ITS

WORKING PAPER
ITS-WP-03-18

Using Classical Inference Methods to reveal individual-specific parameter estimates to avoid the potential complexities of WTP derived from population moments

By

David A Hensher, William H Greene* and John M Rose

October, 2003

*Professor of Economics and Entertainment and Media Faculty Fellow, Department of Economics, Stern School of Business, New York University.

ISSN 1440-3501

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre in Transport Management

The University of Sydney and Monash University

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

NUMBER: Working Paper ITS-WP-03-18

TITLE: Using Classical Inference Methods to reveal individual-specific parameter estimates to avoid the potential complexities of WTP derived from population moments

ABSTRACT: A number of papers have recently contrasted classical inference estimation methods for logit models with Bayesian methods and suggested that the latter are more appealing on grounds of relative simplicity in estimation and in producing individual observation parameter estimates instead of population distributions. It is argued that one particularly appealing feature of the Bayesian approach is the ability to derive individual-specific willingness to pay measures that are claimed to be less problematic than the classical approaches in terms of extreme values and signs. This paper takes a close look at this claim by deriving both population derived WTP measures and individual-specific values based on the classical 'mixed logit' model. We show that the population approach may undervalue the willingness to pay substantially; however individual parameters derived using conditional distributions can be obtained from classical inference methods, offering the same posterior information associated with the Bayesian view. The technique is no more difficult to apply than the Bayesian approach – indeed the individual specific estimates are a by-product of the parameter estimation process. Our results suggest that while extreme values and unexpected signs cannot be ruled out (nor can they in the Bayesian framework), the overall superiority of the Bayesian method is overstated.

KEY WORDS: Classical Inference, Bayesian Inference, Mixed Logit, Stated Choice, Valuation.

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

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

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

DATE: October 2003

1. Introduction

Discrete choice models are the primary sources of estimates of willingness to pay (WTP) for specific attributes such as travel time savings. As choice modelling matures into a suite of models with increasing degrees of behavioural richness typified by the progression from multinomial logit, nested logit (NL), cross-correlated NL and mixed logit, analysts are increasingly exploring the deep parameterisation of WTP as a way of accommodating the heterogeneity of trade-offs in a sampled population. Such distributions of WTP can be derived from a set of moments portraying the population-level profile of a distribution (i.e., a mean and standard deviation with a specific analytical distribution (e.g., normal, lognormal, triangular, uniform etc.) or from parameters that are unique to each sampled individual. Individual-specific parameters are derived from conditional distributions in which known choices are taken into account in the spirit of the Bayesian posterior distributions.

With a growing interest in the Bayesian approach and claims that it is a more attractive paradigm than classical inference methods, the objective of this paper is to show how easy it is to obtain the equivalent information on individual parameters within the classical inference framework, and to derive such rich indicators of WTP distributions. We contrast the Bayesian-like estimates with the population specification more commonly associated with classical inference. From a policy perspective the empirical evidence is very revealing and disturbing, suggesting that the aggregation inherent in the population approach has suppressed the main moments of the distribution, namely the mean WTP in contrast to that associated with the WTP derived from the individual-specific parameterisation. The extant literature (e.g., Sillano and Ortuzar 2003) suggests that there is a greater incidence of negative WTP in the distribution derived from the population moments. We have found in our empirical setting the absence of any negative values under both approaches without imposing any constraints on the analytical distribution for the standard deviation of the random parameters in a mixed logit model.

The paper is organised as follows. We begin with a brief overview of the Bayesian approach, followed by a summary of the mixed logit model that will deliver the parameters. The data setting is then presented (a stated mode choice experiment for non-work trips in Sydney in 2003), followed by the findings and implications for deriving WTP from alternative interpretations of the mixed logit outputs.

2. The Bayesian Approach

Bayesian methods are often promoted as behaviourally different from and preferable to classical estimation methods currently used in estimation of advanced discrete choice models such as mixed logit. Brownstone (2001) provides a useful overview as do Chen et al (2000), Geweke (1999) and Train (2001) of the Bayesian perspective. Use of information on priors (as structural parameters) and posterior individual-specific parameter estimates from conditional utility functions are included as information to capture sources of heterogeneity[1].

[1] We capture within the classical estimation framework the same information that hierarchical Bayes modellers capture.

The key difference between Bayesian and classical statistics is that Bayesians treat parameters as random variables. Bayesians summarise their prior knowledge about parameters, q, by a *prior* distribution, $\pi(\mathrm{q})$. The sampling distribution, or likelihood function, is given by $f(x|q)$. After observing some data, the information about q is given by the *posterior* distribution:

$$p(\mathrm{q}|x) = f(x|\mathrm{q})\pi(\mathrm{q}) \Big/ \int f(x|\mathrm{q})\pi(\mathrm{q})d\mathrm{q} \tag{1}$$

We note for purposes below, that the posterior density is functionally equivalent to the conditional distribution of the parameters given the data. All inference is based on this posterior distribution. The usual Bayes estimator is the mean of the posterior distribution, and Bayesian confidence bands are typically given by the narrowest region of the posterior distribution with the specified coverage probability. Bayesian confidence regions are interpreted as fixed regions containing the random parameter q with the specified coverage probability (i.e., the 'highest posterior density' or HPD interval). This is different from the classical confidence region, which is a region with random endpoints that contain the true value q with the specified probability over independent repeated realisations of the data (Brownstone 2001). Classical inference therefore depends on the distribution of unobserved realisations of the data, whereas Bayesian inference *conditions on* the observed data. Bayesian inference is also exact and does not rely on asymptotic approximations.

