ITLS

WORKING PAPER
ITLS-WP-06-04

Heteroscedastic Control for Random Coefficients and Error Components in Mixed Logit

By

William H Greene* and David A Hensher

March 2005

*Department of Economics
Stern School of Business
New York University
New York USA

(accepted for Transportation Research E)

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-06-04

TITLE: Heteroscedastic Control for Random Coefficients and Error Components in Mixed Logit

ABSTRACT: Developments in simulation methods, and the computational power that is now available, have enabled open-form discrete choice models such as mixed logit to be estimated with relative ease. The random parameter (RP) form has been used to identify preference heterogeneity, which can be mapped to specific individuals through re-parameterisation of the mean and/or variance of each RP's distribution. However this formulation depends on the selection of random parameters to reveal such heterogeneity, with any residual heterogeneity forced into the constant variance condition of the extreme value type 1 distribution of the classical multinomial logit model. In this paper we enhance the mixed logit model to capture additional alternative-specific unobserved variation not subject to the constant variance condition, which is independent of sources revealed through random parameters. An empirical example is presented to illustrate the additional information obtained from this model.

KEY WORDS: ***Mixed logit, error components, discrete choice, elasticities, value of time savings***

AUTHORS: William H Greene & David A Hensher

CONTACT: Institute of Transport and Logistics Studies **(C37)**
An Australian Key Centre
The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itlsinfo@itls.usyd.edu.au
Internet: http://www.itls.usyd.edu.au

DATE: March 2006

1. Introduction

Developments in simulation methods, and the computational power that is now available, have enabled open-form discrete choice models such as mixed logit to be estimated with relative ease (see Train 2003, Hensher *et al*. 2005). The popular random parameter (RP) form has been used to identify preference heterogeneity[1], which can be mapped to specific individuals through re-parameterisation of the mean and/or variance of each RP's distribution (Greene *et al.* 2005). However this formulation depends on the selection of random parameters to reveal such heterogeneity, with any residual heterogeneity resident in the constant variance condition of the EV1 distribution of the multinomial logit model. In this paper we enhance the mixed logit model to capture additional alternative-specific unobserved variation not subject to the constant variance condition, which is independent of sources revealed through random parameters. It can be mapped to specific individuals through deep parameterisation of the variance. Some of the features build on developments by Brownstone and Train (1999) that were extended by Ben-Akiva *et al.* (2001) and Gopinath *et al.* (2004). This further extension of mixed logit may offer the way forward in empirically separating scale from variance in open-form choice models.

The paper is organised as follows. In Section 2 we set out the extensions of the mixed logit model to highlight the extended set of sources of preference heterogeneity across individuals and alternatives. Section 3 presents an application of the full model to a stated mode choice experiment for commuter trips in Sydney in 2003 with empirical evidence set out in Section 4. Conclusions are drawn in Section 5.

2. Heteroscedastic Control for Random Coefficients and Error Components in Mixed Logit

We assume that sampled individuals $q = 1,...,Q$ face a choice among J alternatives, denoted $j = 1,...,J$ in each of T choice settings, $t = 1,...,T$.[2] The canonical random utility model associates utility for individual q

$$U_{q,j,t} = \boldsymbol{\beta}'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} \qquad (1)$$

with each alternative in each choice situation. We assume the individual considers the full choice set in each situation and makes the choice associated with the highest utility. The $K\times1$ vector $\mathbf{x}_{q,j,t}$ defines the full set of explanatory variables including attributes of the choices, socioeconomic characteristics of the individual and descriptors of the decision context and choice task[3] itself in situation t. For convenience, define $\mathbf{X}_{q,t}$ to be the full set of explanatory variables for all choices in choice situation t, and $\mathbf{X}_q$ likewise

[1] Heterogeneity refers to differences across individuals in their preferences. From an econometric perspective there are two types of heterogeneity; that which is related to observed attributes of the individual, called observed heterogeneity, and that which cannot be related to the observed attributes of the individual, called unobserved heterogeneity. Observed heterogeneity is captured by entering the relevant attributes of the individual while unobserved heterogeneity is captured by entering random terms. We acknowledge the request from a referee to explicitly clarify the meaning of heterogeneity.
[2] In our implementation of this model, J may vary across q and t and T may vary across q. But it is notationally convenient to assume both are fixed for now.
[3] If a stated choice experiment is the data source, then the complexity of the choice task as defined by the number of choice situations, number of alternatives, attribute ranges, data collection method, etc. can also be included to condition specific parameters associated with attributes of alternatives.

for the full set of choice situations, t=1,…,T. The components $\boldsymbol{\beta}$ and ε are not observed by the analyst. The conventional departure point for analysis of this model is the assumption of identical independent (IID) type 1 extreme value distributions (EV1) for $\varepsilon_{q,j,t}$, which gives rise to the multinomial logit (MNL) probability model,

$$\text{Prob[choice } j|\text{individual } q, \mathbf{X}_{q,t}, \text{ choice setting } t] = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{q.j.t})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{x}_{q.j.t})} \;. \tag{2}$$

There are (at least) two major problems with this model. First, the IID-EV1 assumption is extremely restrictive and induces the 'independence from irrelevant alternatives' (IIA) property in the model. This is a well known major shortcoming of the MNL model. The second issue is that the canonical MNL model as stated fails to capture preference heterogeneity of any sort not embodied in the individual characteristics and the IID-EV1 disturbances. Various models have been proposed to extend the MNL model in this direction. The mixed logit model provides a rich set of features that will allow the analyst to accommodate observed and unobserved heterogeneity from a variety of sources[4].

We introduce preference heterogeneity in the model by specifying the individual specific random parameters, in equation (3).

$$\beta_{q,k} = \bar{\beta}_k + \boldsymbol{\delta}_k'\mathbf{z}_q + \gamma_{q,k}v_{q,k},\; k = 1,\ldots,K \tag{3}$$

where

$v_{q,k}$ = a random variable with $E[v_{q,k}]=0$ and $\text{Var}[v_{q,k}] = a_k^2$, a known constant
$\gamma_{q,k}$ = $\sigma_k\times\exp[\boldsymbol{\eta}_k'\mathbf{h}_q]$.

The parameters are randomly distributed over individuals with means and variances that can depend on individual characteristics, $\mathbf{z}_q$ and $\mathbf{h}_q$. The components of this random parameters model are

$\bar{\beta}_{q,k}$ = fixed mean (in name only, since the mean includes $\boldsymbol{\delta}_k'\mathbf{z}_q$),

$\boldsymbol{\delta}_k'\mathbf{z}_q$ = observed heterogeneity around the mean,

σ_k = fixed part of the standard deviation of the random parameter $\beta_{q,k}$,

$\exp[\boldsymbol{\eta}_k'\mathbf{h}_q]$ = observed heterogeneity associated with the distribution of $\beta_{q,k}$.

