I T L S

WORKING PAPER
ITLS-WP-05-03

Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models

By
William Greene,* David A Hensher &
John Rose

*William H Greene
Department of Economics
Stern School of Business
New York University
New York USA
wgreene@stern.nyu.edu

January 2005

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-05-03

TITLE: **Accounting for Heterogeneity in the Variance of Unobserved Effects in Mixed Logit Models**

ABSTRACT: The growing popularity of mixed logit to obtain estimates of willingness to pay (WTP) has focussed on the distribution of the random parameters and the possibility of estimating deep parameters to account for heterogeneity around the mean of the distribution. However the possibility exists to add further behavioural information associated with the variance of the random parameter distribution, through parameterisation of its heterogeneity (or heteroskedasticity). In this paper we extend the mixed logit model to account for this heterogeneity and illustrate the implications this has on the moments of the willingness to pay for travel time savings in the context of commuter choice of mode. The empirical study highlights the statistical and behavioural gains but warns of the potential downside of exposing the distribution of the parameterised numerator and/or denominator of the more complex WTP function to a sign change and extreme values over the range of the distribution.

KEY WORDS: ***Mixed logit, willingness to pay, stated choice methods, heterogeneity.***

AUTHORS: William H Green, David A Hensher and John Rose

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

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

DATE: January, 2005

1. Introduction

The mixed logit model has evolved to be a major innovation and a fundamentally useful tool in the analysis of discrete choice (see McFadden and Train, 2000; and Train, 2003). The hierarchical formulation and built in randomness of the model parameters allow the researcher to incorporate both observed and unobserved heterogeneity of individuals in the model. An increasing number of applications (e.g., Train, 2003; Hensher and Greene, 2003) have built this heterogeneity into the means of the distributions of the random parameters, and all but a very few studies, by Bhat[1], have treated the conditional variances of these distributions as constants. Our approach is more general than previous studies in that we allow the variance of any distribution to be a function of individual-specific characteristics that do not have to be related to the mean of the distribution.[2]

In this paper, we explore the further impact of accounting for variance heterogeneity in the distributions of individual specific taste weights.[3] We find in our application that accounting for individual specific variances of the distributions of random parameters brings a significant change in the estimated results. We also introduce the triangular distribution, initially promoted by Train (2001, 2003, pp 314-5), as an alternative to the normal and lognormal in mixed logit models.

The interest in exploring the value of promoting more 'complex' discrete choice models is strictly a matter of behavioural relevance. While it is useful to demonstrate the econometric feasibility of a method, the gains in behavioural relevance must be paramount. In the context of mixed logit models, the justification for the proposed extension is the search for a greater understanding of sources of preference heterogeneity within a sampled population. For a given attribute associated with an alternative, the mixed logit model obtains a distribution of marginal (dis)utilities as a description of the nature of heterogeneity of preferences for that attribute within the sample. This distribution is analytical and can take any specified form such as normal, lognormal, triangular, or Rayleigh. Most distributions are unrestricted over the positive and negative domain but can be constrained by additional assumptions. As an unconditional distribution, we have no *ex post* information beyond random allocation as to the precise location in the distribution, of each sampled individual's marginal (dis)utility. Given the information in the moments of the distribution (primarily the mean and standard deviation), further behavioural insights might be obtained by a knowledge of the relationship between the moments and individual-specific characteristics (e.g., personal income). Although many studies have established the extent to which the mean of the distribution of a random parameter might vary across market segments[4] (i.e., socioeconomic-level or range specific means), it is possible that we could segment along the full distribution without regard for, or in addition to, the mean. Intuitively, we might imagine a number of segments (even latent classes), defined

---

[1] We are aware of three studies by Bhat (1998, 2000) and Bhat and Zhao (2002).

[2] Bhat imposed a lognormal distribution on the random parameters, with the underlying normal having a *constant* variance and a mean that depends on demographics. Since the variance of a lognormal depends on the mean and variance of the underlying normal, the variance of the lognormal is automatically a function of demographics in these applications.

[3] Intuitively, this is analogous to accounting for heteroscedasticity across individuals in a random or fixed effects panel data style regression model. The computations are rather more complex in the discrete choice context, however.

[4] We acknowledge a referee for reminding us of this salient point.

by personal income levels, but within each income class we have further deep identification of preference heterogeneity. Alternatively, a single segment for the entire sample may be the relevant specification (i.e., no decomposition around the mean), where we identify systematic variation in the distribution of otherwise unobserved preference heterogeneity due to personal income levels. This is the behavioural justification of the proposed extension which can importantly accommodate conditioning of the variance of attribute parameters on individual-specific effects. These individual-specific effects can also condition the mean, but they should not be required to do so.

The paper is organised as follows. We first present the extension of the mixed logit model to allow for variance heterogeneity. We then describe the application to the choice of mode for commuting between existing and prospective alternatives. We set out the details of the design of a stated choice experiment, and present the results for the preferred models. The models are estimated on the sample data and so all parameters reflect the behavioural trade-offs made by the sampled individuals. A set of conclusions highlight the behavioural value of accounting for variance heterogeneity in the mixed logit model as well as the challenges that remain in delivering increased behavioural relevance while exposing the choice model to greater risk of an unacceptable range and sign domain of willingness to pay outputs.

2. A Heteroscedastic Mixed Logit Model

We assume that a sampled individual q (q=1,…,Q) faces a choice among J alternatives in each of T choice situations.[5] Individual q is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternative with the highest utility. The utility associated with each alternative j as evaluated by individual q in choice situation t, is represented in a discrete choice model by a utility expression of the general form in (1):

$$U_{jtq} = \mathrm{b}_q' \mathbf{x}_{jtq} + \varepsilon_{jtq}, \tag{1}$$

where $\mathbf{x}_{jtq}$ is the full vector of explanatory variables, including attributes of the alternatives, socioeconomic characteristics of the individual and descriptors of the decision context and choice task itself in choice situation t. The complexity of the choice task in stated choice experiments, as defined by number of choice situations, the number of alternatives, attribute ranges, data collection methods, etc., could also be included in the model to condition certain specific parameters. The components b_q and ε_{jtq} are not observed by the analyst and are treated as stochastic influences. Note that b_q is assumed to vary across individuals.