The Bayesian approach requires the *a priori* specification of prior distributions for all of the model parameters. In cases where this prior is summarising the results of previous empirical research, specifying the prior distribution is a useful exercise for quantifying previous knowledge (such as the alternative currently chosen). In most circumstances, however, the prior distribution cannot be fully based on previous empirical work. The resulting specification of prior distributions based on the analyst's subjective beliefs is the most controversial part of Bayesian methodology. Poirier (1988) argues that the subjective Bayesian approach is the only approach consistent with the usual rational actor model to explain individuals' choices under uncertainty. More importantly, the requirement to specify a prior distribution enforces intellectual honesty on Bayesian practitioners. All empirical work is guided by prior knowledge and the subjective reasons for excluding some variables and observations are usually only implicit in the classical framework. Bayesians are therefore forced to carry out sensitivity analysis (which they rarely do, unfortunately) across other reasonable prior distributions to convince others that their empirical results are not just reflections of their prior beliefs (Brownstone 2001). The simplicity of the formula defining the posterior distribution hides some difficult computational problems, explained in Brownstone (2001)[2].

[2] Computing the posterior distribution typically requires integrating over q and this can be difficult for the number of parameters frequently encountered in choice modelling. Until recently Bayesians solved this problem by working with *conjugate families*. These are a family of prior distributions linked to a family of likelihood functions where the posterior distribution is in the same family as the prior distribution. For example, the Beta family is a conjugate prior for the binomial with fixed number of trials. Koop and Poirier (1993) have developed and applied a conjugate prior for the conditional (and multinomial) logit model, but there do not appear to be tractable conjugate priors for other GEV discrete choice models. Recent applications have circumvented these difficulties through the use of Gibbs Sampling and Markov Chain Monte Carlo Methods.

Huber and Train (2001) have explored the empirical similarities and differences between *hierarchical Bayes* and *classical* estimators in the context of estimating reliable individual-level parameters from sampled population data as a basis of market segmentation. The ability to combine information about the aggregate distributions of preferences with individuals' choices to derive conditional estimates of the individual parameters is very attractive. They conclude, however, that the empirical results are virtually equivalent conditional estimates of marginal utilities of attributes for individuals. What this debate has achieved in particular is to show classical estimation choice modellers that there is indeed more information in their estimation procedure that enables one to improve on the behavioural explanation within sample[3]. Recent developments in classical inference methods that are rich in deep parameters enable the analyst to obtain information that is *Bayesian-like*[4]. The mixed-logit model is one choice specification with this capability.

3. The Mixed Logit Model[5]

In the random utility model of the discrete choice family of models, we assume that a sampled individual ($q = 1,\ldots,Q$) faces a choice among J alternatives in each of T choice situations[6]. The individual is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternative with the highest utility. The (relative) utility associated with each alternative j as evaluated by each individual q in choice situation t is represented in a discrete choice model by a utility expression of the general form in (2).

$$U_{jtq} = \beta_q' \mathbf{x}_{jtq} + \varepsilon_{jtq} \tag{2}$$

where $\mathbf{x}_{jtq}$ is a vector of explanatory variables that are observed by the analyst (from any source) and include attributes of the alternatives, socio-economic characteristics of the respondent and descriptors of the decision context and choice task itself (e.g., task complexity in stated choice experiments as defined by number of choice situations, number of alternatives, attribute ranges, data collection method etc) in choice situation t.

---

[3] Within-sample priors such as the actual choice can help a great deal. When applying a model out-of-sample then Bayesians need some subjective priors.

[4] It is important to reinforce the fact that the nature of the randomness in Bayesian and classical approaches is different. In the classical view, the randomness is part of the model; it is the heterogeneity of the taste parameters, across individuals. In the Bayesian approach, the randomness 'represents' the uncertainty in the mind of the analyst (conjugate priors notwithstanding). Therefore, from the classical viewpoint, there is a 'true' distribution of the parameters across individuals. From the Bayesian viewpoint, in principle, there could be two analysts with two different, both legitimate but substantially different priors, who therefore could obtain very different, albeit both legitimate, posteriors. The idea that Bayesian estimation is exact in finite samples and has to cope with that proposition is worrying. We do not sense in the literature that this is actually considered appropriately.

[5] It is also referred to in various literatures as random parameter logit (RPL), mixed multinomial logit (MMNL), kernel logit, hybrid logit and error components logit.

[6] A single choice situation refers to a set of alternatives (or choice set) from which an individual chooses one alternative. They could also rank the alternatives but we focus on first preference choice. An individual who faces a choice situation on more than one occasion (e.g., in a longitudinal panel) or a number of choice sets, one after the other as in stated choice experiments, is described as facing a number of choice situations. Louviere et al (2000) provide a useful introduction to discrete choice methods that use data derived from repeated choice situations, commonly known as stated choice methods. Note that the assumption of a fixed choice set size, J, is made purely for convenience at this point; it is inessential.

The components b_q and ε_{jtq} are not observed by the analyst and are treated as stochastic influences.