The data vectors $\mathbf{z}_q$ and $\mathbf{h}_q$ contain individual specific characteristics such as socio-demographic factors[5] which may overlap or be identical. Parameters to be estimated are $\bar{\beta}_k$, $\boldsymbol{\delta}_k$, σ_k and $\boldsymbol{\eta}_k$, k = 1,…,K. The structural random variable $v_{q,k}$ endows the random parameter with its statistical properties. In principle, all K parameters may be random and heterogeneously distributed in this fashion, although in the typical model, not all

[4] Revelt and Train (1998) and Train (1998) developed the first mixed logits on repeated choices. Revelt and Train (1998) were the first authors, to our knowledge, to introduce the term 'mixed logit'.
[5] In general these characteristics can be any source that is observation specific (as distinct from alternative specific).

parameters will be random.[6] In some cases, the model may also specify certain restrictions on δ_k, σ_k and/or η_k. For example, homoscedasticity is achieved by imposing $\eta_k = \mathbf{0}$, while a hierarchical model specification with non-random parameters is specified with $\sigma_k = 0$ (which implies that $\eta_k = 0$). The assumption that Var[$v_{q,k}$] is known is not substantive, as the unknown scaling is contained in $\gamma_{q,k}$. For example, if $v_{q,k}$ is normal, a_k equals 1; if it is standard uniform, it equals 1/12.

Collecting all K parameters in a column vector, we have

$$\beta_q = \bar{\beta} + \Delta \mathbf{z}_q + \Gamma_q \mathbf{v}_q \tag{4}$$

where $E[\mathbf{v}_q] = \mathbf{0}$ and Var[$\mathbf{v}_q$] = $\mathbf{A}$, where $\mathbf{A}$ is a known diagonal matrix. For example, if all parameters are normally distributed, then $\mathbf{A} = \mathbf{I}$. We now allow two additional elements to enter this random parameters specification. First, the parameters may be correlated, accommodated by specifying a mixing of the primitive underlying random variables:

$$\mathbf{v}_q = \mathbf{R}\mathbf{v}_q^* \tag{5}$$

where $\mathbf{R}$ is a lower triangular matrix with ones on the main diagonal and v_q^* is the vector of independent variables, with zero means and known variances a_k^2. The correlation is parameterized by the nonzero elements below the diagonal. These below diagonal elements are additional parameters to be estimated. The covariance matrix of the random parameter vector is then

$$\text{Var}[\beta_q \mid \mathbf{z}_q, \mathbf{h}_q] = \Gamma_q \mathbf{R}\mathbf{A}\mathbf{R}'\Gamma_q. \tag{6}$$

This is an unrestricted matrix, subject to the structural form specified earlier for Γ_q. A second extension allows preferences to evolve over 'time' or over the sequence of choices in a repeated choice experiment setting. The 'autocorrelation' is specified as:

$$v_{q,k,t}^* = \rho_k \, v_{q,k,t-1}^* + w_{q,k,t}^*. \tag{7}$$

where $w_{q,k,t}^*$ is now the underlying structural random variable. This adds K autocorrelation parameters, $\rho_1,\ldots,\rho_K$, to the model, also to be estimated.[7] Since β_q can contain alternative-specific constants which may be correlated, this specification can induce correlation across alternatives. It follows immediately that the model does not impose the IIA assumption.[8] Restrictions can be imposed at numerous points in the model to produce a wide variety of specifications.

For convenience, we collect all the structural parameters, $\bar{\beta}$, Δ, $\mathbf{R}$, (σ_k,η_k,ρ_k, k=1,…,K), in a parameter set $\mathbf{\Omega}$. The marginal distribution of β_q is induced by the distribution of $\mathbf{v}_q$, which we denote as[9]:

$$f_\beta(\beta_q \mid \mathbf{\Omega}, \mathbf{z}_q, \mathbf{h}_q) = f_v(\bar{\beta} + \Delta \mathbf{z}_q + \Gamma_q \mathbf{v}_q \mid \mathbf{\Omega}, \mathbf{z}_q, \mathbf{h}_q). \tag{8}$$

---

6 This selection of a subset of random parameter is potentially a weakness of the pure random parameter specification of mixed logit in that all sources of preference heterogeneity are totally dependent on the random parameter selection. Additional sources of preference heterogeneity that are not dependent on the random parameters for their revelation that are specified herein is an appealing extension.

7 With this change, assuming stationarity, the matrix A in Var[$\beta_q \mid z_q, h_q$] is replaced with A(ρ) = diag[$a_k^2/(1-\rho_k^2)$,k=1,…,K].

8 See McFadden and Train (2000) for details.

9 The Jacobian of the transformation Γ_q^{-1}, from v to β would appear in the density as well.

where $\boldsymbol{\Gamma}_q =$ the matrix of structural parameters, Γ multipled by a diagonal matrix, say Λ_q which contains the observation specific standard deviations on the diagonals. This is how we build heteroscedasticity in the parameters into the model.

The reduced form of the choice model is now

$$U_{q,j,t} = \boldsymbol{\beta}_q'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} \tag{9}$$

and

$$\text{Prob}_{q,t}\,[j|\mathbf{X}_{q,t},\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q\,] = \frac{\exp(\boldsymbol{\beta}_q'\mathbf{x}_{q.j.t})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}_q'\mathbf{x}_{q.j.t})}\ . \tag{10}$$

Any remaining heterogeneity not already accounted for is treated as unobserved and resides in $\varepsilon_{j,t,q}$. This will also capture misspecification effects such as treating as fixed a parameter which should be random.