Individual heterogeneity is introduced into the utility function through b_q. We allow the 'individual-specific'[6] parameter vector to vary across individuals both randomly and

[5] In our implementation of the model, the number of alternatives could be different for different individuals, and could vary across choice situations as well. We assume the choice set size is fixed across both dimensions purely for expositional convenience.

[6] Strictly, as explained after equation (11), these individual-specific parameters are draws from a conditional distribution of the sub-sample of observations that have the same choice alternative.

systematically with observable variables, $\mathbf{z}_q$. If the random parameters are assumed to be uncorrelated, then the model may be written as

$$\begin{aligned} \mathrm{b}_q &= \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{S}^{1/2}\mathbf{v}_q \\ &= \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{h}_q \end{aligned} \tag{2}$$

or

$$\beta_{qk} = \beta_k + \mathrm{d}_k'\mathbf{z}_q + \eta_{qk},$$

where β_{qk} is the random coefficient for the k[th] attribute faced by individual q. The term b + D$\mathbf{z}_q$ accommodates heterogeneity in the mean of the distribution of the random parameters. The random vector h_q endows the random parameter with its stochastic properties. For convenience in isolating the model components, we define $\mathbf{v}_q$ to be a primitive vector of uncorrelated random variables with known variances and denote the matrix of known variances of the random draws as **W**. The actual scale factors which provide the unknown standard deviations of the random parameters are then arrayed on the diagonal of the diagonal matrix $\mathrm{S}^{1/2}$. Thus, for example, if (conditioned on $\mathbf{z}_q$) β_{qk} is normally distributed, then v_{qk} would be drawn from a standard normal distribution, $\mathbf{W}_{kk}$ would equal 1.0 as the known variance of v_{qk} would be 1.0, and σ_k would be the unknown scale factor. If, on the other hand, β_{qk} were assumed to be uniformly distributed, then $\mathbf{W}_{kk}$ would equal 1/12, and, once again, σ_k would be the unknown scale factor.[7] In order to allow the random parameters to be correlated, we now introduce the lower triangular matrix G, and extend the model to:

$$\mathrm{b}_q = \mathrm{b} + \mathrm{D}\mathbf{z}_q + \mathrm{G}\mathrm{S}^{1/2}\mathbf{v}_q.$$

Since the unknown scaling of the random components is already provided by the terms σ_k, the diagonal elements of G are normalized at one. The uncorrelated parameters model assumed in (2) then arises by assuming G = **I**. The conditional variance of b_q is now

$$\mathrm{Var}[\mathrm{b}_q \mid \mathbf{z}_q] = \mathrm{G}\mathrm{S}^{1/2}\mathbf{W}\mathrm{S}^{1/2}\mathrm{G}'.$$

For the individual variance terms, denoting the kth row of G as g_k, we have

$$\mathrm{Var}[\beta_{qk} \mid \mathbf{z}_q] = \mathrm{g}_k\,\mathrm{S}^{1/2}\mathbf{W}\mathrm{S}^{1/2}\,\mathrm{g}_k' = \Sigma_{i=1}^{k}\gamma_{ki}^2(\sigma_i w_i)^2 \text{ where } \gamma_{kk} = 1 \tag{3}$$

and

$$\mathrm{Cov}[\beta_k,\beta_m \mid \mathbf{z}_q] = \mathrm{g}_k\,\mathrm{S}^{1/2}\mathbf{W}\mathrm{S}^{1/2}\,\mathrm{g}_m' = \Sigma_{i=1}^{k}\gamma_{ki}\gamma_{mi}(\sigma_i w_i)^2 \text{ where } \gamma_{ki} = 0 \text{ if } i > k.$$

where g_k is the k[th] row of G.

[7] In each case, we impose a normalization on $\mathbf{W}_{kk}$ by assuming that the underlying variable v_k is standardized, in the normal case, to mean zero and variance one and in the uniform case, to the range [0,1] which gives a variance of 1/12.

The distribution of β_{qk} over individuals depends in general on underlying structural parameters (β_k,d_k,σ_k,g_k), the observed data, $\mathbf{z}_q$ and the unobserved vector of K random components in the set of utility functions $h_q = GS^{1/2}\mathbf{v}_q$. The last of these represents a stochastic element that enters the utility functions in addition to the J random elements in e_{tq}. Since b_q may contain alternative specific constants, covariation in η_{qk} induced by $G \neq \mathbf{I}$ will induce correlation of the random elements in the model across choices. Note that b_q, its component structural parameters $W = (b,D,G,S,\mathbf{W})$ and the characteristics of the person, $\mathbf{z}_q$ are choice situation invariant. They do not vary across choice situations or across choices.

Previous applications of the random parameter logit (RPL) model, and the specification above, have assumed homoscedasticity in the parameter distributions. We will introduce variance heterogeneity into the model as follows: Let $S_q^{1/2} = \text{Diag}[\sigma_{q1},\sigma_{q2},\ldots,\sigma_{qK}]$ where

$$\sigma_{qk} = \sigma_k \times \exp(q_k{¢}\mathbf{h}_q) \tag{4}$$

and $\mathbf{h}_q$ is a vector of M variables such as demographic characteristics that enters the variances (and possibly the means as well). This adds a $K{\times}M$ matrix of parameters, Q, to the model whose k^{th} row is the elements of q_k. With this explicit scaling, the full model for the variances in our model is now

$$\text{Var}[b_q|W, \mathbf{z}_q,\mathbf{h}_q] = F_q = GS_q^{1/2}\,\mathbf{W}S_q^{1/2}\,G¢.$$

The conditional variance of any specific parameter is now

$$\text{Var}[\beta_{qk}|W, \mathbf{z}_q,\mathbf{h}_q] = \Sigma_{i=1}^{k}\gamma_{ki}^2[\sigma_i w_i \exp(q_i'\mathbf{h}_q)]^2 \tag{5}$$

where w_k is the known scale factor $\mathbf{W}_{kk}^{1/2}$ and the covariance of any two parameters is

$$\text{Cov}[\beta_{qk},\beta_{ml}] = \Sigma_{i=1}^{k}\gamma_{ki}\gamma_{mi}[\sigma_i w_i \exp(q_i'\mathbf{h}_q)]^2\,. \tag{6}$$

