I T L S

WORKING PAPER

ITLS-WP-07-16

Combining RP and SP data: Biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects

By

David A Hensher, John Rose & William H Greene*

***Stern School of Business, New York University**

September 2007

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-07-16

TITLE: **Combining RP and SP data: Biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects**

ABSTRACT: It has become popular practice that joint estimation of choice models that use stated preference (SP) and revealed preference (RP) data requires a way of adjusting for scale to ensure that parameter estimates across data sets are not confounded by differences in scale. The nested logit 'trick' presented in Hensher and Bradley (1993) continues to be widely used, especially by practitioners, to accommodate scale differences. This modelling strategy has always assumed that the observations are independent, a condition of all GEV models, which is not strictly valid within a stated preference experiment with repeated choice sets and between each SP observation and the single RP data point. This paper promotes the replacement of the NL 'trick' method with an error components model that can accommodate correlated observations as well as reveal the relevant scale parameter for subsets of alternatives. Such a model can also incorporate "state" or reference dependence between data types and preference heterogeneity on observed attributes. An example illustrates the difference in empirical evidence.

KEY WORDS: ***Combined SP and RP data, nested logit trick, scale parameters, error component mixed logit, reference dependence***

AUTHORS: David A Hensher, John M Rose & William H Greene

CONTACT: Institute of Transport and Logistics Studies **(C37)**

The Australian Key Centre in Transport Management

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: September 2007

1. Introduction

It is common practice in estimation of many choice models which combine multiple data sources (e.g., RP and SP data sets) to use a nested logit (NL) structure as a 'trick or mechanism to reveal differences in scale between data sources' (Bradley and Daly 1992, 1997, Hensher and Bradley 1993). It is a trick in the sense that the underlying conditions to comply with utility maximisation such as the 0-1 bound on the inclusive value variable linking two levels in a nest (McFadden 1981), while applicable between alternatives within SP and within RP choice sets, are not relevant *between* data sets – the scale differences between data sets (typically normalising to unity on one data source) is the only agenda.

In the majority of NL applications, the predictability of the set of NL structures studied is driven by the revelation of differences in SP-RP scale parameters and/or the partitioning of alternatives within a given data set in what is best described as 'commonsense' or intuitive partitions; for example, the marginal choice between car and public transport and then the choice between bus and train, conditional on choosing public transport. Hensher (1999) generalised the role of scale parameters through the use of the HEV search engine to allow for differences in scale, not only between data sets but between alternatives within and between data sets.

The NL model is a member of GEV models (McFadden 1981) which cannot accommodate a number of specification requirements of data that has repeated observations from the same respondent. This occurs with SP choice sets which exhibit potential correlation due to repeated observations or panel data. We need mixing of some kind for this, using the GEV model as a kernel.

In addition to potential observation correlation, joint RP-SP estimation induces a 'state' or reference dependence effect, defined as the influence of the actual (revealed) choice on the stated choices of the individual. Reference dependence can manifest itself as a positive or negative effect of the choice of an alternative on the utility associated with that alternative in the stated responses (Bhat and Castelar 2002). In a real sense it is a reflection of accumulated experience and the role that reference dependency plays in choosing, in the spirit of prospect theory (Kahneman and Tversky 1979, Hensher 2006).

It is possible that the effect of reference dependence is positive for some individuals and negative for others (see Ailawadi *et al.*, 1999), suggesting that an unconstrained analytical distribution for the random parameterisation of state dependence is appropriate. A positive effect may be the result of habit persistence, inertia to explore another alternative, or learning combined with risk aversion. A negative effect could be the result of variety seeking or the result of latent frustration with the inconvenience associated with the currently used alternative (Bhat and Castelar 2002).

Thus, joint RP-SP estimations should not only recognise state dependence, but also accommodate heterogeneity in the reference dependence effect. Most RP-SP studies disregard reference dependence and adopt fixed parameters (i.e. homogeneity of attribute preference). Bhat and Castelar (2002) accommodate such unobserved heterogeneity in the reference dependence effect of the RP choice on SP choices. Brownstone *et al.* (1996), on the other hand, accommodate observed heterogeneity in the reference dependence effect by interacting the RP choice dummy variable with sociodemographic characteristics of the individual and SP choice attributes.