The ‘kernel logit’ model suggested by Ben-Akiva *et al.* (2001), based on an idea first proposed by Brownstone and Train (1999)[10], incorporates additional unobserved heterogeneity through effects that are associated with the individual’s preferences within the choices. These appear as $M \leq J$ additional random effects,

$$U_{q,j,t} = \boldsymbol{\beta}_q'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} + c_{j1}W_{q,1} + c_{j2}W_{q,2} + \ldots + c_{jM}W_{q,M}, \tag{11}$$

where the $W_{m,q}$ are normally distributed effects with zero mean, $m = 1,\ldots,M \leq J$ and $c_{jm} = 1$ if m appears in utility function j.[11] This specification can produce a simple ‘random effects’ model if all J utilities share a single error component,

$$U_{q,j,t} = \boldsymbol{\beta}_q'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} + W_q,\ \ j = 1,\ldots,J. \tag{12}$$

or an error components sort of model if one and only one alternative-specific parameter appears in each utility function, as in (13).

$$U_{q,j,t} = \boldsymbol{\beta}_q'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} + W_{q,j},\ j = 1,\ldots,J. \tag{13}$$

If groups of utility functions each contain a common subset of the error components across specific nests of alternatives, then we can specify the ‘nested’ system in (14):

$$\begin{aligned} U_{q,1,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,1,t} + \varepsilon_{q,1,t} + W_{q,1} \\ U_{q,2,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,2,t} + \varepsilon_{q,2,t} + W_{q,1} \\ U_{q,3,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,3,t} + \varepsilon_{q,3,t} + W_{q,2} \\ U_{q,4,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,4,t} + \varepsilon_{q,4,t} + W_{q,2} \end{aligned} \tag{14}$$

[10] The Ben-Akiva et al paper was a reaction to the suggestion in Brownstone and Train, pointing out that identification can be difficult to assess in mixed models with these kinds of error components for alternatives and nests.
[11] Issues of specification and identification are discussed in Ben-Akiva *et al.* (2001).

and even a cross nested model if the groups of error components overlap, as in the following example (15),

$$
\begin{aligned}
U_{q,1,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,1,t} + \varepsilon_{q,1,t} + W_{q,1} + W_{q,2},\\
U_{q,2,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,2,t} + \varepsilon_{q,2,t} + W_{q,1} + W_{q,2},\\
U_{q,3,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,3,t} + \varepsilon_{q,3,t} + W_{q,2} + W_{q,3} + W_{q,4},\\
U_{q,4,t} &= \boldsymbol{\beta}_q'\mathbf{x}_{q,4,t} + \varepsilon_{q,4,t} + W_{q,3} + W_{q,4}.
\end{aligned}
\tag{15}
$$

This extension of the mixed logit model entails capturing additional unobserved variance that is alternative-specific through a mixture formulation which imposes a normal distribution on such information across the sampled population. The standard deviation of these normals can be parameterised for each alternative with special cases in which there are cross-alternative equality constraints on the estimated standard deviations. Through cross-alternative constraints we can permit an alternative to appear in more than one subset of alternatives, giving it the appearance of a nested structure.

Our generalization extends the Brownstone and Train (1999) model in two respects. First, we allow the same kind of variance heterogeneity in the error components for alternatives and nests of alternatives as in the random parameters part:

$$\mathrm{Var}[W_{m,q}] = [\theta_m \exp(\boldsymbol{\tau}_m'\mathbf{h}_q)]^2 \tag{16}$$

Second, we combine this specification with the full random parameters model laid out earlier.[12] Collecting all results, the full mixed logit model is given by equations (17-22).

$$U_{q,j,t} = \boldsymbol{\beta}_q'\mathbf{x}_{q,j,t} + \varepsilon_{q,j,t} + \sum_{m=1}^{M} c_{j,m} W_{q,m} \tag{17}$$

$$\boldsymbol{\beta}_q = \overline{\boldsymbol{\beta}} + \boldsymbol{\Delta}\mathbf{z}_q + \boldsymbol{\Gamma}_q\mathbf{v}_q \tag{18}$$

$$\mathbf{v}_q = \mathbf{R}\mathbf{v}_q^* \tag{19}$$

$$v_{q,k,t}^* = \rho_k\, v_{q,k,t-1}^* + w_{q,k,t}^* \tag{20}$$

$$\mathrm{Var}[v_{q,k}^*] = [\sigma_k \times \exp(\boldsymbol{\eta}_k'\mathbf{h}_q)]^2 \tag{21}$$

$$\mathrm{Var}[W_{m,q}] = [\theta_m \exp(\boldsymbol{\tau}_m'\mathbf{h}_q)]^2 \tag{22}$$

The conditional choice probability is now

[12] Ben-Akiva *et al* (2001) extend the basic model somewhat by imposing a factor analytic structure on the set of kernels. This achieves a small amount of generality in allowing the variables that appear in the utility functions to be correlated. With respect to the behavioural model, little is actually obtained by this, since the assumed independent kernels above may be mixed in any fashion in the utility functions.

$$L_{q,j,t} \; = \; \text{Prob}_{q,t}\,[j|\,\mathbf{X}_{q.t},\boldsymbol{\Omega},\,\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q,\mathbf{W}_q\,] = \frac{\exp(\boldsymbol{\beta}_q'\mathbf{x}_{q.j.t} + \Sigma_{m=1}^{M} c_{jm}W_{mq})}{\sum_{j=1}^{J}\exp(\boldsymbol{\beta}_q'\mathbf{x}_{q.j.t} + \Sigma_{m=1}^{M} c_{jm}W_{mq})}\;. \tag{23}$$

The *unconditional* choice probability is the expected value of this logit probability over all the possible values of $\boldsymbol{\beta}_q$ and $\mathbf{W}_q$, that is, integrated over these values, weighted by the joint density of $\boldsymbol{\beta}_q$ and $\mathbf{W}_q$. We assume that $\mathbf{v}_q$ and $\mathbf{W}_q$ are independent, so this is just the product. Thus, the unconditional choice probability is

$$P_{jtq}(\mathbf{X}_{t,q},\mathbf{z}_q,\mathbf{h}_q,\boldsymbol{\Omega}) \; = \; \text{Prob}_{q,t}\,[j\,|\mathbf{X}_{t,q},\boldsymbol{\Omega},\,\mathbf{z}_q,\mathbf{h}_q] \; =$$

$$\int_{\mathbf{W}_q}\int_{\boldsymbol{\beta}_q} L_{q,j,t}(\boldsymbol{\beta}_q\,|\,\mathbf{X}_{q,t},\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q,\mathbf{W}_q)\,f(\boldsymbol{\beta}_q\,|\,\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q)\,f(\mathbf{W}_q\,|\,\boldsymbol{\Omega},\mathbf{h}_q)\,d\boldsymbol{\beta}_q d\mathbf{W}_q\,. \tag{24}$$