The *mixed logit* class of models assumes a general distribution for β_{qk} and an IID extreme value type 1 distribution for ε_{jtq}. That is, β_{qk}?can take on different distributional forms such as normal, lognormal, uniform or triangular. For a given value of b_q, the *conditional* (on $\mathbf{z}_q$, $\mathbf{h}_q$ and $\mathbf{v}_q$) probability for choice j in choice situation t is multinomial logit, since the remaining random term, e_{tjq}, is IID extreme value:

$$\mathbf{P}_{jtq}(\text{choice } j\ |W,\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q) = \exp(b_q¢\mathbf{x}_{jtq}) / \Sigma_j\exp(b_q¢\mathbf{x}_{jtq}) \tag{7}$$

where the full set of attributes and characteristics is gathered in $\mathbf{X}_{tq} = [\mathbf{x}_{1tq},\mathbf{x}_{2tq},\ldots,\mathbf{x}_{Jtq}]$. For convenience, denote this as $P_{jtq}(b_q\ |\mathbf{B}_{qt},\mathbf{v}_q)$. Denote the marginal joint density of $[\beta_{q1},\beta_{q2},...,\beta_{qK}]$ by $f(b_q\ |W,?\mathbf{z}_q,\mathbf{h}_q)$ where the elements of W are the underlying parameters of the distribution of b_q, $(b,D,G,S,Q,\mathbf{W})$ and $(\mathbf{z}_q,\mathbf{h}_q)$ are observed data specific to the individual that enter the determination of b_q, such as socio-demographic characteristics. The density, itself, is induced by the transformation of the primitive random vector, $\mathbf{v}_q$ in $b_q = b + D\mathbf{z}_q + GS_q^{1/2}\mathbf{v}_q$.

We label as the *unconditional* choice probability the expected value of the logit probability over all the possible values of b_q, that is, integrated over these values, weighted by the density of b_q (it is still conditioned on the observable demographic information ($\mathbf{z}_q$,$\mathbf{h}_q$), but not on the unobservable $\mathbf{v}_q$). From (2), we see that this probability density is induced by the random component in the model for b_q, $\mathbf{v}_q$ (Hensher and Greene, 2003). Thus, the unconditional choice probability is

$$\mathbf{P}_{jtq}(\text{choice } j \mid \mathrm{W},\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q) = \int_{b_q} P_{jtq}(b_q \mid \mathbf{O},\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q) f(b_q \mid \mathbf{O},\mathbf{z}_q,\mathbf{h}_q) d b_q \tag{8}$$
$$= \int_{\mathbf{v}_q} P_{jtq}(b_q \mid \mathrm{W},\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q) f(\mathbf{v}_q \mid \mathbf{W}) d\mathbf{v}_q$$

where, once again, $b_q = b + D\mathbf{z}_q + GS_q^{1/2}\mathbf{v}_q$ (a Jacobian term would appear in the expression. The form given is intended to be generic). 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 b_q. The random variation in b_q is induced by the random vector $\mathbf{v}_q$; hence, that is the variable of integration in (8). The log likelihood function for estimation of the structural parameters is built up from these unconditional probabilities, aggregated for individual *q* over the *T* choice situations and the choices actually made:

$$\log L = \sum_{q=1}^{Q} \log \int_{\mathbf{v}_q} \prod_{t=1}^{T} P_{jtq}(b_q \mid \mathrm{W},\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_q) f(\mathbf{v}_q \mid \mathbf{W}) d\mathbf{v}_q . \tag{9}$$

The log likelihood function in (9) cannot be evaluated because the integrals will not have a closed form solution. But, it can be satisfactorily approximated by simulation. The simulated log likelihood function is given in equation (10).

$$\log L_S = \sum_{q=1}^{Q} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} P_{jtq}(b_{rq} \mid \mathrm{W},\mathbf{X}_{tq},\mathbf{z}_q,\mathbf{h}_q,\mathbf{v}_{rq}) . \tag{10}$$

where *R* is the number of draws in the simulation,

$$b_{rq} = b + D\mathbf{z}_q + GS_q^{1/2}\mathbf{v}_{rq} \tag{11}$$

and $\mathbf{v}_{rq}$ is the rth primitive random draw from the marginal population that generates $\mathbf{v}_q$. Maximum simulated likelihood estimates are obtained by maximizing $\log L_S$ with respect to all the unknown parameters in W. Details on estimation of the parameters of the mixed logit model by maximum simulated likelihood may be found in Train (2003).

One can construct estimates of 'individual-specific preferences' by deriving the conditional distribution based (within-sample) on known choices (i.e., prior knowledge), as originally shown by Revelt and Train (2000) (see also Train, 2003 chapter 11). These conditional parameter estimates are strictly 'same-choice-specific' parameters, or the mean of the parameters of the subpopulation of individuals who, when faced with the same choice situation would have made the same choices. This is an important distinction[8] since we are not able to establish for each individual, their

[8] Discussion with Ken Train is appreciated.

unique set of estimates, but rather we are able to identify a mean (and standard deviation) estimate for the sub-population who make the same choice. For convenience, let $\mathbf{Y}_q$ denote the observed information on choices by individual q, and let $\mathbf{X}_q$ denote all elements of $\mathbf{x}_{jtq}$ for all j and t. Using Bayes Rule, we find the conditional density for the random parameters,

$$f(\mathrm{b}_q|\mathrm{W}, \mathbf{Y}_q,\mathbf{X}_q,\mathbf{z}_q,\mathbf{h}_q) = \frac{f(\mathbf{Y}_q \,|\, \mathrm{b}_q, \mathrm{W}, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q) P(\mathrm{b}_q \,|\, \mathrm{W}, \mathbf{z}_q, \mathbf{h}_q)}{f(\mathbf{Y}_q \,|\, \mathrm{W}, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q)}. \tag{12}$$