The paper outlines a very general mixed logit model which brings together the many recent contributions in the literature that deliver a flexible structure that can account for between-alternative error structure including correlated choice sets, RP-SP scale difference, unobserved preference heterogeneity, and reference dependency. An empirical example is used to illustrate the behavioural differences between the traditional NL-trick model and the flexible mixed logit model.

2. The Mixed Logit Framework

We begin with the basic form of the multinomial logit model, with alternative specific constants α_{ji} and attributes x_{ji}, for individuals i = 1,...,N in choice setting t

$$\text{Prob}(y_{it} = j_t) = \frac{\exp\left(\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit}\right)}{\sum_{q=1}^{J_i}\exp\left(\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit}\right)} \quad (1)$$

The random parameter model emerges as the form of the individual specific parameter vector, $\boldsymbol{\beta}_i$ is developed. The most familiar, simplest version of the model specifies

$$\beta_{ki} = \beta_k + \sigma_k v_{ik},$$
and (2)
$$\alpha_{ji} = \alpha_j + \sigma_j v_{ji},$$

where β_k is the population mean for the k^{th} attribute (k=1,…,K), v_{ik} is the individual specific heterogeneity, with mean zero and standard deviation one, and σ_k is the standard deviation of the distribution of β_{ik}'s around β_k. The term 'mixed logit' is increasingly used in the literature (e.g., Revelt and Train 1998, Train 2003, Hensher *et al.* 2005) for this model. The choice specific constants, α_{ji} and the elements of $\boldsymbol{\beta}_i$ are distributed randomly across individuals with fixed means.

The v_{ij}'s are individual and choice specific, unobserved random disturbances - the source of the heterogeneity. For the full vector of *K* random coefficients in the model, we may write the full set of random parameters as

$$\boldsymbol{\rho}_i = \boldsymbol{\rho} + \boldsymbol{\Gamma}\mathbf{v}_i. \quad (3)$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix which contains σ_k on its diagonal. For convenience we gather the parameters, choice specific or not, under the subscript '*k*.'. We can allow the random parameters to be correlated by allowing $\boldsymbol{\Gamma}$ to be a triangular matrix with nonzero elements below the main diagonal, producing the full covariance matrix of the random coefficients as $\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Gamma}'$. The standard case of uncorrelated coefficients has $\boldsymbol{\Gamma} = \text{diag}(\sigma_1,\sigma_2,\ldots,\sigma_k)$. If the coefficients are freely correlated, $\boldsymbol{\Gamma}$ is a full, unrestricted, lower triangular matrix and $\boldsymbol{\Sigma}$ will have nonzero off diagonal elements.

An additional layer of individual heterogeneity may be added to the model in the form of the error components. The full model with all components is given in (4), based on Greene and Hensher (2007).

$$\text{Prob}(y_{it} = j) = \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_i'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m E_{im}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_i'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m E_{im}\right]} \quad (4)$$

$(\alpha_{ji},\boldsymbol{\beta}_i) = (\alpha_j,\boldsymbol{\beta}) + \boldsymbol{\Gamma\Omega}_i\mathbf{v}_i$ are random alternative-specific constants and taste parameters; $\boldsymbol{\Omega}_i = \text{diag}(\omega_{i1}, \omega_{i2}, \ldots)$ and $\boldsymbol{\beta},\alpha_{ji}$ are constant terms in the distributions of the random taste parameters. Elements ω of the variance-covariance matrix represent the full generalized matrix. Uncorrelated parameters with homogeneous means and variances are defined by $\beta_{ik} = \beta_k + \sigma_k v_{ik}$ when $\boldsymbol{\Gamma} = \mathbf{I}$, $\boldsymbol{\Omega}_i = \text{diag}(\sigma_1,\ldots,\sigma_k)$, $\mathbf{x}_{jit}$ are observed choice attributes and individual characteristics, and $\mathbf{v}_i$ is random unobserved taste variation, with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$[1]**.** The individual specific underlying random error components are introduced through the term E_{im}, $m = 1,\ldots,M$, $E_{im} \sim N[0,1]$, given $d_{jm} = 1$ if E_{im} appears in utility for alternative j and 0 otherwise, and θ_m is a dispersion factor for error component m.