Thus, the *unconditional* probability that individual q will choose alternative j given the specific characteristics of their choice set and the underlying model parameters, is equal to the expected value of the conditional probability as it ranges over the possible values of $\boldsymbol{\beta}_q$. and $\mathbf{W}_q$. Finally, the contribution of individual q to the likelihood for the full sample is the product of the T conditionally (on $\mathbf{v}_q$ and $\mathbf{W}_q$) independent choice probabilities. The log likelihood is then formed as usual. The contribution of individual q is

$$P_q\,(\mathbf{X}_q,\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q) \; =$$

$$\int_{\mathbf{W}_q}\int_{\boldsymbol{\beta}_q}\prod_{t=1}^{T} L_{q,j,t}(\boldsymbol{\beta}_q\,|\,\mathbf{X}_{q,t},\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q,\mathbf{W}_q)\,f(\mathbf{v}_q\,|\,\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q)\,f(W_q\,|\,\boldsymbol{\Omega},\mathbf{h}_q)\,d\mathbf{v}_q d\mathbf{W}_q \tag{25}$$

and the full log likelihood is

$$\log L(\boldsymbol{\Omega}) = \sum_{q=1}^{Q}\log\int_{\mathbf{W}_q}\int_{\boldsymbol{\beta}_q}\begin{matrix}\prod_{t=1}^{T} L_{q,j,t}(\boldsymbol{\beta}_q\,|\,\mathbf{X}_{tq},\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q,\mathbf{W}_q)\times \\ f(\mathbf{v}_q\,|\,\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q)\,f(\mathbf{W}_q\,|\,\boldsymbol{\Omega},\mathbf{h}_q)\end{matrix}\,d\mathbf{v}_q d\mathbf{W}_q \tag{26}$$

The integrals in (26) cannot be computed analytically because there are no closed forms solutions. However, the full expression is in the form of an expectation, which suggests that it can be approximated satisfactorily with Monte Carlo integration. Let $\mathbf{v}_{qr}$ denote the rth of R random draws from the population of $\mathbf{v}_q$ and $\mathbf{W}_{qr}$ be an accompanying random draw from the M-variate standard normal population. Using these draws, the logit probability is calculated. This process is repeated for many draws, and the mean of the resulting simulated likelihood values is taken as the approximate choice probability giving the simulated log likelihood,

$$\log L_S\,(\boldsymbol{\Omega}) \; = \; \sum_{q=1}^{Q}\log\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T} L_{q,j,t}(\boldsymbol{\beta}_q\,|\,\mathbf{X}_{q,t},\boldsymbol{\Omega},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_{qr},\mathbf{W}_{qr},)$$

$$= \; \sum_{q=1}^{Q}\log\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}\frac{\exp[(\overline{\boldsymbol{\beta}}+\boldsymbol{\Delta}\mathbf{z}_q+\boldsymbol{\Gamma}_q\mathbf{v}_{q,r})'\mathbf{x}_{q,j,t}+\Sigma_{m=1}^{M}c_{j,m}W_{q,m,r}]}{\sum_{m=1}^{J}\exp[(\overline{\boldsymbol{\beta}}+\boldsymbol{\Delta}\mathbf{z}_q+\boldsymbol{\Gamma}_q\mathbf{v}_{q,r})'\mathbf{x}_{q,m.t}+\Sigma_{m=1}^{M}c_{j,m}W_{q,m,r}]}\,. \tag{27}$$

This function is smooth and continuous in the elements of $\boldsymbol{\Omega}$ and can be maximized by conventional methods. Gourieroux and Monfort (1996) or Train (2003) provide a discussion of this form of maximum simulated likelihood estimation. With sufficiently large *R* (number of draws), the simulated function provides an adequate approximation to the actual function for likelihood based estimation and inference.[13]

3. An Empirical Application

Our empirical study uses a mode choice data set of commuting trips of a sample of residents of the north-west sector of the Sydney metropolitan area interviewed in 2003. The principle aim of the study was to establish the preferences of residents within the study area for private and public transport modes for commuting. Once known, the study called for the preferences to be used to forecast patronage levels for currently non-existing transport modes, specifically possible new heavy rail, light rail or busway modes. Independent of the 'new' mode type, the proposed infrastructure is expected to be built along the same corridor.

To capture information on the preferences of residents, a stated choice (SC) experiment[14] was generated and administered using computer aided personal interview (CAPI) technology. Sampled residents were invited to review a number of alternative main and access modes (both consisting of public and private transport options) in terms of levels of service and costs within the context of a recent trip and to choose the main mode and access mode that they would use if faced with the same trip circumstance in the future. Each sampled respondent completed 10 choice tasks under alternative scenarios of attribute levels, choosing the preferred main and access modes in each instance.

The experiment was complicated by the fact that alternatives available to any individual respondent undertaking a hypothetical trip depended not only on the alternatives the respondent had available at the time of the 'reference' trip, but also on the destination of the trip. If the trip undertaken was intra-regional, then the existing busway and heavy rail modes could not be considered viable alternatives as neither mode travels within the bounds of the study area. If on the other hand, the reference trip was inter-regional (e.g., to the CBD), then respondents could feasibly travel to the nearest busway or heavy rail train station (outside of the origin region) and continue their trip using these modes. Furthermore, not all respondents had access to a private vehicle for the reference trip, either due to a lack of ownership or non-availability at the time when the trip was made. Given that the objective of the study was to derive an estimate of patronage demand, the lack of availability of privately-owned vehicles (either through random circumstance or non ownership) should be accounted for in the SC experiment. More details are given in Hensher and Rose (in press).

The mode choice experimental design has 47 attributes (46 in four levels and one in six levels for the blocks) in 60 runs; that is there are six blocks of ten choice sets each. The design is D-optimal and almost orthogonal, with correlations between ± 0.06. This design allows the estimation of all alternative-specific main effects, assuming linear marginal utilities. Within each block the order of the runs has been randomised to control for order effect biases.

---

[13] In our application, we use Halton sequences rather than random draws to speed up and smooth the simulations. See Bhat (2001), Train (2003) or Greene (2003) for discussion.
[14] Readers unfamiliar with stated choice methods should consult Louviere *et al.* (2000) or Hensher *et al.* (2005).

The trip attributes associated with each mode are summarised in Table 1. Whereas the times and costs associated with currently available modes[15] are obtained from the respondents in earlier survey screens, the attribute levels for the currently non-existent public transport modes are established from other sources. The levels shown in Table 2 were provided by the Ministry of Transport as their best estimates of the *most likely* fare and service levels. To establish the likely access location to the new modes, respondents were also asked to view a map that listed potential new stations and asked to choose a particular station, which is used in the software to derive the access and linehaul travel times and fares.

Table 1: Trip Attributes in Stated Choice Design

For existing public transport modes	**For new public transport modes**	**For the existing car mode**
Fare (one-way)	Fare (one-way)	Running Cost
In-vehicle travel time	In-vehicle travel time	In-vehicle Travel time
Waiting time	Waiting time	Toll Cost (One way)
Access Mode: Walk time	Transfer waiting time	Daily Parking Cost
Car time	Access Mode: Walk time	Egress time
Bus time	Car time	
Bus fare	Bus time	
Egress time	Access Mode Fare (one-way)	
	Bus fare	
	Egress time	

All design attributes had four levels. These were chosen as the following variations around the base level: -25%, 0%, +25%, +50%. An example of a stated choice screen is shown as Figure 1, derived as one row of a D-optimal stated choice design. Each respondent was asked to review the access modes and choose one associated with each main mode alternative, and then to choose the preferred main mode. This was repeated 10 times.