The left hand side gives the conditional density of the random parameter vector given the underlying parameters and all of the data on individual q. In the numerator of the right hand side, the first term gives the choice probability in the conditional likelihood – this is in (8). The second term gives the marginal probability density for the random b_q implied by (2) with the assumed distribution of $\mathbf{v}_q$. The denominator is the unconditional choice probability for the individual – this is given by (8). Note that the denominator in (12) is the integral of the numerator. This result can be used to estimate the 'common-choice-specific' parameters, utilities, and willingness to pay values or choice probabilities as a function of the underlying parameters of the distribution of the random parameters. Estimation of the individual specific value of b_q is done by computing an estimate of the mean of this conditional distribution. Note that this conditional mean is a direct analog to its counterpart in the Bayesian framework, the mean of the posterior distribution, or the posterior mean. More generally, for a particular function of b_q, $g(\mathrm{b}_q)$, such as b_q itself, the conditional mean function is

$$\mathbf{E}[\mathrm{g}(\mathrm{b}_q) \,|\, \mathrm{W}, \mathbf{Y}_q,\mathbf{X}_q,\mathbf{z}_q,\mathbf{h}_q] = \int_{\mathrm{b}_q} \frac{g(\mathrm{b}_q) f(\mathbf{Y}_q \,|\, \mathrm{b}_q, \mathrm{W}, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q,) P(\mathrm{b}_q \,|\, \mathrm{W}, \mathbf{z}_q, \mathbf{h}_q)}{f(\mathbf{Y}_q \,|\, \mathrm{W}, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q)} d\mathrm{b}_q \tag{13}$$

The various integrals mentioned above generally cannot be calculated exactly because the integrals will not have a closed form solution. But, like the likelihood function, they can be accurately approximated by simulation. For given values of the parameters, W, and the observed data, $(\mathbf{Y}_q,\mathbf{X}_q,\mathbf{z}_q,\mathbf{h}_q)$ a value of b_q is drawn from its distribution based on (2). For example, using this draw, the logit formula (13) for $\mathrm{L}_{jtq}(\mathrm{b}_q)$ is calculated. This process is repeated for many draws, and the mean of the resulting $L_{jtq}(\mathrm{b}_q)$'s is taken as the approximate choice probability giving the simulated probability,

$$\hat{P}(\mathbf{Y}_q \,|\, \mathbf{X}_q, \mathbf{z}_q, \mathrm{W}) = \frac{1}{R} \sum_{r=1}^{R} L_{jq}(\mathrm{b}_{qr} \,|\, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q, \mathrm{W}, \mathrm{h}_{qr}) \tag{14}$$

R is the number of replications (i.e., draws of b_{qr}), b_{qr} is the r^{th} draw, and the right hand side is the simulated probability that an individual chooses alternative j[9]. Then, for example, the simulation estimator of the conditional mean for β_q is

[9] By construction, this is a consistent estimator of P_j for any R; its variance decreases as R increases. It is strictly positive for any R, so that $\ln(SP_j)$ is always defined in a log-likelihood function. It is smooth (i.e., twice differentiable) in parameters and variables, which helps in the numerical search for the maximum of the likelihood function. The simulated probabilities sum to one over alternatives. Train (1998) provides further commentary on this.

$$\hat{E}_S[\beta_q \mid Individual\ q] = \frac{(1/R)\sum_{r=1}^{R}\beta_{q,r}L(\beta_{q,r} \mid data_q)}{(1/R)\sum_{r=1}^{R}L(\beta_{q,r} \mid data_q)}. \qquad (15)$$

3. Empirical Application

In 2003, the Institute of Transport Studies (ITS) (University of Sydney), on behalf of the New South Wales State government, undertook a patronage demand study as part of an evaluation of possible investment options in public transport infrastructure in the north-west sector of metropolitan Sydney[10]. 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 and non-commuting trip purposes. 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 (Figure 1).

To capture information on the preferences of residents, a stated choice (SC) experiment was generated and administered using computer aided program 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 (M2) 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.

[10] The north-west sector is approximately 25 kilometres from the Sydney central business district (CBD). It is the fastest growing sector of Sydney in terms of residential population and traffic build up. It is also one of the wealthiest areas with high car ownership and usage and a very poor public transport service with the exception of a busway system along the M2 tollroad into the CBD of Sydney.

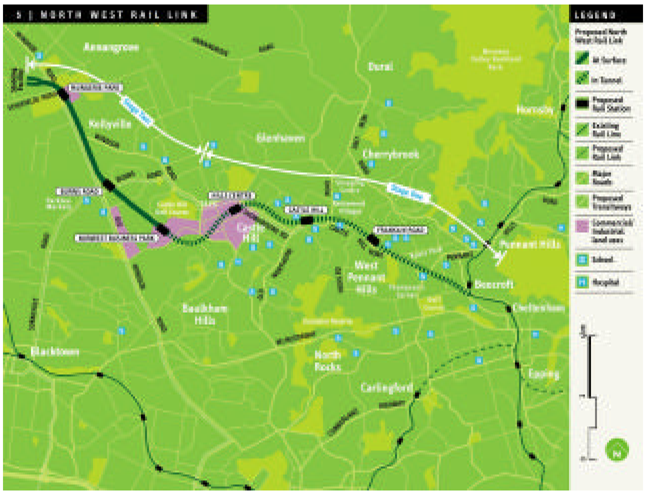

Figure 1: The North-West Sector and Proposed new rail link

The master experimental design for the mode SC study required a total of 47 attributes (46 in four levels and one in six levels for the blocks) and had 60 runs; that is, there are six blocks of ten choice sets each. The design was constructed using a procedure that simultaneously optimised the minimisation of the D-error of the design as well as the correlations. The final design had correlations no greater than ± 0.06. The design allowed for the estimation of all main mode and access mode alternative specific main effects. Within each block, the order of the choice sets has been randomised to control for order effect biases. The experiment consisted of different task configurations designed to reflect the alternatives realistically available to a respondent, given the reference trip circumstance reported by the respondent earlier in the CAPI interview. The configurations consisted of: (i) with/without car, (ii) inter/intra regional trips, (iii) new light rail versus new heavy rail, new light rail versus new busway and new heavy rail versus new busway. These configurations were included to provide more realism in the scenarios shown to individual respondents. To maintain efficiency and minimise correlations within the data set, a maximum number of designs have to be completed within each configuration. Using the CAPI program, if the first respondent had a car available for an intra-regional trip with new light rail and heavy rail alternatives present, she was assigned to block one for that configuration. If the second respondent is in the exact same configuration, she was assigned to the second block; otherwise, she was assigned to block one of the appropriate design configuration. Once a configuration has all blocks completed, the process started at block one again.