Within a logit context we impose the condition that ε_{jtq} is independent and identically distributed (IID) extreme value type 1. The IID assumption is restrictive in that its does not allow for the error components of different alternatives to be correlated. We would want to be able to take this into account in some way. One way to do this is to partition the stochastic component additively into two parts. One part is correlated over alternatives and heteroskedastic, and another part is IID over alternatives and individuals as shown in equation (3) (ignoring the *t* subscript for the present):

$$U_{jq} = b_q{'}\mathbf{x}_{jq} + [\eta_{jq} + \varepsilon_{jq}] \tag{3}$$

where η_{jq} is a random term with zero mean whose distribution over individuals and alternatives depends in general on underlying parameters and observed data relating to alternative j and individual q; and ε_{jq} is a random term with zero mean that is IID over alternatives and does not depend on underlying parameters or data.

The Mixed Logit class of models assumes a general distribution for η_{jq} and an IID extreme value type 1 distribution for ε_{jq} [7]. That is, η_{jq} can take on a number of distributional forms such as normal, lognormal, or triangular. Denote the joint density of $[\eta_{1q}, \eta_{2q},..., \eta_{Jq}]$ by $f(h_q|W)$ where the elements of W are the fixed parameters of the distribution and h_q denotes the vector of J random components in the set of utility functions. For a given value of h_q, the conditional probability for choice j is logit, since the remaining error term is IID extreme value type 1:

$$L_{jq}(b_q|\eta_q) = \exp(b_q{'}\mathbf{x}_{jq} + \eta_{jq}) / \textstyle\sum_j \exp(b_q{'}\mathbf{x}_{jq} + \eta_{jq}). \tag{4}$$

The unconditional choice probability would be this logit probability integrated over all values of h_q weighted by the density of h_q is as shown in equation (5):

$$P_{jq}(b_q|W) = \int_{\eta_{1q}} \int_{\eta_{2q}} ... \int_{\eta_{Jq}} L_{iq}(b_q|h_q) f(h_q|W) d\eta_{Jq} ... d\eta_{2q} d\eta_{1q} . \tag{5}$$

Models of this form are called *mixed logit* because the choice probability P_{jq} is a mixture of logits with f as the mixing distribution. The probabilities do not exhibit the questionable independence from irrelevant alternatives property (IIA), and different substitution patterns may be obtained by appropriate specification of f. This is handled in two ways. The first, known as the random parameters specification, involves specifying each element of b_q associated with an attribute of an alternative as having both a mean and a standard deviation (i.e., it is treated as a random parameter instead of a fixed parameter[8]). The second, known as the error components approach, treats the unobserved information as a single separate error component in the random component (as done above in equations (4) and (5)).

[7] The proof in McFadden and Train (2000) that mixed logit can approximate any choice model including any multinomial probit model is an important message. The reverse cannot be said: the multinomial probit model cannot approximate all mixed logit models since the multinomial probit relies critically on normal distributions. If a random term in utility is not normal, then mixed logit can handle it and multinomial probit cannot. The description of mixed logit follows that given by Brownstone and Train (1999).

[8] A fixed parameter essentially treats the standard deviation as zero such that all the behavioural information is captured by the mean.

The standard deviation of an element of the b_q parameter vector, which we denote σ_{qk}, accommodates the presence of preference heterogeneity in the sampled population. While one might handle this heterogeneity in the context of a fixed β_{qk} parameter through data segmentation (e.g., a different model for each trip length range, age, gender and income of each traveller) and/or attribute segmentation (e.g., separate β_{qk}s for different trip length ranges), in contrast to treating it all as random, the challenge of these (deterministic) segmentation strategies is in picking the right segmentation criteria and range cut-offs that account for statistically significant sources of preference heterogeneity. The random parameters representation of preference heterogeneity is more general. However such a specification carries a challenge in that these parameters have a distribution that is unknown. Selecting such a distribution presents a challenge. The concern that one might not know the location of each individual's preferences on the distribution can be accommodated by retrieving estimates of individual-specific preferences by deriving the individual's conditional distribution based (within-sample) on their choices (i.e., prior knowledge). Using Bayes Rule, we first define the conditional choice probability as in equation (6):

$$\mathbf{H}_{jq}(b_q|W) = L_{jq}(b_q)g(b_q|W)/ P_{jq}(b_q| W) \tag{6}$$

where $L_{jq}(b_q)$ is now the likelihood of an individual's choice if they had this specific b_q, $g(b_q|W)$ is the distribution in the population of b_qs, and $P_{jq}(W)$ is the choice probability function defined in open-form as (seeTrain (2003)):

$$P_{jq}(W) = \int_{b_q} L_{jq}(b_q)g(b_q|W)\, db_q. \tag{7}$$

This shows how one can estimate the person specific choice probabilities as a function of the underlying parameters of the distribution of the random parameters.

The choice probability in (5) or (7) generally cannot be calculated exactly because the integral will not have a closed form. The integral is approximated through simulation. For a given value of the parameters, W, a value of b_q is drawn from its distribution. Using this draw, the logit formula (4) for $L_{iq}(b_q)$ is calculated. This process is repeated for many draws, and the mean of the resulting $L_{jq}(b_q)$'s is taken as the approximate choice probability giving the simulated probability in equation (8).

$$SP_{jq}(W) = (1/R)\sum_{r=1}^{R} L_{jq}(b_{qr}) \tag{8}$$

where R is the number of replications (i.e., draws of b_{qr}), b_{qr} is the r^{th} draw, and SP_{jq} is the simulated probability that an individual chooses alternative i.[9] It remains to specify the structure of the random vector b_q. In our application of this model, we will use the structure $b_q = b + D\mathbf{z}_q + G\mathbf{v}_q$ where the fixed underlying parameters are W = (b,D,G), b is the fixed mean of the distribution, $\mathbf{z}_q$ is a set of person-specific characteristics, D is a matrix of parameters, $\mathbf{v}_q$ is a vector of uncorrelated random variables with known variances on the diagonals of S, and G is a lower triangular matrix which, because

[9] By construction, SP_j 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.