Table 2: Base times and costs for new public transport modes

Station Location:	**Dollars** **$**	**Busway** **min**	**Heavy rail** **min**	**Light Rail** **min**
Mungerie Park	1.8	33	22	33
Burns Road	1	27	18	27
Norwest Business Park	1	22.5	15	22.5
Hills Centre	1	18	12	18
Castle Hill	0.2	13.5	9	13.5
Franklin Road	0.2	7.5	5	7.5
Beecroft				

15 Respondents were shown generic pictures to assist them in understanding the modes.

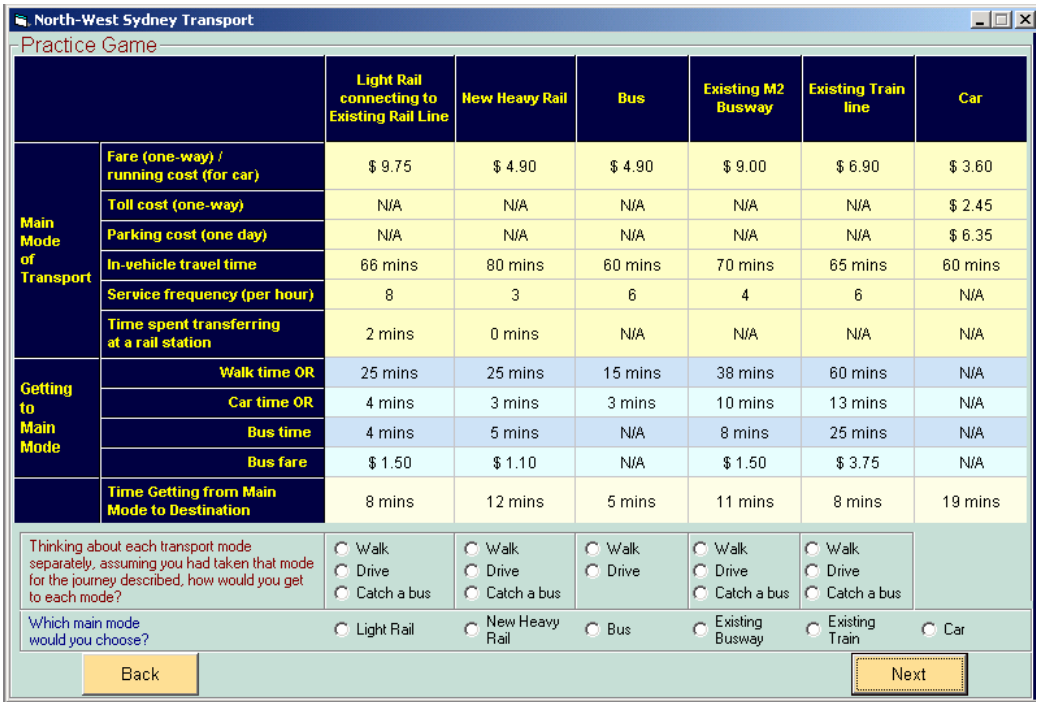

Figure 1: Example stated preference choice screen

The sample comprises 223 commuters or 2230 observations for model estimation. Table 3 shows the descriptive statistics for the work segment. The average age is 43.1 years with a mean personal income of $Aud64,100. 89.24 percent of the sample had access to a car for the surveyed trip.

Table 3: Descriptive statistics for Commuters

	Mean	Std. Deviation	Minimum	Maximum
Age	43.1	12.5	24	70
Hours worked per week	37.6	14.6	0	70
Annual Personal Income	64.1	41.8	0	140
Household size	3.78	2.30	1	8
No. of children in HH	1.05	1.09	0	4
Gender	50.4	-	0	1

4. The Results

Five mixed logit models were estimated, beginning with the base model and building up to the most general model. The models are identified as follows and the results summarised in Table 4:

ML1: Base model with random parameters only
ML2: ML1 plus observed heterogeneity around the mean of random parameters
ML3: ML2 plus heteroskedasticity around the standard deviations of random parameters
ML4: ML3 plus standard deviation of error components for alternatives and nests of alternatives

ML5: ML4 plus heterogeneity in variance of error components for alternatives and nests of alternatives

Table 4: Summary of Empirical Results[16]: Commuter Trips

Note: All public transport = (new heavy rail, new light rail, new busway, bus, train, busway); time is in minutes; fares and cost is in dollars ($2003). T-values in brackets 2230 observations, 200 Halton draws.

Attribute	Alternatives	ML1	ML2	ML3	ML4	ML5
New light rail constant	New light rail	2.411	3.313	2.978	4.442	5.011
		(5.0)	(6.1)	(5.7)	(4.68)	(5.3)
New busway constant	New busway	1.019	1.933	1.561	2.939	3.487
		(2.1)	(3.5)	(2.8)	(3.1)	(3.7)
Existing bus constant	Bus	1.393	2.273	1.852	3.255	3.808
		(3.0)	(4.4)	(3.6)	(3.5)	(4.1)
Train constant	Existing and new Train	1.709	2.609	2.246	3.657	4.213
		(3.6)	(4.9)	(4.4)	(3.9)	(4.5)
Existing busway constant	busway	1.266	2.183	1.801	3.178	3.714
		(2.7)	(4.1)	(3.4)	(3.4)	(4.0)
Random parameters - constrained triangular:						
Main mode fares	All public transport	-0.2505	-0.3536	-0.3512	-0.3723	-0.3853
		(-12.1)	(-10.1)	(-10.4)	(-9.3)	(-9.4)
Car mode running and toll cost	Car	-0.1653	-0.1764	-0.1876	-0.2152	-0.2182
		(-3.3)	(-3.3)	(-3.4)	(-2.8)	(-2.9)
Car parking cost	Car	-0.0340	-0.0377	-0.0443	-0.0571	-0.0558
		(-2.7)	(-2.7)	(-3.0)	(-2.7)	(-2.8)
Main mode in-vehicle time	All public transport	-0.0640	-0.0744	-0.0713	-0.0773	-0.0785
		(-19.3)	(-14.4)	(-18.5)	(-15.3)	(-15.1)
Access and wait time	All train and light rail	-0.0699	-0.0716	-0.0762	-0.0811	-0.0828
		(-9.5)	(-9.5)	(-9.8)	(-9.1)	(-9.2)
Access time	All bus and busway	-.0756	-0.0808	-0.0839	-0.0929	-0.0942
		(-6.9)	(-7.1)	(-6.5)	(-6.4)	(-6.3)
Wait time	All bus and busway	-0.0882	-0.0907	-0.1026	-0.1034	-0.1048
		(-3.1)	(-3.06)	(-3.2)	(-3.0)	(-3.0)
Main mode in-vehicle time	Car	-.0728	-0.0791	-0.0859	-0.0796	-0.0732
		(-6.2)	(-6.2)	(-7.0)	(-5.2)	(-4.8)
Egress travel time	All public transport	-0.0145	-0.0142	-0.0151	-0.0169	-0.0181
		(-2.6)	(-2.5)	(-2.7)	(-2.8)	(-3.0)
Egress travel time	Car	-0.0814	-0.0876	-0.0892	-0.1193	-0.1084
		(-3.5)	(-3.5)	(-2.8)	(-2.5)	(-2.5)