The trip attributes associated with each mode are summarised in Table 1. These were identified from extensive reviews of the literature and through input from a technical advisory committee chaired by the NSW Ministry of Transport.

Table 1: Trip Attributes in the 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	

For existing modes, the attribute levels where pivoted off the attribute levels captured from respondents for a reference trip. Respondents were asked to complete information on the reference trip not only for the mode used for the reference trip, but also for the other modes they had available for that trip. Whilst asking respondents to provide information for non-chosen alternatives may potentially provide inaccurate attribute levels, choices made by individuals are based on their perceptions of the attribute levels of the available alternatives and not the reality of the attribute levels of those same alternatives. As such, asking respondents what they thought the levels were for the non-chosen alternatives was preferable than imposing those levels on the experiment based on some heuristic, given knowledge of the attribute levels for the actual chosen alternative.

The design attributes used in the SC experiment, each had four levels. These were chosen as the following variations around the 'reference' trip base levels: -25 percent, 0 percent, +25 percent, +50 percent. The times and costs associated with currently non-existent public transport modes were 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 the map (Figure 1 above) and choose a particular station[11], which is used in the software to derive the access and linehaul travel times and fares. Example SC screens are shown in Figures 2 (inter-regional trip with car) and 3 (intra-regional trip with car).

[11] The map was shown so that respondents will be able to locate the proposed stations in relation to their residential address. This is essential so that we can gather the most reliable measure of access times to the stations/bus stops. The attribute levels of the access times (the *getting to main mode* attributes) will be 'switched or pivoted off' these reported levels. Without these, the attribute levels for the access modes cannot be determined and that part of the model cannot be estimated.

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

	Dollars $	Busway (minutes)	Heavy rail (minutes)	Light Rail (minutes)
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				

A total of 223 commuters completed the survey. The average survey response time was 34 minutes, including preliminary screening questions. Respondents were not offered any incentive to participate.

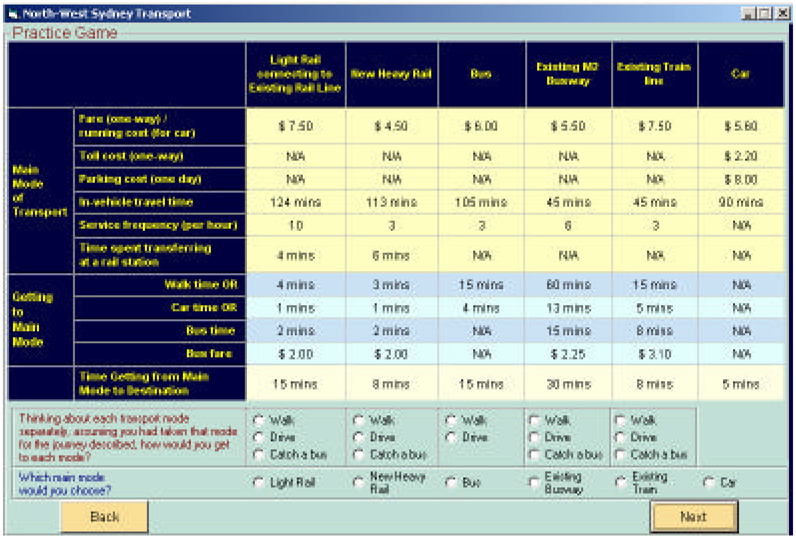

Figure 2: Example inter-regional stated choice screen

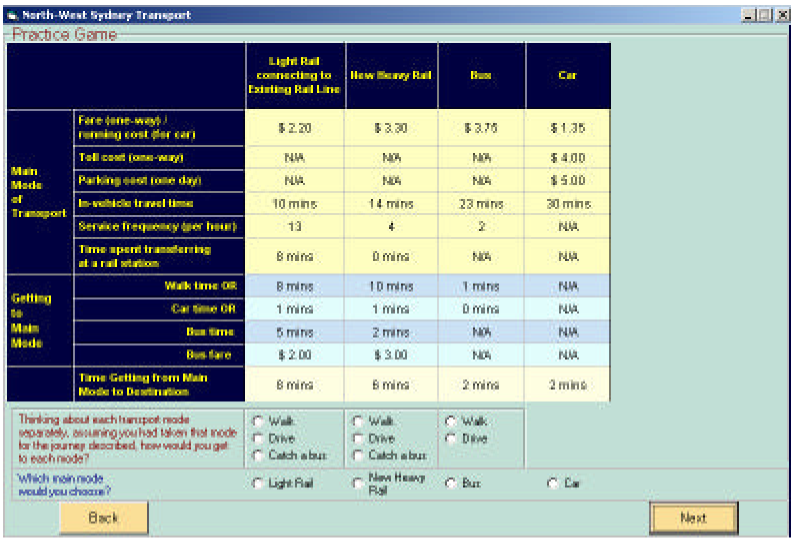

Figure 3: Example intra-regional stated choice screen

The empirical data on commuter trips is drawn from a larger study on all trip purposes.[12] Table 3 shows the descriptive statistics for the work segment. The mean age is 43.1 years with an average annual gross personal income of $64,100. The proportion of males to females is equally split. Of the 223 commuters interviewed, 199 (or 89.42 percent) had access to a car for the surveyed trip.

Table 3: Descriptive statistics for Work segment

	N	Mean	Std. Deviation	Minimum	Maximum
Age	223	43.1	12.5	24	70
Hours worked per week	223	37.6	14.6	0	70
Annual Personal Income ($000's)	223	64.1	41.8	0	140
Household size	223	3.78	2.30	1	8
No. of children in household	223	1.05	1.09	0	4
Gender (male =1)	223	50.4	-	0	1

4. Results

Table 4 presents the model results for the experiment. Four models are summarised in the table. The first model is a multinomial logit (MNL) model whilst the remaining three are mixed logit (ML) models. Mixed logit models 1 and 2 do not allow for variance heterogeneity whilst ML model 3 does. Within all three ML models, the cost, travel time and egress time parameters for both the car and public transport modes are specified as random parameters drawn using 500 Halton draws. With the exception of the waiting and access time non-random parameters, the public transport parameters were specified as generic across the alternatives.