The probabilities defined above are conditioned on the random terms, $\mathbf{v}_i$ and the error components, $\mathbf{E}_i$. The unconditional probabilities are obtained by integrating v_{ik} and E_{im} out of the conditional probabilities: $P_j = E_{\mathbf{v},\mathbf{E}}[P(j|\mathbf{v}_i,\mathbf{E}_i)]$. This multiple integral, which does not exist in closed form, is approximated by sampling *nrep* draws from the assumed populations and averaging. See for example Bhat (2003), Revelt and Train (1998), Train (2003) and Brownstone *et al.* (2000) for discussion. Parameters are estimated by maximizing the simulated log likelihood given in (5).

$$\log L_s = \sum_{i=1}^{N} \log\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i} \frac{\exp\left[\alpha_{ji} + \boldsymbol{\beta}_{ir}'\mathbf{x}_{jit} + \Sigma_{m=1}^{M} d_{jm}\theta_m E_{im,r}\right]}{\sum_{q=1}^{J_i}\exp\left[\alpha_{qi} + \boldsymbol{\beta}_{ir}'\mathbf{x}_{qit} + \Sigma_{m=1}^{M} d_{qm}\theta_m E_{im,r}\right]} \quad (5)$$

with respect to $(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}, \theta)$, where R= the number of replications, $\boldsymbol{\beta}_{ir} = \boldsymbol{\beta} + \boldsymbol{\Gamma\Omega}_i\mathbf{v}_{ir}$ is the rth draw on $\boldsymbol{\beta}_i$, $\mathbf{v}_{ir}$ is the rth multivariate draw for individual i, and $E_{im,r}$ is the rth univariate normal draw on the underlying effect for individual i. The multivariate draw,

[1] Although not considered in the empirical case study, this model can accommodate correlated parameters with homogeneous means through defining $\beta_{ik} = \beta_k + \Sigma_{s=1}^{k}\Gamma_{ks} v_{is}$ when $\Gamma \neq$ I, and $\Omega_i = \text{diag}(\sigma_1,\ldots,\sigma_k)$, with Γ defined as a lower triangular matrix with ones on the diagonal that allows correlation across random parameters when $\Gamma \neq$ I.

v_{ir} is actually K independent draws. Heteroscedasticity is induced first by multiplying by Ω_i, then the correlation is induced by multiplying $\Omega_i v_{ir}$ by Γ.

The alternative-specific constants in (5) are linked to the extreme value Type 1 (EV1) distribution for the random terms, after accounting for unobserved heterogeneity induced via distributions imposed on the observed attributes, and the unobserved heterogeneity that is alternative-specific and accounted for by the error components. The error components account for correlated observations across choice sets that are administered to individual i as well as unobserved (to the analyst) differences across decision-makers in the intrinsic preference for a choice alternative (or preference heterogeneity). The parameter associated with each error component is defined as $\delta\sigma$, neither of which appears elsewhere in the model. We induce meaning by treating this parameter pair as θ which identifies the variance of the alternative-specific heterogeneity. What we are measuring is variation around the mean[2], hence the reference to a dispersion parameter.

Some specific features of the model of interest in joint estimation with multiple data sets are the possibility of 'state (reference) dependence' engendered in the SP data as a derivative of an RP market context; and the differences in the scale parameters for the SP data relative to the RP data. Formally, reference dependence is defined as (Bhat and Castelar 2002):

$$\varphi_q(1-\delta_{qt,RP}) \tag{6}$$

where $\delta_{qt,RP}$ = 1 if an RP obs, 0 otherwise, and φ_q is the parameter estimate of reference dependence which can be fixed or random. This variable enters the utility expression for each SP alternative, with the capability to select a generic specification.