[16] A referee comments that "multiple mixture runs must be conducted and a measure of variation in parameter coefficients must be report (e.g., std. deviation in parameter estimates). Theoretically, we cannot make statistical inference from a distribution using a single point." Unfortunately, this suggestion does not address a relevant issue in our estimations. We used Halton draws to perform our integrations, so there is no simulation variance. If we repeated the estimation, we would get the exact same estimates. In fact, there is no simulation variance because the estimates are not based on simulations. We have used the Halton technique to evaluate certain integrals. There will be an approximation error, of course. Our only control over that is to use many Halton draws, which we have done. The interpretation of the estimates as a sample of one is not correct, however. There are many other settings in which researchers must resort to approximations to evaluate integrals, such as random effects probit models which use Hermite quadrature to approximate integrals, and even the most mundane univariate probit model which uses a ratio of polynomials to approximate the standard normal cdf. The MLEs obtained in these settings are not samples of one; they are maximisers of an approximation to the log likelihood function which cannot be evaluated exactly.

Access bus mode fare	Where bus is access mode	-0.1067 (-2.9)	-0.1916 (-3.65)	-0.1981 (-3.8)	-0.2118 (-3.8)	-0.2167 (-3.8)
Non-random parameters:						
Inside Study area	New heavy rail	-1.119 (-2.8)	-1.207 (-2.9)	-1.300 (-3.3)	-1.497 (-3.4)	-1.419 (-3.2)
Inside Study area	New light rail	-1.104 (-3.1)	-1.164 (-3.2)	-1.249 (-3.6)	-1.443 (-3.7)	-1.383 (-3.5)
Personal income	All public transport	-0.0122 (-4.1)	-0.0278 (-6.2)	-0.0211 (-4.9)	-0.0266 (-4.2)	-0.0326 (-5.0)
Gender (male = 1)	All Public transport	1.905 (7.1)	1.969 (6.9)	2.662 (7.6)	3.437 (6.2)	3.493 (6.7)
Heterogeneity around Mean:						
In-vehicle time * personal income	All public transport		0.000124 (2.6)			
Main mode fares * personal income	All public transport		0.00135 (3.8)	0.00078 (1.90)	0.00079 (1.8)	0.00090 (2.0)
Access bus fare * personal income	All public transport except existing bus		0.00143 (2.4)	0.00146 (2.3)	0.00160 (2.3)	0.00164 (2.4)
Heterogeneity around Standard deviation:						
In-vehicle time *gender	All public transport			0.4007 (4.3)	0.4195 (4.5)	0.4270 (4.6)
Main mode fares * gender	All public transport			0.6384 (4.4)	0.7049 (4.4)	0.6727 (4.2)
Error components for alternatives and nests of alternatives parameters:						
Standard deviation	New light rail, new heavy rail, new busway, existing busway				0.8659 (2.2)	1.010 (2.9)
Standard deviation	Existing bus and heavy rail				0.2068 (.32)	0.0814 (.13)
Standard deviation	Car				3.021 (4.0)	11.158 (2.3)
Heterogeneity around standard deviation of error components effect:						
Age of commuter	Car					-0.0366 (-2.3)
Log-likelihood at convergence		-2464.3	-2451.7	-2442.1	-2435.9	-2428.7
Pseudo-R^2		0.3101	0.3135	0.3161	0.3176	0.3195

The access mode travel time relates to the chosen access mode associated with public transport main mode.

Table 5 summaries the mean and standard deviation of the parameter estimates associated with each of the specifications for each attribute and its associated decomposition.

Table 5: Mean and Standard deviation of random parameter estimates for entire representation of each attribute from relatively simple to more complex models (note: except for ML1 which has a single parameter, the other models are complex representations of multiple parameters from Table 4)

Mean	invtpt	costpt	acwt	eggt	crpark	accbusf	waittb	acctb	crcost	crinvt	creggt
Ave ML1	-0.0640	-0.2505	-0.0699	-0.0145	-0.0340	-0.1067	-0.0882	-0.0756	-0.1653	-0.0727	-0.0814
Std Dev ML1	0.0245	0.1003	0.0278	0.0059	0.0139	0.0435	0.0359	0.0306	0.0670	0.0283	0.0329
Ave ML2	-0.0680	-0.2744	-0.0733	-0.0148	-0.0407	-0.1125	-0.0916	-0.0831	-0.1821	-0.0869	-0.0907
Std Dev ML2	0.0287	0.1385	0.0291	0.0060	0.0166	0.1014	0.0373	0.0336	0.0738	0.0367	0.0366
Ave ML3	-0.0788	-0.3581	-0.0819	-0.0258	-0.0558	-0.1241	-0.1071	-0.0859	-0.2789	-0.1146	-0.1297
Std Dev ML3	0.0391	0.2771	0.0269	0.0465	0.0075	0.1112	0.0664	0.0002	0.1868	0.0556	0.0034
Ave ML4	-0.0923	-0.4232	-0.0942	-0.0286	-0.0763	-0.1366	-0.1231	-0.1006	-0.3866	-0.0864	-0.1735
Std Dev ML4	0.0476	0.3374	0.0299	0.0489	0.0373	0.1157	0.0979	0.0007	0.3003	0.0016	0.0161
Ave ML5	-0.0923	-0.4152	-0.0941	-0.0348	-0.1151	-0.1340	-0.1137	-0.1001	-0.2678	-0.1140	-0.1524
Std Dev ML5	0.0473	0.3236	0.0317	0.0620	0.1165	0.1051	0.0837	0.0006	0.1445	0.0602	0.0061

invtpt = invehicle time for public transport (PT); costpt = public transport fares; acwt = access and wait time for light and heavy rail; eggt = aggress time for PT; crpark = car parking cost; accbusf =access bus mode fare; waittb = wait time for bus and busway; acctb =access time for bus and busway; crcost = car running cost; crinvt = car invehicle time; creggt = egress time from car.