All random parameter estimates for ML model 1 were drawn from constrained triangular distributions[13] as were the random parameter estimates associated with the car alternative in ML models 2 and 3. Hensher and Greene (2003) have shown that for the triangular distribution, when the mean parameter is constrained to equal its spread (i.e., $\beta_{jk} = \beta_k + |\beta_k|\ T_j$, where T_j is a triangular distribution ranging between -1 and +1), the density of the distribution rises linearly to the mean from zero before declining to zero again at twice the mean. Therefore, the distribution lies between zero and some estimated value (i.e., the β_{jk}). As such, all parameter estimates are constrained to be of the same sign. Empirically the distribution will be symmetrical about the mean which not only allows for ease of interpretation, but also avoids the problem of long tails often associated with drawing from a log-normal distribution as in Bhat (1998, 2000).

[12] See Greene *et al*. (2005) or Hensher and Rose (2004) for details of results based on the non-commuter data.

[13] For example, the usual specification in terms of a normal distribution is to define $\beta_i = \beta + \sigma v_i$ where v_i is the random variable. The constrained specification would be $\beta_i = \beta + \beta v_i$ when the standard deviation equals the mean or $\beta_i = \beta + h\beta v_i$ when h is the coefficient of variation taking any positive value. We would generally expect *h* to lie in the 0-1 range since a standard deviation greater than the mean estimate *typically* results in behaviourally unacceptable parameter estimates.

Table 4: Model Results

		MNL		ML (1) – constrained T		ML (2) – unconstrained T		ML (3)- unconstrained T	
Attribute	**Alternative**	**Coeff.**	**t-ratio**	**Coeff.**	**t-ratio**	**Coeff.**	**t-ratio**	**Coeff.**	**t-ratio**
		Random parameter means							
Fare	Main Mode Public Transport	-0.175	-9.20	-0.233	-10.66	-0.342	-9.20	-0.358	-10.20
In-vehicle Travel Time	Main Mode Public Transport	-0.053	-13.77	-0.065	-17.98	-0.073	-13.77	-0.075	-14.93
Egress Time	Main Mode Public Transport	-0.020	-3.93	-0.025	-3.87	-0.029	-3.93	-0.027	-3.84
Cost (Running and Toll)	Main Mode Car	-0.109	-3.02	-0.168	-3.19	-0.218	-3.02	-0.204	-3.03
In-vehicle Travel Time	Main Mode Car	-0.032	-5.22	-0.063	-5.50	-0.084	-5.22	-0.084	-5.72
Egress Time	Main Mode Car	-0.054	-3.18	-0.081	-3.17	-0.105	-3.18	-0.106	-2.58
		Random parameter spread							
Fare	Main Mode Public Transport	-	-	0.233	10.66	0.763	6.84	0.653	5.67
In-vehicle Travel Time	Main Mode Public Transport	-	-	0.065	17.98	0.088	6.09	0.074	5.64
Egress Time	Main Mode Public Transport	-	-	0.025	3.87	0.029	3.93	0.027	3.84
Cost (Running and Toll)	Main Mode Car	-	-	0.168	3.19	0.218	3.02	0.204	3.03
In-vehicle Travel Time	Main Mode Car	-	-	0.063	5.50	0.084	5.22	0.084	5.72
Egress Time	Main Mode Car	-	-	0.081	3.17	0.105	3.18	0.106	2.58
		Non Random parameters							
Constant	New Light rail	2.925	8.41	2.528	5.46	2.906	5.20	3.306	6.11
Constant	Heavy Rail Modes	2.266	6.80	1.791	3.95	2.088	3.82	2.514	4.76
Constant	New Busway	1.825	4.83	1.445	2.94	1.765	3.03	2.109	3.64
Constant	Bus	2.198	6.63	1.747	3.86	1.953	3.59	2.292	4.31
Constant	Busway	1.924	5.63	1.475	3.19	1.756	3.15	2.082	3.82
Access and wait time	All Rail Modes	-0.060	-9.70	-0.067	-10.01	-0.076	-9.95	-0.079	-10.93
Wait time	Bus modes	-0.101	-3.62	-0.111	-3.71	-0.124	-3.80	-0.115	-3.34
Access time	Bus modes	-0.055	-5.72	-0.064	-6.06	-0.070	-6.00	-0.070	-5.92
Access bus fare	All Public Transport	-0.075	-2.12	-0.081	-2.15	-0.099	-2.42	-0.109	-2.76
Parking cost	Car	-0.017	-1.94	-0.031	-2.46	-0.041	-2.37	-0.038	-2.16
		Heteroscedasticity in Random Parameters							

Fare | Income	Main Mode Public Transport	-	-	-	-	-	-	0.910	2.95
Fare | Household Size	Main Mode Public Transport	-	-	-	-	-	-	0.046	1.83
In-vehicle Travel Time | Household Size	Main Mode Public Transport	-	-	-	-	-	-	0.050	2.18
Egress Time | Household Size	Main Mode Public Transport	-	-	-	-	-	-	0.141	3.77
				Model Fits					
LL(0)		-2769.127		-2769.127		-2769.127		-2769.127	
LL(B)		-1929.411		-1912.880		-1899.230		-1888.288	
Chi-square		1679.432		1712.494		1739.794		1761.678	
Adj. pseudo R^2		0.298		0.308		0.313		0.316	
Observations		1840		1840		1840		1840	

In contrast to ML(1), in ML(2) and ML(3), the random parameter estimates associated with the generic fare, travel time and egress time random parameters for the public transport modes were drawn from an unconstrained triangular distribution. Although the unconstrained triangular distribution allows for parameter estimates either side of zero, the introduction of interaction terms (namely personal income and household size) designed to uncover variance heterogeneity[14] can no longer guarantee that the parameter distribution will be limited to one side of zero[15], despite constraints imposed on the underlying distribution. To explain why, let the marginal utility associated with attribute *k* be given as

$$\hat{b}_k = [\bar{b}_k v_k + s_k \times \exp(q_k h_q)] \times h \tag{17}$$

where $\bar{b}_k$ is the mean of the random parameter distribution, v_k is the random variable, s_k is the standard deviation or spread of the random parameter distribution, q_k is the variance heterogeneity parameter uncovered through the interaction between h_q and the random parameter distribution and h represents a draw from a known empirical distribution (e.g., normal, lognormal, triangular, uniform).