The scale parameter for one data set (or set of alternatives) relative to the other set, the latter normalized to 1.0, is obtained through the introduction in one data set, the SP data[3], of a set of alternative-specific constants (ASCs) that have a zero mean and free variance (Brownstone *et al.* 2000). The scale parameter is calculated using the formula in equation (7).

$$\lambda_{qt} = [(1-\delta_{qt,RP})\lambda] + \delta_{qt,RP} \tag{7}$$

[2] The idea that beta is the coefficient on the unmeasured heterogeneity might be strictly true, but the concept does not work in other models that have error components in them, so we should not try to impose it here. For example, in the linear model, we have an unmeasured variable epsilon, and we write the model y = a +x'b + sigma*epsilon where, strictly speaking, epsilon is the unmeasured heterogeneity and sigma is the coefficient. But, sigma is the standard deviation of the unmeasured heterogeneity, not the "coefficient" on the unmeasured heterogeneity.

[3] We select the SP data set in the empirical application but the RP data set could have been chosen.

where $\delta_{qt,RP}$ is as defined above and λ is inversely proportional to the estimated standard deviation of the ASC of an alternative, according to the EV1 distribution, where $\lambda = \pi / \sqrt{6} StdDev$ = 1.28255/Std Dev of ASC.

This model with error components for each alternative is identified. Unlike other specifications (e.g., Ben-Akiva *et al.* 2001) that apply the results to identifying the scale factors in the disturbances in the marginal distributions of the utility functions, the logic does not apply to identifying the parameters on the attributes; and in the conditional distribution we are looking at here, the error components are acting like attributes, not disturbances. We are estimating the θ parameters as if they were weights on attributes, not scales on disturbances, and hence the way that the conditional distribution is presented. The parameters are identified in the same way that the βs on the attributes are identified. Since the error components are not observed, their scale is not identified. Hence, the parameter on the error component is ($\delta_m\sigma_m$), where σ_m is the standard deviation. Since the scale is unidentified, we would normalize it to one for estimation purposes, with the understanding that the sign and magnitude of the weight on the component are carried by θ. The sign of δ_m is not identified, since the same set of model results will emerge if the sign of every draw on the component were reversed – the estimator of δ would simply change sign with them. As such, we normalize the sign to plus, and estimate $|\delta_m|$, with the sign and the value of σ_m normalized for identification purposes.

3. The Data

In this section we illustrate the ideas discussed above using choice data in a transport context. The data were drawn from a stated-choice experiment that was conducted in six Australian capital cities: Sydney, Melbourne, Brisbane, Adelaide, Perth and Canberra (Hensher *et al*. 2005). The universal choice set comprised the currently available modes plus two 'new' modes, light rail and bus-based transitway (often referred to as a busway). Respondents evaluated scenarios describing ways to commute between their current residence and workplace locations using different combinations of policy-sensitive attributes and levels. The purpose of the exercise was to observe and model their observed coping strategies in each scenario.

Four alternatives appeared in each travel choice scenario: a) car (drive alone), b) car (ride share), c) bus or busway, and d) train or light rail. Twelve types of showcards described scenarios involving combinations of trip length (3) and public transport pairs (4): bus vs. light rail, bus vs. train (heavy rail), busway vs. light rail, and busway vs. train. Appearance of public transport pairs in each card shown to respondents was based on an experimental design.

Five three-level attributes were used to describe public transport alternatives: a) total in-vehicle time, b) frequency of service, c) closest stop to home, d) closest stop to destination, and e) fare. The attributes of the car alternatives were: a) travel time, b) fuel cost, c) parking cost, d) travel time variability, and for toll roads e) departure times and f) toll charge. The design allows orthogonal estimation of alternative-specific main effect models for each mode option: a) car no toll, b) car toll road, c) bus, d) busway, e) train, and f) light rail.

The master design for the travel choice task was a 27 x 3^{30} orthogonal fractional factorial, which produced 81 scenarios or choice sets. The 27-level factor was used to block the design into 27 versions each with three choice sets containing two alternatives. Versions were balanced such that each respondent saw every level of each attribute exactly once. The 3^{30} portion of the master design is an orthogonal main effects design, which permits independent estimation of all effects of interest. Two 2-level attributes were used to describe bus/busway and train/light rail modes, such that bus/train options appear in 36 scenarios and busway/light rail in 45.