All random parameters have a triangular distribution[17]. We investigated model specifications with unconstrained and constrained[18] triangular distributions and found that the specification that constrained the standard deviation to equal the mean gave the better overall model fit, as well as ensuring a behaviourally meaningful sign for estimated parameters across the entire distribution[19]. The entire set of modal attributes are specified with random parameters and all are statistically significant and of the expected sign. The models progressively improve in overall goodness of fit from a pseudo-r^2 of 0.3101 for the base model through to 0.3195 for the fully generalised model ML5. We find that personal income is a statistically significant source of influence on preference heterogeneity for main mode public transport fares and across all models (ML2-ML5), and for bus fares on the access mode (which is available to all public modes except current bus), reducing the marginal disutility of fares for all public transport modes as personal income increases. Although personal income has an influence on public transport in-vehicle time in ML2, it is statistically non-significant when we move to incorporate heterogeneity around the standard deviation of the random parameter and the error components for alternatives and nests of alternatives.

When we add in observed heterogeneity around the standard deviations of random parameters, we find that gender has a statistically significant influence on in-vehicle travel time and fare for all public transport modes. The positive sign on both travel time and fares suggests that male commuters are much more heterogeneous in terms of the

[17] The triangular distribution was first used for random coefficients by Train and Revelt (2000) and Train (2001), later incorporated into Train (2003). Hensher and Greene (2003) also used it and it is increasingly being used in empirical studies. Let c be the centre and s the spread. The density starts at c-s, rises linearly to c, and then drops linearly to c+s. It is zero below c-s and above c+s. The mean and mode are c. The standard deviation is the spread divided by $\sqrt{6}$; hence the spread is the standard deviation times $\sqrt{6}$. The height of the tent at c is 1/s (such that each side of the tent has area s×(1/s)×(1/2)=1/2, and both sides have area 1/2+1/2=1, as required for a density). The slope is $1/s^2$.

[18] The constrained triangular has only one parameters that is its mean and spread.

[19] The mean weighted average elasticities were also statistically equivalent.

marginal disutility associated with public mode travel time and fares compared to females.

Adding in the error components for alternatives and nests of alternatives in model ML4 is way of allowing for additional sources of preference heterogeneity that is not accounted for by the random parameterisation and its associated decomposition[20]. Importantly however, whereas the random parameters can account for differences across individuals and alternatives, the error components for alternatives and nests of alternatives focus is on the heterogeneity profile of additional unobserved effects associated with each alternative. The standard deviation parameters associated with each alternative capture this. Although each alternative can in theory have its own unique standard deviation parameter, grouping of the modes into car, existing bus and heavy rail and the 'remaining modes' produced the best model fit (the latter being the new modes and existing busway)[21]. However only two of the standard deviation effects are statistically significant, with the car having the largest standard deviation parameter.

What this suggests is that there is a noticeable amount of preference heterogeneity associated with the car alternative that is not accounted for by the random parameters of car-specific attributes, compared to the public modes. The local environment in the North-West of Sydney has a high incidence of car usage compared to public transport use and so the heterogeneity across the population might be expected to be greater for the car segment. We have explored the possible reasons for the strong error components for alternatives and nests of alternatives for the car mode and its decomposition, to reveal sources of observed heterogeneity. We find that the age of the commuter has a statistically significant influence on preference heterogeneity. We could not find any significant effects for the public transport modes. All other effects being equal, the age effect suggests that as the age of the commuter increases, the standard deviation of the error components for alternatives and nests of alternatives decreases, leading to a reduction in preference heterogeneity from these unobserved effects.

To gain a richer understanding of the behavioural implications of increasingly more complex models as we progress from Ml1 to ML5, we present the matrix of direct elasticities (Table 6)[22]. The direct elasticities take into account every element of equations (17) - (22) that contribute to the percentage change in the attribute and the percentage change in the choice probability. We have selected main mode in-vehicle time to illustrate the differences in behavioural response across the five models[23]. The absolute values for the direct elasticities should be treated with caution since they are derived from an un-calibrated[24] stated choice model. Their purpose is simply to establish the behavioural response implications of alternative mixed logit specifications.

---

[20] Ben Akiva *et al* (2001) show that a variance term can be estimated for each of the alternatives due to the fact that in the kernel logit model '…a perfect trade-off does not exist because of the slight difference between the Gumbel and Normal distributions'. By contrast probit, probit kernel and extreme value logit require that one of the variances is constrained.

[21] Existing busway is in one sense a relatively new mode and it is interesting how its variance effect aligns closer to modes under consideration, all of which have their own infrastructure, with new busway being no more than a geographical extension of the existing busway.

[22] Model fit on its own is not the best indicator of the advantages of a more complex structure. Indeed, the improvements in fit may be quite small, but the findings in terms of elasticities can be quite different.

[23] Evidence for other attributes is similar and available on request.

[24] Elasticities are strictly meaningful, in a behavioural sense, after a model has been calibrated to reproduce the known population shares. Stated choice models whose alternative-specific constants have not been calibrated after estimation to reproduce population (in contrast to sample) shares are related to sample shares only.

Table 6: Direct Elasticities (probability weighted)
Main Mode In-vehicle Time

Direct Elasticities

Elasticity of in-vehicle time for	With respect to	ML1	ML2	ML3	ML4	ML5
New light rail	New light rail	-1.800	-1.778	-1.763	-1.781	-2.182
New heavy rail	New heavy rail	-1.759	-1.764	-1.764	-1.720	-1.909
New busway	New busway	-2.323	-2.311	-2.282	-1.366	-2.092
Existing Bus	Existing Bus	-1.829	-1.798	-1.771	-2.316	-3.079
Existing Busway	Existing Busway	-1.673	-1.676	-1.686	-2.379	-3.010
Existing heavy rail	Existing heavy rail	-1.486	-1.495	-1.500	-2.000	-2.639
Car	Car	-1.204	-1.214	-1.202	-1.129	-1.036

The mean estimates of the direct elasticities vary marginally as we move from the base model (ML1) through to accounting for heterogeneity in the mean and standard deviations of random parameters (ML3). However when we introduce the error components for alternatives and nests of alternatives (ML4), the elasticities change substantially for four public modes, with three increasing (existing bus, busway and heavy rail) and one decreasing (new busway). When we account for the commuter age effect (ML5), all public mode direct elasticities further increase. Although this empirical application is a single assessment of the extended mixed logit model, it does suggest that the introduction of the error components for alternatives and nests of alternatives and its decomposition has a (potentially) significant influence on the behavioural responsiveness of the model, in contrast to the refinements centred around the random parameters alone. The cross elasticity evidence (available on request) tells a similar story with some elasticities increasing and others decreasing as we include the error components for alternatives and nests of alternatives.