A significant, non zero value for q_k may allow for parameter estimates of either sign, given that the parameter estimate is no longer solely dependent on the draw (conditional or otherwise) from h, but also upon the additional information imparted through $s_k \times \exp(q_k h_q)$. As such, even if all draws from h are constrained to one side of zero, the addition of $s_k \times \exp(q_k h_q)$ within equation (17) allows for the possibility that some random parameter estimates will not be of the desired sign. Although we do not show it here, this same issue exists when decomposing the mean of random parameter distributions to uncover sources of heterogeneity. Whilst the literature has identified the need to employ distributions that dictate the sign of random parameters, research on the impact of accommodating heterogeneity around the mean of random parameter distributions and variance heterogeneity appears to be absent (see Hensher, 2004). Given the above, there does not appear to be any *guaranteed* advantage in using a distribution that constrains the random parameter estimates to one side of zero when deeper parameter estimates are included within the model.

For the models reported in Table 5, all parameter estimates associated with the design attributes are statistically significant and of the expected sign. A direct test to determine whether ML model 1 statistically represents the data better than the MNL model is not possible given the use of constrained triangular distributions in the estimation of the random parameters in ML model 1. This is because the spread parameters of the random parameter distributions are constrained to equal the means of the distributions, and hence no additional parameters are estimated in the ML model. As such, any statistical test will have zero degrees of freedom. The freely estimated spread parameters in ML model 2 and

[14] In the current empirical study we found no statistically significant individual-specific characteristics influencing the mean of any random parameter.
[15] Hensher (2004) investigated the imposition of a fully exponential marginal (dis)utility function for any analytical distribution that captures all of the inputs as shown in equation (17). Estimation leading to convergence proved to be problematic. After extensive investigation with four data sets and five analytical distributions, we were only able to establish a final model for the Rayleigh distribution on one data set. The findings are however problematic for other reasons such as very low mean values.

the additional interaction terms introduced to uncover variance heterogeneity in ML model 3, however, allow for tests of statistical significance between these models and the MNL model and ML model 1, as well as between each other. The log-likelihood ratio tests are summarised in Table 5. ML models 2 and 3 are superior statistically to both the MNL model and ML model 1, whilst ML model 3 provides a better model fit to that provided from ML model 2.

Table 5: Log-likelihood ratio tests

Test	c^2	Degrees of Freedom
ML (2)-MNL	60.36	6
ML (2)-ML (1)	27.3	6
ML (3)-MNL	82.25	10
ML (3)-ML (1)	49.18	10
ML (3)-ML (2)	21.88	4

Three of the four variance heterogeneity parameters are statistically significant at the 95 percent level, whilst the remaining variance heterogeneity interaction parameter is significant at the 90 percent level, demonstrating the presence of variance heterogeneity within the model. Comparison of the means of the random parameter estimates for the public transport fare, travel and egress time attributes across ML models 1 and 2 suggest that accounting for variance heterogeneity produces differences in the means of the parameter estimate distributions, however, some of these discrepancies may be attributable to differences in the distributions employed in the estimation of the two models (i.e., constrained versus unconstrained triangular distributions). The differences in the marginal (dis)utilities for the public transport random parameters is demonstrated when the respective marginal utilities are written out in full, as below.

ML Model 1 (where T_c is a draw from a constrained triangular distribution)

MU(Fare) = {-0.233 + 0.233×T_c}

MU(Travel Time) = {-0.065 + 0.065×T_c}

MU(Egress Time) = {-0.025 + 0.025×T_c}

ML Model 2 (where T is a draw from an unconstrained triangular distribution)

MU(Fare) = {-0.342 + 0.763×T}

MU(Travel Time) = {-0.073 + 0.088×T}

MU(Egress Time) = {-0.029 + 0.029×T}

ML Model 3 (where T is a draw from an unconstrained triangular distribution)

MU(Fare) = {-0.358 + 0.653×[(exp(0.910×personal income) + exp(0.046×household size)]×T}

MU(Travel Time) = {-0.075 + 0.074 ×(exp(0.050×household size)×T}

MU(Egress Time) = {-0.027 + 0.027×(exp(0.141×household size)×T}

Behavioural values of travel time savings (VTTS) for the MNL model and ML models are summarised in Table 6. The means of the VTTS distributions derived from all four models are intuitively plausible in absolute values[16], however, the range of VTTS obtained from ML models 2 and 3 show both negative and large positive VTTS estimates. Removing the negative VTTS and positive VTTS which are greater than three times the standard deviation from the mean of the VTSS distributions produces much more sensible ranges for the VTTS values for both ML models 2 and 3, (as shown in Table 6). Table 7 summarises the number of cases removed from the distributions which represent less than 3% of the total number of cases (i.e., individuals × choice sets). Note that it was not necessary to remove any observations for the car alternatives given that the random parameter estimates were derived using constrained triangular distributions, and hence, all VTTS estimates for this alternative were of the correct sign. The findings from a policy perspective suggest that accounting for heterogeneity in the variance of the unobserved effects tends to significantly reduce the mean VTTS; with the reduction less significant when the negative VTTS and extreme positive values are removed.

Although our main focus is on promoting the value of allowing for sources of observation-specific influence on the variance of the unobserved effects in choice models, as an additional way of linking unobserved heterogeneity to specific characteristics of sampled individuals, the supplementary contribution illustrates the challenge that remains in establishing behavioural distributions of willingness to pay outputs that are meaningful across the entire distribution. Sign changes and extreme values[17] remain a challenge to estimation of random parameter logit models that pursue a deeper understanding of the dimensionality of heterogeneity that underlie these distributions. The VTTS evidence reinforces this.

[16] It is noteworthy how similar the mean estimates are across MNL and ML2 where the comparison is straightforward (since there are no influences on unobserved variance). We often see papers only reporting the mean VTTS for mixed logit without acknowledging the range and we suspect sizeable negative and positive values at the extremes.
[17] The extreme values problem is especially noticeable for the lognormal distribution.