Var[b_q] = GSG¢, allows free correlation of the parameters. Thus, a 'draw' from the distribution of b_q consists of a 'draw' from the distribution of $\mathbf{v}_q$ which is then used to compute b_q as shown above.

The simulation method was initially introduced by Geweke (and improved by Keane, McFadden, Börsch-Supan and Hajivassiliou - see Geweke et al 1994, McFadden and Ruud 1994) to compute random variates from a multivariate truncated normal distribution. The method produces consistent estimators of the choice probabilities. The cumulative distribution function in their research is assumed to be multivariate normal and characterised by the covariance matrix **M**. The approach is quick and generated draws and simulated probabilities depend smoothly on the parameters b and **M**. This latter dependence enables one to use conventional numerical methods such as quadratic hill climbing or gradient methods to solve the first order conditions for maximising the simulated likelihood function (equation 8) across a sample of $q = 1,\ldots,Q$ individuals; hence the term maximum simulated likelihood (MSL) (Stern 1997).

4. An Empirical Example

In the application of models, the posterior information accounts for the parameter variation across the sampled population, with the standard deviation (or spread) of each random β and the correlated inclusion for alternatives and choice situations being taken into account. This information is ignored in the priors. The procedure to distinguish prior and posterior information within sample is applied to a mode choice data set of 230 non-commuting trips of a sample of residents of the north-west sector of the Sydney metropolitan area interviewed in 2003. The centrepiece of the data collection activity is a stated choice experiment in which each sampled individual reviewed 10 mode choice sets. The main mode alternatives are car, existing bus, existing train, existing busway, new light rail (LR), new heavy rail (HR) and new busway (BW). Each public transport mode has an access and an egress component (see below). The data is collected using a Computer Aided Survey Instrument (CAPI), with all data being automatically captured in the CAPI into a data base formatted for immediate choice model estimation. More details are given in Hensher and Rose (2003).

Table 1 shows the descriptive statistics for the sample. The median age of the sample is 50, and the median employment category is casual. The average household size is 3.3 and the annual personal income level is \$10,000-\$15,000. 83.04 percent of the non-work segment had a car available for the surveyed trip (Table 2).

Table 1: Descriptive statistics for Non-Work segment

	N	Mean	Median	Std. Deviation	Minimum	Maximum
Age	230	N/A	50	70.97	0	70
No. of employed household members	230	N/A	3	1.25	1	4
Hours worked per week	230	15.60	8	18.98	0	80
Annual Personal Income (\$000's)	230	N/A	15	28.59	0	140
Household size	230	3.30	3	1.46	1	8
No. of children in household	230	0.73	0	1.04	0	4
Gender (male =1)	230	N/A	0	0.48	0	1

Table 2: Percentage of Non-Work segment who had a motor vehicle available for the trip

Car available	Frequency	Percent
yes	191	83.04
no	39	16.96
Total	230	100

The majority of the Non-Work trips consisted of Social/recreational trips (Table 3).

Table 3: Frequencies of trip purposes for Non-Work segment

Trip purpose	**Frequency**
Shopping	31
Visiting friends/relatives	27
Education	49
Social/recreational	81
Personal business	32
Other	10
Total	230

5. Stated Choice Experimental Design

The experimental design has 47 variables (46 in four levels and one in six levels for the blocks) in 60 runs; yielding six blocks of ten scenarios each. The design is almost orthogonal with maximum correlations between ± 0.06. The design allows the estimation of all alternative-specific main effects. Within each block the order of the runs has been randomised to control for order effect. There are different task configurations: with/without car, inter/intra regional, new LR & New HR versus new HR & new BW. A maximum number of complete designs have to be filled within each configuration. This is achieved in the field as follows: if the first respondent has a car on an intra regional trip with new LR & HR he is randomly assigned to a block (e.g., block three). If the second respondent is in the exact same configuration she sees the next immediate block (e.g., block four) otherwise she sees another randomly assigned block in one of the other configurations. Once all blocks in a configuration have been viewed, we randomly start at with another block.

The trip attributes associated with each mode are summarised in Table 4.

