4 Practical advice

G.4.1 For practical reasons, and with reference to the earlier discussion, it is useful to fix at least one coefficient during estimation as:

- this helps stabilise the estimation process
- makes the interpretation of the willingness to pay easier
- avoids the potential for unfeasibly large willingness to pay

The usual candidate for the fixed coefficient is cost as it is conceivable that wrong signed time coefficients might exist (i.e. people like spending time travelling) but it is implausible that people would exhibit wrong signed cost coefficients (i.e. people like paying more).