In addition to the stated preference data, each respondent provided details of a current commuting trip for the chosen mode and one alternative mode. This enabled us to estimate a joint SP and revealed preference (RP) model of choice of mode for the journey to work. The data and detailed descriptions of the sampling process and data profile are provided in Hensher *et al*. (2005).

4. Results

Table 1 reports the final models for the NL-trick model and the unified mixed logit (UML) models. The UML model is a statistically significant improvement in overall goodness of fit after controlling for different number of parameters. The level of service variables are generic within the car and public transport (PT) modes in recognition of differences in marginal disutilities between car and PT attributes that are commonly reported in the wider literature, and travel cost is generic across all modes. Preference heterogeneity for each of these attributes is accounted for by random parameters. We investigated a large number of analytical distributions, including, normal, lognormal and triangular and found that the constrained triangular distribution gave the best fit as well as satisfying the negative sign condition of each parameter estimate[4].

The reference dependence effect was treated as a random parameter and was also assessed as a constrained and an unconstrained normal and triangular distribution. We were unable to establish any statistically significant influence of the actual (revealed) choice on the stated choices of the individual, reporting the constrained triangular results in Table 1.

The scale parameters for subsets of the SP alternatives were found to be statistically significant and greater than one for the bus and train modes; although not statistically

4 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/s2. The mean weighted average elasticities were also statistically equivalent.

significantly different from 1.0, the normalized value for the RP data. The car modes had a scale parameter of 2.963, suggesting a much lower variance on the unobserved effects associated with the EV1 random component; however it has a t-ratio of 0.89. What this suggests is that there is no serious violation of scale differences between the RP and SP data. This may be due in part to the capturing of relevant unobserved heterogeneity through attributes (i.e. random parameters) and alternatives (i.e., error components).

A number of alternative groupings of alternatives in the error components found that combining drive alone mode across the RP and SP data sets and the ride share gave statistically significant dispersion parameters than distinguishing RP and SP data. We found that separate error components for each of the car drive alone and car ride share alternatives across both RP and SP data sets gave statistically significant parameter estimate of 2.877 and 1.845 respectively in contrast to the fixed 1.0 for the full set of public modes (in RP and SP data sets). This suggests that there is substantial unobserved heterogeneity associated with the car alternatives, and especially the car drive alone alternative, that is greater than that associated with the public transport alternatives. This finding is intuitive given the large body of literature that suggests that the influencing attributes on car use are often more extensive (especially when including socio-demographic conditioning) than the set that determine public transport use. Given the generally dominant role of the car in many cities (notably 70-85 percent modal share in Australian cities), one might expect greater preference heterogeneity in the car choosers and hence an increasing likelihood of greater unobserved heterogeneity. Interestingly the scale parameter in the NL model of 0.7321 for public transport suggests greater unobserved heterogeneity for public transport modes than the car; however while this may be the appropriate interpretation for this model, the absence of accounting for correlated choice sets, random preference heterogeneity in the attributes and in the alternatives makes the comparison somewhat trite.

Table 1: Model Results for 'Nested Logit' trick vs. Panel Mixed Logit for combined Revealed Preference (RP) and Stated Preference (SP) Choice Data. The models were estimated in Nlogit 4. (cost is in dollars $AUD, time is in minutes)