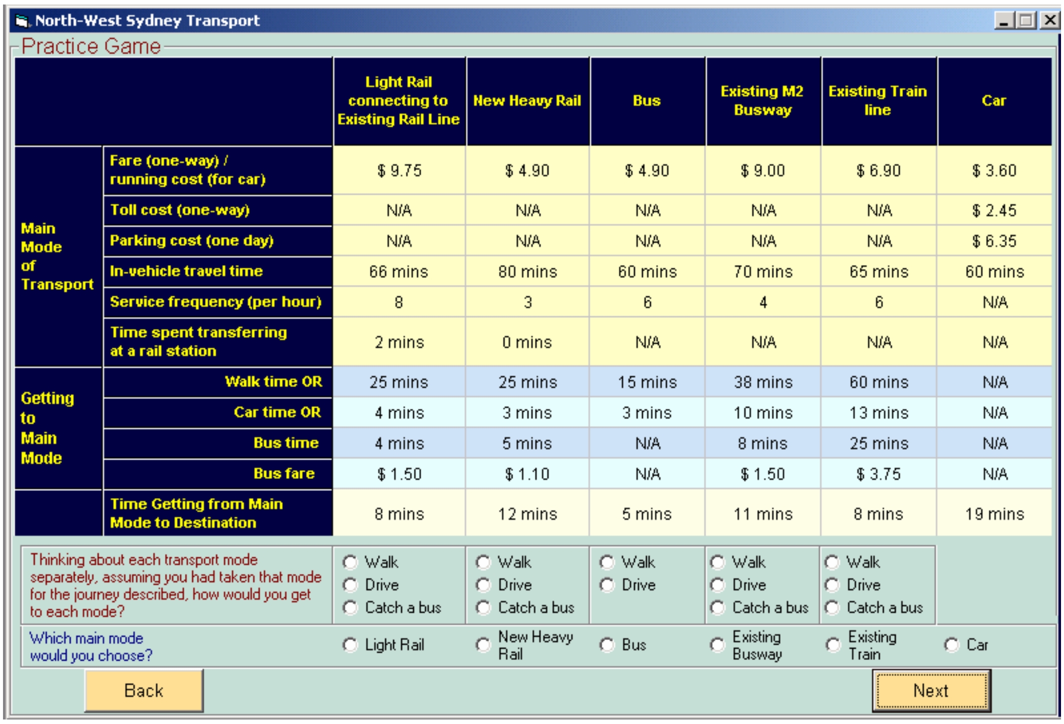
Table 4: Trip Attributes in Stated Choice Design

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	

Each design attribute has four levels. These were chosen as the following variations around the base level: -25%, 0%, +25%, +50%. The base times and costs used for new modes are shown in Table 5 where the locations are rail or busway stations. An example of a stated choice screen is shown as Figure 1.

Table 5: Base times and costs for new public transport modes

	Dollars	Busway	Heavy rail	Light Rail
	$	min	min	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				



North-West Sydney Transport
Practice Game

		Light Rail connecting to Existing Rail Line	New Heavy Rail	Bus	Existing M2 Busway	Existing Train line	Car
Main Mode of Transport	Fare (one-way) / running cost (for car)	$ 9.75	$ 4.90	$ 4.90	$ 9.00	$ 6.90	$ 3.60
	Toll cost (one-way)	N/A	N/A	N/A	N/A	N/A	$ 2.45
	Parking cost (one day)	N/A	N/A	N/A	N/A	N/A	$ 6.35
	In-vehicle travel time	66 mins	80 mins	60 mins	70 mins	65 mins	60 mins
	Service frequency (per hour)	8	3	6	4	6	N/A
	Time spent transferring at a rail station	2 mins	0 mins	N/A	N/A	N/A	N/A
Getting to Main Mode	Walk time OR	25 mins	25 mins	15 mins	38 mins	60 mins	N/A
	Car time OR	4 mins	3 mins	3 mins	10 mins	13 mins	N/A
	Bus time	4 mins	5 mins	N/A	8 mins	25 mins	N/A
	Bus fare	$ 1.50	$ 1.10	N/A	$ 1.50	$ 3.75	N/A
	Time Getting from Main Mode to Destination	8 mins	12 mins	5 mins	11 mins	8 mins	19 mins
Thinking about each transport mode separately, assuming you had taken that mode for the journey described, how would you get to each mode?		Walk / Drive / Catch a bus	Walk / Drive / Catch a bus	Walk / Drive	Walk / Drive / Catch a bus	Walk / Drive / Catch a bus	
Which main mode would you choose?		Light Rail	New Heavy Rail	Bus	Existing Busway	Existing Train	Car

Back Next

6. Findings

The final multinomial logit and mixed logit models are given in Table 6. The overall goodness of fit is similar to MNL and mixed logit. Although mixed logit is not a statistically significant improvement overall on the multinomial logit model, the statistically significant standard deviation parameter for in-vehicle time for public transport (with an unconstrained triangular distribution[10]) suggests that there is a

[10] For the triangular distribution, the density function looks like a tent: a peak in the centre and dropping off linearly on both sides of the centre. 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 $\sqrt{6}$.

structural advantage in selecting the mixed logit specification. In-vehicle cost and in-vehicle time for all public modes were specified as generic and separated from car costs and times. For car, the combination of running cost and toll was selected with parking cost treated separately.

A generic access time parameter (and likewise for wait time) best represents the marginal disutility of access time for all trips where the main mode is public transport. We found that the number of transfers enters into the public transport utility expressions as a separate effect. The age of the respondent and their gender both have a statistically significant influence on choice between public transport and car. All others things equal, the probability of choosing public transport decreases as age increases, more so for males.

Table 6: Summary of Empirical Results for Non-work Trips

Note: All public transport = (new heavy rail, new light rail, new busway, bus, train, busway); time is in minutes and cost is in dollars ($2003). T-values in brackets in columns 3 and 4.