In addition to elasticities, we derived willingness to pay (WTP) distributions for each random parameter. We have selected three WTP estimates (Table 7) to illustrate the influence of model specification on the mean, standard deviation, range and incidence of negative WTP for values of travel time savings (VTTS) for public transport in-vehicle time, access wait time and egress time. The VTTS are based on the ratio of the parameter estimates associated with each individual observation as drawn from the distributions for the numerator attribute and the denominator attribute, the latter being the public transport main mode fare). The numerator varies in complexity in terms of deep parameterisation as we move from ML1 through to ML5.

Table 7: Value of Travel Time Savings ($/person hour)

		ML1	ML2	ML3	ML4	ML5
Public transport in-vehicle time	Mean	18.63	18.52	18.29	18.53	18.47
	Standard deviation	4.47	4.34	4.32	4.51	4.46
	Minimum	-3.13	-0.94	-5.77	-6.50	-6.31
	Maximum	29.82	29.73	31.86	34.42	34.86
	Percent Negative	0.2174	0.1630	0.3804	0.4891	0.4348
Access walk time	Mean	18.18	18.05	17.96	18.10	18.23
	Standard deviation	0.69	0.63	0.67	0.62	0.72
	Minimum	14.01	14.20	13.92	14.14	13.74
	Maximum	20.28	19.99	20.02	19.95	20.63
	Percent Negative	0	0	0	0	0
Egress time	Mean	6.61	6.61	6.65	6.32	6.46
	Standard deviation	2.26	2.30	2.29	1.97	2.06
	Minimum	-8.71	-9.04	-9.04	-7.42	-7.78
	Maximum	15.15	15.38	15.53	13.83	14.21
	Percent Negative	1.739	1.739	1.739	1.630	1.684

The evidence shows a very flat profile of values across the models, suggesting little if any behavioural enhancement when progressing from the relatively simple (‘parsimonious’) model I to the more complex model V. We might not expect any significant difference between models IV and V since the heterogeneity around the standard deviation of the error components for alternatives and nest of alternatives is car-specific. There are however statistically significant effects attributed to the enhancements across Models I-IV, which appear to ‘rearrange’ the contributing effects without changing the overall absolute values of travel time savings. What this suggests is that we might expect different valuation for specific segments of the sample (associated with the personal income and gender), which averages out to a similar overall estimate for the entire sample across all models. This is an important finding since it supports a position that there are systematic variations in tastes attribute to person-specific effects which, while not overly important when applying the findings to a sampled population as a whole, are important when evaluating the influence of policy of specific socio-economic segments.

5. Conclusion

This paper has added additional behavioural dimensionality to the mixed logit model for a cross section to account for an increasing number of sources of preference heterogeneity in choice making. We have set out the econometric extensions of the model and empirically illustrated how the increased dimensionality of the utility expressions adds to the revelation of sources of preference heterogeneity. The extended mixed logit model has rich behavioural properties, increasing the analyst’s capability to investigate sources of heterogeneity around both the mean and variance of random parameters as well as alternative-specific variance effects.

We find that preference heterogeneity can and should be captured via both the mean and variance domains underlying the utility expressions for sampled individuals, and that influences often rejected in one domain (e.g., the influence of gender of mean parameter estimates), resurfaces through their influence on the standard deviation parameter estimate.

The econometric feasibility now available to estimate such increasingly more complex models must be placed with the context of the gains in behavioural realism. Progressing from a base model up through additions of complexity provides a sensible way of revealing whether the increased generality of the model does make a difference in terms of the behavioural outputs of interest. In promoting this approach to good model estimation and interpretation practice, we recognise that the potential gains in behavioural realism may well be hampered by the quality of the data being used in estimation. Data that lacks the richness in variability in candidate influences as reflected in real markets or in experimental settings will almost always fail to offer an empirical setting within which the analyst can do justice to the increased realism available in the extended mixed logit model.

As a final reminder, analysts should always be circumspect in the number of random elements and parameterizations that can be estimated from data. One particular issue that needs to be recognized clearly is that there is only so much random structure that can be extracted from things that we do not observe.

References

Ben-Akiva, M., Bolduc, D. and Walker, J (2001) Specification, identification and estimation of the logit kernel (or continuous mixed logit) model, MIT Working paper, Department of Civil Engineering.

Bhat, C.R. (2001) Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model, *Transportation Research* Part B, 35, 677-693.

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

Gopinath, D., Schofield, M., Walker, J. and Ben-Akiva, M. (2004) Comparative analysis of discrete choice models with flexible substitution patterns, paper prepared for the 2005 *Annual meeting of the Transportation Research Board*, Washington D.C.

Gourieroux, C., and A. Monfort. *Simulation-Based Methods Econometric Methods*. Oxford: Oxford University Press, 1996.

Greene, W. H. (2003) Interpreting estimated parameters and measuring individual heterogeneity in random coefficient models, Department of Economics, Stern School of Business, New York University.

Greene, W. H., Hensher, D.A. and Rose, J. (2006) Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models, *Transportation Research B*, 40(1), 75-92.

Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: state of practice, *Transportation*, 30 (2), 133-176.

Hensher, D.A. and Rose, J. (in press) Development of Commuter and non-Commuter Mode Choice Models for the Assessment of New Public Transport Infrastructure Projects: A Case Study, *Transportation Research A*

Hensher, D.A., Rose, J. and Greene, W. H. (2005) *Applied Choice Analysis: A Primer,* Cambridge University Press, Cambridge.

Louviere, J.J., Hensher, D.A. and Swait, J.F. (2000) *Stated Choice Methods and Analysis*, Cambridge University Press, Cambridge.

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

Revelt, D. and K. Train (1998) Mixed Logit with repeated choices: households' choices of appliance efficiency level, *Review of Economics and Statistics*, 80, 1-11.

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

Train, K. (2001) A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit, working paper, Department of Economics, University of California, Berkeley, http://elsa.berkeley.edu/~train/compare.pdf

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

Train, K. and Revelt, D. (2000) Customer-specific taste parameters and mixed logit, working paper, Department of Economics, University of California, Berkeley, http://elsa.berkeley.edu/wp/train0999.pdf