Attribute	Alternative(s)	NL	Mixed Logit (RP EC Panel)
In-vehicle cost	All	-0.5802 (-14.7)	R: -.8534 (-14.17)*
Main mode time	RP and SP - DA, RS	-0.0368 (-6.4)	R: -0.1119 (-13.7)*
Main mode time	RP and SP - BS,TN,LR,BWY	-0.0566 (-8.2)	R: -0.0679 (-8.42)*
Access & egress mode time	RP and SP - BS,TN,LR,BWY	-0.0374 (-8.5)	R: -0.0524 (-9.72)*
Reference Dependence	DA, RS, BS, TN	ns	R: -.0917 (-.81)*
Personal income	RP and SP - DA	0.0068 (2.30)	0.01638 (3.46)
Drive alone constant	DA - RP	0.7429 (2.48)	2.3445 (7.26)
Ride share constant	RS - RP	-0.8444 (-3.1)	-0.9227 (-2.91)
Drive alone constant	DA - SP	0.0598 (0.36)	
Ride share constant	RS - SP	-0.2507 (-1.8)	
Train specific constant	TN -SP	0.1585 (1.40)	
Light rail specific constant	LR - SP	0.3055 (2.81)	
Busway specific constant	BWY - SP	-0.016 (-0.14)	
Bus specific constant	BS - RP	0.0214 (0.81)	0.1383 (0.51)
Random Parameter standard deviations:			
In-vehicle cost	All		.8534 (-14.17)*
Main mode time	RP and SP - DA, RS		0.1119 (-13.7)*
Main mode time	RP and SP - BS,TN,LR,BWY		0.0679 (-8.42)*
Access & egress mode time	RP and SP - BS,TN,LR,BWY		0.0524 (-9.72)*
Reference Dependence	DA, RS, BS, TN		.0917 (-.81)*
Sp to RP Scale Parameter	DA, RS		2.963 (0.89)
	BS, BWY		1.077 (6.48)
	TN, LR		1.058 (6.31)
Error Component (alternative specific heterogeneity)	RP and SP - DA,		2.877 (13.2)
	RP and SP - RS		1.845 (8.5)
Scale parameters	RP and SP – DA, RS	1.00 (fixed)	
	RP and SP - BS,TN,LR,BWY	0.7321 (8.85)	
Sample size		2688	2688
Log-likelihood at convergence		-2668.1	-2324.7
Value of travel time savings:		$ per person hour	
Main mode time	RP and SP - DA, RS	$3.81	$7.87
Main mode time	RP and SP - BS,TN,LR,BWY	$5.85	$4.77
Access & egress mode time	RP and SP - BS,TN,LR,BWY	$3.87	$3.68

Notes: * = constrained triangular random parameter, R = random parameter mean estimates, DA = drive alone, RS – ride share, BS – bus, TN – train, LR – light rail, BWY – busway. We used 500 Halton draws to perform our integrations, so there is no simulation variance. Ns: the reference dependence term was extremely insignificant.

Given the limitations of direct interpretation of parameter estimates from discrete choice models, behavioural contrasts are best made through outputs such as elasticities and willingness to pay estimates for specific attributes. The direct elasticities for invehicle cost and main mode time are given in Table 2, with values of travel time savings (VTTS) reported in Table 1. The VTTS for the unified mixed logit model are based on the conditional distributions (i.e., conditional on the alternative chosen). There are significant mean differences between the VTTS for main mode times for both car and public modes between the NL and UML models. There is significant under valuation for car under NL and slight over valuation for public modes. The access and egress VTTS for public modes are essentially the same.

When we look at the direct elasticities, we get some significant differences. We focus only on the elasticities associated with the revealed preference alternatives where the alternatives currently exist, since the SP estimates are somewhat meaningless given that the probabilities in the elasticity formula are based on hypothetical choice responses and hence market shares. The RP shares are calibrated market shares. For the new modes we have to rely on the SP estimates. On average there are some noticeable differences, especially for ride share where the mean cost elasticity is significantly lower for the UML in contrast to the NL-trick model; yet significantly higher for main mode time. Practitioners will use mean estimates only and indeed will obtain noticeably different market share predictions when using mean elasticity estimates from each model.

Table 2: Summary of Direct Choice Elasticities for Cost and In-vehicle Time

(mean and standard deviation across the sample)