Attribute	Alternatives	Multinomial Logit	Mixed Logit
New light rail constant	New light rail	3.3289 (7.54)	3.3701 (7.43)
New busway constant	New busway	1.6664 (3.74)	1.6988 (3.72)
Existing bus constant	Bus	1.5640 (4.02)	1.5705 (3.92)
Train constant	Existing and new Train	2.5478 (6.38)	2.5678 (6.25)
Existing busway constant	busway	1.8105 (4.52)	1.8270 (4.43)
Main mode fares	All public transport	-.3056 (-18.99)	-.3095 (-18.60)
Car mode running and toll cost	Car	-.20638 (-6.88)	-.2185 (-6.61)
Car parking cost	Car	-.12309 (-6.89)	-.1265 (-6.79)
Main mode in-vehicle time	All public transport	-.0350 (-17.62)	-.03679 (-14.26)
Number of transfers	All public transport	-.3707 (-5.90)	-.3753 (-5.90)
Access time	All public transport	-.0304 (-7.51)	-.03078 (-7.49)
Wait time	All public transport	-.04628 (-4.44)	-.04728 (-4.48)
Main mode in-vehicle time	Car	-.03455 (-6.8)	-.03715 (-6.43)
Egress travel time	All public transport	-.05011 (-10.85)	-.05061 (-10.78)
Egress travel time	Car	-.05637 (-3.50)	-.05765 (-3.47)
Access bus mode fare	Where bus is access mode	-.1746 (-5.36)	-.1773 (-5.38)
Age of respondent (yrs)	Public transport	-.0152 (-2.37)	-.01652 (-2.46)
Gender (male = 1)	Public transport	-.4472 (-2.26)	-.4484 (-2.18)
Standard deviation of parameters (triangular distribution)			
In--vehicle time	All public transport		.02871 (2.14)
In-vehicle time	Car		.00079 (.02)
Log-likelihood at zero		-3777.01	-3777.01
Log-likelihood at convergence		-2433.71	-2432.8
Pseudo-R^2		0.3542	0.3544
Sample Size		1941	1941

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

Behavioural values of travel time savings (VTTS) are summarised in Table 7 for the mixed logit non-work trip model using the population parameter estimates (i.e., priors). The only statistically significant random parameter is related to in-vehicle time for public transport. We limited the investigation of random parameters to in-vehicle time, access time and wait time. The mean VTTS's for in-vehicle time are intuitively plausible in absolute values and as a percentage of the average gross wage rate, with car higher than public transport. Wait time valuation is 2.1 times greater than the main

The height of the tent at c is 1/s (such that each side of the tent has area $s\times(1/s)\times(1/2)=1/2$, and both sides have area 1/2+1/2=1, as required for a density). The slope is $1/s^2$. See Evans et al (1993) for formal proofs.

mode in-vehicle time value while the access time (essentially in vehicle except for walking) is 1.36 times greater than the main mode time. Egress values are much higher than in-vehicle values for both public transport (2.25 times higher) and car (1.55 times higher). Overall there is evidence to support the position that VTTS for non main mode in-vehicle time savings is higher than for main mode in-vehicle time.

Table 7: Behavioural Values of travel Time savings ($/person hour): mixed logit model, non-work trips (mean gross personal income per hour = $14.27)

Willingness to Pay Attribute	VTTS ($/person Hour)	VTTS as Percentage of average gross wage rate
In-vehicle time – public transport	4.37 **(1.61)**	**30.6**
In-vehicle time – car	**10.21**	**71.55**
Waiting time – all public transport	**9.18**	**64.3**
Access time – all public transport	**5.97**	**41.8**
Egress time – all public transport	**9.83**	**68.9**
Egress time - car	**15.87**	**112.2**

Of particular interest herein is the derivation of the conditional individual-specific parameter estimates and the associated values of travel time savings for each individual. As described in Train (2002), we can obtain the conditional estimator for any individual by using Bayes Theorem. The estimator will be

$$\begin{aligned}
\mathrm{E}[\beta_q | \, data_q] &= \int_{\beta_q} \beta_q \, p(\beta_q \,|\, data_q) d\beta_q \\
&= \int_{\beta_q} \beta_q \frac{p(data_q \,|\, \beta_q) \, p(\beta_q)}{p(data_q)} d\beta_q \\
&= \int_{\beta_q} \beta_q \frac{p(data_q \,|\, \beta_q) \, p(\beta_q)}{\int_{\beta_q} p(data_q \,|\, \beta_q) \, p(\beta_q) d\beta_q} d\beta_q \qquad (9) \\
&= \frac{\int_{\beta_q} \beta_q \, p(data_q \,|\, \beta_q) \, p(\beta_q) d\beta_q}{\int_{\beta_q} p(data_q \,|\, \beta_q) \, p(\beta_q) d\beta_q}.
\end{aligned}$$

This is the empirical counterpart to (1). The prior density, $p(\beta_q)$ is specified after (8) where the distribution is induced by the stochastic specification of v_q. The conditional density is the contribution of individual *q* to the likelihood function. The denominator in the conditional mean is the theoretical contribution of individual *q* to the likelihood function for the observed data. That is, the choice probability defined in (7). The numerator of the expectation is a weighted mixture of the values of β_q over the range of β_q where the weighting function is, again, the likelihood function. Since the integrals cannot be computed analytically, we compute them, once again, by simulation.