Table 6: Behavioural Values of Travel Time Savings ($/person hour) (t-ratios) for full sample

VTTS	Alternative	Mean	Std. Dev.	Range
		MNL		
In-vehicle Travel Time	Main Mode Public Transport	$17.98	-	-
Egress Time	Main Mode Public Transport	$7.00	-	-
Travel Time	Main Mode Car	$17.59	-	-
Egress Time	Main Mode Car	$29.90	-	-
		A: Models Preserving the Total Data Set		
		ML (1) Constrained Triangular		
Travel Time	Main Mode Public Transport	$16.62 (291.29)	$2.45	$5.92-28.79
Egress Time	Main Mode Public Transport	$6.44 (448.15)	$0.62	$4.90-$12.27
Travel Time	Main Mode Car	$22.57 (499.69)	$1.94	$9.98-26.75
Egress Time	Main Mode Car	$29.04 (1100.77)	$1.13	$26.57-$38.21
		ML (2) Unconstrained Triangular		
Travel Time	Main Mode Public Transport	$17.86 (3.87)	$198.14	-$1776.91-$8127.61
Egress Time	Main Mode Public Transport	$7.74 (2.75)	$120.96	-$1044.23-$5037.28
Travel Time	Main Mode Car	$23.03 (504.77)	$1.96	$10.61-$27.24
Egress Time	Main Mode Car	$28.78 (1078.95)	$1.14	$25.55-$38.99
		ML (3) Unconstrained Triangular		
Travel Time	Main Mode Public Transport	$12.05 (3.84)	$134.66	-$4116.19-$1816.29
Egress Time	Main Mode Public Transport	$3.25 (1.89)	$73.60	-$2747.81-$622.36
Travel Time	Main Mode Car	$24.61 (494.86)	$2.13	$11.29-$29.32
Egress Time	Main Mode Car	$31.12 (1166.07)	$1.14	$27.24-$41.74
		B: Models that Eliminate Negative and Extreme Positive Values		
		ML (2)[1] Unconstrained Triangular		
Travel Time	Main Mode Public Transport	$15.31 (58.09)	$11.14	$2.39-$105.18
Egress Time	Main Mode Public Transport	$6.10 (55.16)	$4.68	$2.35-$46.75
Travel Time	Main Mode Car	$23.03 (504.77)	$1.96	$10.61-$27.24
Egress Time	Main Mode Car	$28.78 (1078.95)	$1.14	$25.55-$38.99
		ML (3)[1] Unconstrained Triangular		
Travel Time	Main Mode Public Transport	$15.08 (57.15)	$11.15	$0.95-$102.60
Egress Time	Main Mode Public Transport	$5.64 (48.10)	$4.95	$1.26-$93.73
Travel Time	Main Mode Car	$24.61 (494.86)	$2.13	$11.29-$29.32
Egress Time	Main Mode Car	$31.12 (1166.07)	$1.14	$27.24-$41.74

[1] *VTTS obtained after removing negative values and outliers*

Table 7: Summary of individual-specific VTTS removed

			Number of negative VTTS estimates		
Model	**VTTS**	**Alternative**	**# removed**	**Total #**	**(% of total)**
ML (2)	In-vehicle Travel Time	Main Mode Public Transport	37	1840	2.01%
ML (2)	Egress Time	Main Mode Public Transport	37	1840	2.01%
ML (3)	In-vehicle Travel Time	Main Mode Public Transport	39	1840	2.12%
ML (3)	Egress Time	Main Mode Public Transport	39	1840	2.12%
			Number of positive VTTS estimates		
Model	**VTTS**	**Alternative**	**# removed**	**Total #**	**(% of total)**
ML (2)	In-vehicle Travel Time	Main Mode Public Transport	15	1840	0.82%
ML (2)	Egress Time	Main Mode Public Transport	13	1840	0.71%
ML (3)	In-vehicle Travel Time	Main Mode Public Transport	17	1840	0.92%
ML (3)	Egress Time	Main Mode Public Transport	15	1840	0.82%

5. Conclusion

The underlying behavioural assumptions used in the estimation of models of discrete choice are being increasingly relaxed, allowing not only for the possibility of identifying sources of heterogeneity associated with the mean of population parameters, but also the variances associated with random parameter distributions. It is the issue of variance heterogeneity that is the focus of this paper. We introduce the Heteroscedastic Mixed Logit model which allows for a decomposition of variance heterogeneity in the random parameter estimates via an interaction with individual specific characteristics contained within the data. We show that accounting for variance heterogeneity within the random parameter distributions, conditioned on person specific variables, produces better model fits as well as behaviourally sensible outputs in terms of the *means* of VTTS distributions. Unfortunately, the procedure outlined here is likely to produce unacceptable ranges in the behavioural outputs, including negative VTTS estimates. This remains a significant challenge within the literature; how best to decompose the various sources of heterogeneity that may exist within a data set whilst maintaining behaviourally sensible outputs in terms of sign and range. Hensher (2004) investigated this issue with limited empirical success.

Acknowledgements. We thank the NSW Ministry for permission to re-analyse the data from the NW Transport study and to include the findings in this paper. Comments on an earlier version by two referees are greatly appreciated.

References

Bhat (1998) Accommodating Variations in Responsiveness to Level-of-Service Variables in Travel Mode Choice Modeling, *Transportation Research Part A*, 32 (7), pp. 495-507.

Bhat (2000) Incorporating Observed and Unobserved Heterogeneity in Urban Work Mode Choice Modeling, *Transportation Science*, 34 (2), pp. 228-238.

Bhat & Zhao (2002) The Spatial Analysis of Activity Stop Generation, *Transportation Research Part B*, 36 (6), pp. 557-575.

Greene, W, Hensher, DA & Rose, J (2005) Using Classical Simulation Based Estimators to Estimate Individual Willingness to Pay Values: A Mixed Logit Case Study of Commuters, invited paper for *Applications of Simulation Methods in Environmental and Resource Economics*, Editors Anna Alberini and Riccardo Scarpa, Kluwer Academic Publisher, New York

Hensher, DA (2004) Imposing a globally signed constraint on a willingness to pay function, Institute of Transport Studies, The University of Sydney.

Hensher, DA & Greene, WH (2003) Mixed logit models: state of practice, *Transportation*, 30 (2), 133-176.

Hensher, DA & Rose, J (2003) *The North-West Transport Study Patronage Survey: Stated Choice Model Estimation for Work and Non-Work Travel*, Report prepared for the North-West Sydney Transportation Study, Institute of Transport Studies, The University of Sydney, July.

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

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

Train, K (1998) Unobserved taste variation in recreation demand models. *Land Economics*, 74(2), 230-239.

Train (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.