Attribute	Alternative(s)	NL	Relativity	Mixed Logit (RP - EC Panel)
In vehicle Cost				
	RP DA	*-.108 (.195)*	>	*-.081 (.114)*
	RP RS	*-.363 (.419)*	>	*-.269 (.293)*
	RP BS	*-.510 (.516)*	>	*-.499 (.469)*
	RPTN	*-.638 (.538)*	<	*-.673 (.532)*
	SP LR	*-.544 (.384)*	<	*-.595 (.409)*
	SP BWY	*-.575 (.389)*	>	*-.559 (.382)*
	SP DA	-.812 (.630)		-.551 (.406)
	SP RS	-.957 (.639)		-.751 (.509)
	SP BS	-.545 (.410)		-.542 (.409)
	SP TN	-.568 (.409)		-.604 (.432)
Main mode time				
	RP DA	*-.110 (.190)*	<	*-.188 (.223)*
	RP RS	*-.396 (.387)*	<	*-.602 (.509)*
	RP BS	*-.571 (.583)*	>	*-.450 (.430)*
	RPTN	*-.763 (.664)*	>	*-.678 (.542)*
	SP LR	*-.657 (.325)*	>	*-.584 (.266)*
	SP BWY	*-.618 (.354)*	>	*-.481 (.264)*
	SP DA	-.536 (.372)		-.694 (.393)
	SP RS	-.639 (.375)		-.933 (.458)
	SP BS	-.586 (.339)		-.472 (.262)
	SP TN	-.701 (.330)		-.599 (.269)

5. Conclusions

This paper has set out a more general discrete choice model than the still very popular nested logit specification used by many practitioners and researchers to account for scale differences between multiple data sets, commonly a revealed preference data set and a stated choice set. The nested logit approach is limiting in that is not capable of accounting for the potential correlation induced through repeated observations on one or more pooled data sets. Nor does it recognize the role that various sources of heterogeneity play in influencing choices outcomes, either via the random parameterization of observed attributes and via parameterization of error components associated with a single or a sub-set of alternatives (alternative-specific heterogeneity).

The unified mixed logit model presented herein is capable of allowing for these influencing dimensions, observed or unobserved) in addition to accounting for scale differences (that are equivalent to the scale revealed in the NL model). The empirical example illustrates the additional outputs from the UML and the differences in key behavioural outputs.

References

Ailawadi, K.L., Gedenk, K., and S.A. Neslin (1999). Heterogeneity and purchase event feedback in choice models: an empirical analysis with implications for model buildings. *International Journal of Research in Marketing*, 16, 177-198.

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., (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences, *Transportation Research B*, 37(9), 837-855.

Bhat, C.R. and S. Castelar (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research B*, 36, 577–669.

Bradley, M. A. and Daly A. J. (1992) Uses of the logit scaling approach in stated preference analysis, paper presented at the *7th World Conference on Transport Research*, Lyon, July.

Bradley, M. A. and Daly, A. J .(1997) Estimation of logit choice models using mixed stated preference and revealed preference information, in Stopher, P.R. and Lee-Gosselin, M. (Eds) *Understanding Travel Behaviour in an Era of Change*, Pergamon, Oxford, 209-232.

Brownstone, D., Bunch, D.S., Golob, T., Ren (1996). A transactions choice model for forecasting demand for alternative-fuel vehicles. In S. McMullen, ed., *Research in Transportation Economics*, **4**, 87-129, JAI Press Inc.

Brownstone, D., Bunch, D.S., Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research, 34B, 315 338.

Greene, W. H. and Hensher, D.A. (2007) Mixed Logit and Heteroscedastic Control for Random Coefficients and Error Components *Transportation Research E*

Hensher, D.A. (1999) HEV Choice Models as a Search Engine for Specification of Nested Logit Tree Structures, *Marketing Letters*, 10 (4), 333-343.

Hensher, D.A. (2006) How do Respondents Handle Stated Choice Experiments? - Attribute processing strategies under varying information load, *Journal of Applied Econometrics*, 21, 861-878.

Hensher, D.A. and Bradley, M. (1993) Using stated response data to enrich revealed preference discrete choice models, *Marketing Letters*, 4(2), 139-152.

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

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

Kahneman, D. and Tversky, A. (1979) Intuitive prediction: biases and corrective procedures. In S. Makridakis and S. C. Wheelwright, Eds., *Studies in the Management Sciences: Forecasting*, 12 (Amsterdam: North Holland).

McFadden, D.L. (1981) Econometric models of probabilistic choice in *Structural Analysis of Discrete Data*, Manski, C.F. and McFadden, D.L. (eds.) MIT Press, Cambridge Massachusetts, 198-271.

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