The simulation estimator of the conditional mean for β_q is

$$\hat{E}_S[\beta_q] = \frac{(1/R)\sum_{r=1}^{R}\beta_{q,r}L(\beta_{q,r} \mid data_q)}{(1/R)\sum_{r=1}^{R}L(\beta_{q,r} \mid data_q)}. \tag{10}$$

where the weighting function in each case is the contribution to the likelihood function (not its log), computed at the r^{th} draw of $\beta_{q,r}$ in the simulation (see equation (8)). The approach in (10) can also be used to estimate the conditional variance or standard deviation of β_q by estimating the expected square and subtracting the square of the mean. This estimated conditional variance will be smaller than the average variance obtained simply by computing the sample variance of the estimated conditional means, as the latter is averaged over all the data in the sample while the former is averaged with respect only to the data for individual *q*.

The moments for the individual-specific parameter-derived VTTS are summarised in Table 8 for public transport in-vehicle time. The mean VTTS is $7.13 (49.9% of the average gross wage rate) in contrast to $4.37 based on the unconditional population-level distribution. The population distribution derived VTTS has a range of $1.56 to $7.13 with a standard deviation of $1.61; in contrast the range for the individual-specific VTTS is $4.31-$8.90 and a standard deviation of $0.41. Figure 2 graphs the VTTS distribution for the conditional distribution.

Table 8: Profile of VTTS based in Individual-Specific Parameters (Public transport in-vehicle time)

Mean	7.131
Standard Error	0.0093
Median	7.172
Mode	6.365
Standard Deviation	0.411
Sample Variance	0.169
Kurtosis	4.645
Skewness	-0.695
Range	4.586
Minimum	4.308
Maximum	8.895
Sample size	1941

Figure 3 shows a sample of the observations on individual willingness to pay. The figure shows the estimate and an interval defined by the mean plus and minus 2.5 conditional standard deviations. This is roughly the counterpart to the Bayesian HPD interval. Since the conditional distribution need not be symmetric, this interval may not be the narrowest interval containing 99% of the distribution, and, indeed, may contain slightly less than 95% of the distribution. However, it will contain at least 95% in any event.

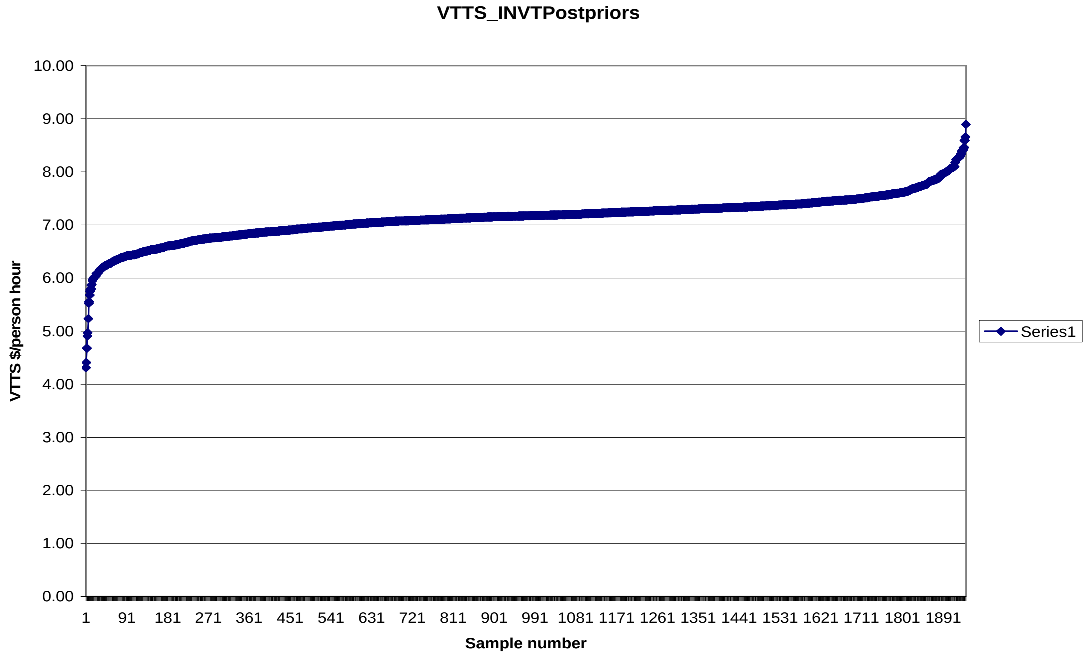


Figure 2: VTTS Distribution from Individual-Specific Parameters

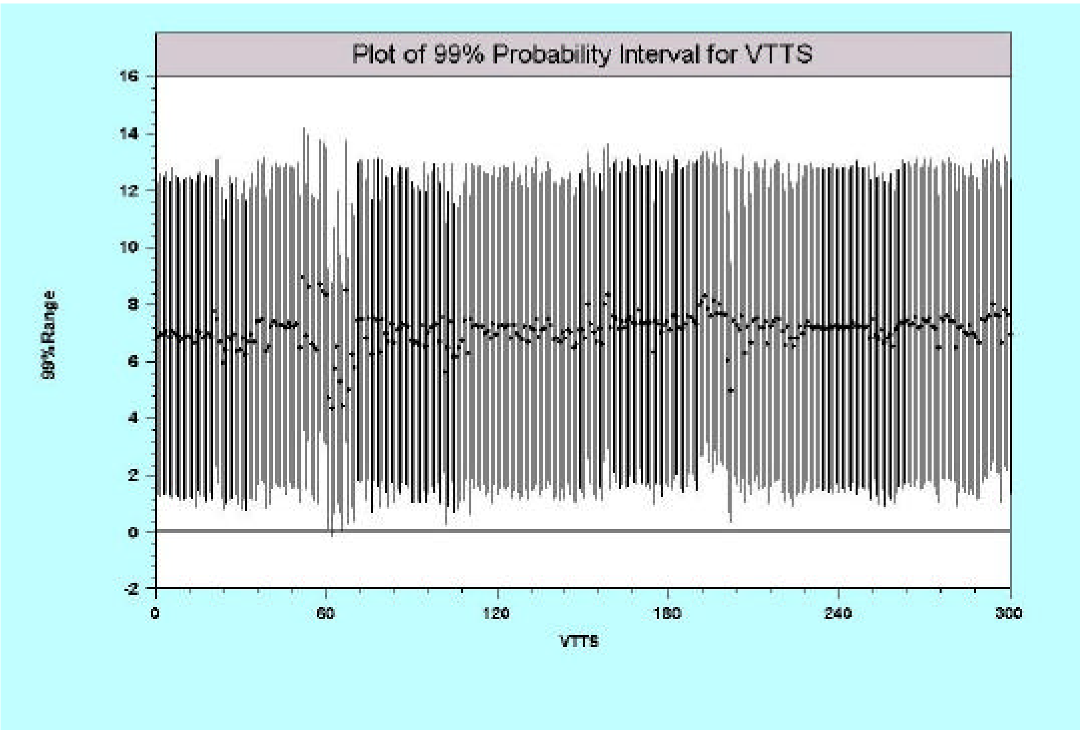


Figure 3: A random sample of VTTS from Individual-Specific Parameters

7. Conclusions

Allenby and Rossi (1999) have carried out an extensive Bayesian analysis of discrete brand choice and discussed a number of methodological issues relating to the estimation of individual level preferences. In comparison of the Bayesian and classical methods, they state the simulation based classical methods are likely to be extremely cumbersome and are approximate whereas the Bayesian methods are much simpler and are exact in addition. Both of these are overstated. The main premise of this study is to illustrate the simplicity of estimating individual level parameters in the random parameters discrete choice model. This can be applied to measures of willingness to pay, or other functions of the model components (see Greene 2003 for another example). The computation of individual level functions such as WTP or part worths is a simple by product of the computation of the simulation estimator (and is already incorporated in NLOGIT 3.0 and LIMDEP 8.0).

As to whether the Bayesian estimates are exact while sampling theory estimates are approximate, one must keep in mind what is being characterised by this statement. The two estimators are not competing for measuring the same population quantity with alternative tools. In the Bayesian approach, the 'exact' computation is of the analysts posterior belief about the distribution of the parameter (conditioned, one might note on a conjugate prior virtually never formulated based on prior experience), not an exact copy of some now revealed population parameter. The sampling theory 'estimate' is of an underlying 'truth' also measured with the uncertainty of sampling variability. The virtue of one over the other is not established on any but methodological grounds – no objective, numerical comparison is provided by any of the preceding or the received literature.

References

Allenby, G.M. and Rossi, P.E. (1999) Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89, 57-78.

Brownstone, D. (2001) Discrete choice modelling for transportation, in Hensher, D.A. (ed.) *Travel Behaviour Research: The Leading Edge*, Pergamon Press, Oxford, 97-124.

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

Chen, M-H., Shao, Q-M and Ibrahim, J. (2000) *Monte Carlo Methods in Bayesian Computation*, New York, Springer.

Evans, M., N Hastings and B. Peacock (1993) *Statistical Distributions*, New York, John Wiley and Sons.

Geweke, J. F. (1999) Using simulation methods for Bayesian econometric models: inference, development, and communication (with discussion and reply). *Econometric Reviews*, 18, 1-126

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.

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. (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.

Hensher, D.A., Louviere, J.J. and Swait, J. (1999) Combining Sources of Preference Data, *Journal of Econometrics*, 89, 197-221.

Huber, J. and Train, K. (2001) On the similarity of classical and Bayesian estimates of individual mean partworths, *Marketing Letters*, 12, August, 259-270.

Koop, G. and D. J. Poirier (1993) Bayesian analysis of logit models using natural conjugate priors. *Journal of Econometrics*, 56, 323-340

Louviere, J.J., Hensher, D.A. and Swait, J.F. (2000) *Stated Choice Methods and Analysis*, Cambridge University Press, Cambridge. (reprinted 2003).

McFadden, D. and Ruud, P.A. (1994) Estimation by simulation, *Review of Economics and Statistics*, LXXVI, (4), 591-608.

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

Poirier, D. J. (1988) Frequentist and subjectivist perspectives on the problems of model building in economics (with discussion and reply). *Journal of Economic Perspectives*, 2, 121-170.

Sillano, M. and Ortuzar, J. de Dios (2003) WTP estimation with mixed logit models: some new evidence, Department of Transport Engineering, Pontificia Universidad Católica de Chile, Santiago.

Stern, S. (1997) Simulation-based estimation, *Journal of Economic Literature*, XXXV. December, 2006-2039.

Train, K. (2001) A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit, Paper presented at the *Asilomar Invitational Choice Symposium*, California, June.

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