



ITLS

WORKING PAPER
ITLS-WP-05- 05

The signs of the times:
Imposing a globally signed
condition on willingness to pay
distributions

By

David A Hensher

July 2005

*Discussions with Bill Greene, John Rose and
Ken Train have been invaluable.*

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-05-05

TITLE: The signs of the times: Imposing a globally signed condition on willingness to pay distributions

ABSTRACT: A feature of recent developments in choice models that enable estimation of the distribution of willingness to pay (WTP), is that the sign of the distribution can change over the range. Behaviourally this often makes little sense for attributes such as travel time on non-discretionary travel, despite a growing recognition of positive utility over some travel time ranges. This can in part be attributed to the analytical distribution that is selected (except the cumbersome lognormal), many of which are unconstrained over the full range. Although a number of analysts have imposed constraints on various distributions for random parameters that can satisfy the single-sign condition, these restrictions are, with rare exception, only satisfied for the mean and the standard deviation estimates of a random parameter. When heterogeneity around the mean and/or heteroscedasticity around the standard deviation is allowed for, however, the constraint condition is often not satisfied. Given the popularity of distributions other than the lognormal, in order to satisfy the sign condition under the most general form of parameterisation, we need to impose a global sign condition. In this paper we show how this might be achieved in the context of the valuation of travel time savings for car commuters choosing amongst an offered set of route-specific travel times and costs. We illustrate the impact of the constraint under a globally constrained Rayleigh distribution for total travel time parameterisation, contrasting the evidence with a multinomial logit model and a range of other distributional assumptions within the mixed logit framework

KEY WORDS: *Willingness to pay distributions, random parameter distributions, mixed logit*

AUTHOR/S: David A Hensher

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

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

DATE: July 2005

1. Introduction

Willingness to pay (WTP) distributions are promoted in studies used to evaluate the user benefits of transport investments. The distribution of the value of travel time savings (VTTS), for example, is increasingly being derived in toll road patronage forecasting studies in many countries (see Hensher and Goodwin 2004, Ben-Akiva *et al.* 1993), especially in Europe and Australasia, as input into the generalised cost used in traffic assignment and in the establishment of monetary user benefits. With the widespread availability of software capable of estimating discrete choice models with random parameters, analysts are now able to empirically derive WTP distributions with relative ease. However, this new capability means that the possibility of producing distributions that are behaviourally questionable is also very high (see Hensher and Greene 2003, Sillano and Ortuzar 2005 for detailed reviews).

We are seeing studies generate such distributions using a myriad of unconstrained analytical distributions (e.g., uniform, normal, triangular, Rayleigh), often with little thought to the behavioural implications of unbounded distributions¹. The lognormal distribution is one exception; however its very long tail often results in a small number of very high VTTS (Ben-Akiva *et al.* 1993, Bhat 1998, 2000, Small *et al.* 2002, Revelt and Train 1998, and Brownstone and Train 1999). Although some authors (e.g., Meijer and Rouwendal 2000, Train and Sonnier 2002, Rigby and Burton 2004, Hess *et al.* 2005) have recognised this by imposing constraints on the standard deviation of the random parameter, which we refer to as a local constraint (e.g., a constrained triangular), to satisfy a single-sign WTP condition², this restriction has to date been demonstrated only when there are *no additional* deep parameters in the specification of the WTP. Such deep parameters exist, for example, when the analyst introduces heterogeneity around the mean of the random parameter (e.g. making travel time a function of personal income³) and/or introduces heteroskedasticity of the standard deviation of the random parameter. The sign constraint is not guaranteed (except for the lognormal) by restrictions on the standard deviation relative to the mean of a random parameter, in the presence of the additional underlying parameters, which we refer to as a global constraint.

This paper proposes a global constraint on the *entire* marginal disutility function for an attribute that restricts the contribution of the numerator and/or denominator of the WTP formula to guarantee a WTP distribution in the positive domain, assuming that this is behaviourally based. We introduce this constraint within the context of a mixed logit model in which additional deep parameters are permissible around the mean and standard deviation parameters associated with the travel time attribute in a study designed to estimate the VTTS distribution for car commuters in Sydney.

¹ Some analysts have noted this but have 'solved' it ex post, arbitrarily, by eliminating information that is incorrectly signed and/or removed very high values where there is a long positive tail. Sillano and Ortuzar (2005, page 541) make the important observation that '...we are simulating countless numbers of values for people who do not even exist'. Fosgerau (2004) makes a similar comment. The challenge however, is to decide on cut-offs, which is somewhat arbitrary. The author prefers to impose behavioural constraints on the analytical distributions, especially in terms of sign and also long tails.

² We acknowledge that some individuals might be willing to pay money to lose time up to a threshold, given the desire to have a minimum time commute (based on recent research by Mohktarian), but in the main, for commuters and other non-discretionary travel, the value of travel time savings is expected to be positive.

³ Adopting an alternative strategy of interacting travel time with income as a separate interaction term that is not part of the global constraint does not resolve this.

We begin the paper with a brief discussion of the mixed logit model to identify the components of a marginal (dis)utility expression where the sign restriction is imposed. We then discuss the way that the VTTS is obtained from ‘individual’ parameters estimated for each respondent, taking into account the prior information of their chosen alternative. The empirical study is then presented followed by the empirical evidence. The selection of a Rayleigh distribution is strictly illustrative; the focus herein is not on the merits of a specific analytical distribution but on the behavioural appeal of the imposition of a global sign condition⁴. We do, however present findings for the lognormal and an unconstrained normal distribution. Five models are compared: the standard multinomial logit (MNL – model 1) and four mixed logit models (locally unconstrained and constrained, and globally unconstrained and constrained, the latter two models including additional heterogeneity terms). We conclude the paper with caveats and warnings.

2. The Heteroscedastic Mixed Logit Model⁵

We assume that a sampled individual q ($q=1, \dots, Q$) faces a choice among J alternatives in each of T choice situations. 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 each individual q in choice situation t , is represented in a discrete choice model by a utility expression of the general form in (1).

$$U_{jtq} = \beta'_q \mathbf{x}_{jtq} + \varepsilon_{jtq}, \quad (1)$$

where \mathbf{x}_{jtq} is a vector of explanatory variables, including attributes of the alternatives, socio-economic characteristics of the individual and descriptors of the decision context in choice situation t . The components β_q and ε_{jtq} are not observed by the analyst and are treated as stochastic influences. ε_{jtq} is white noise following a Weibull distribution. The first of these, unlike its counterpart in the other discrete choice models usually examined, is assumed to vary across individuals.

Individual heterogeneity is introduced into the utility function through β_q . We allow the ‘individual-specific’⁶ parameter vector to vary across individuals both randomly and systematically with observable variables, \mathbf{z}_q . If the random parameters are assumed to be uncorrelated, then the model may be written as

$$\begin{aligned} \beta_q &= \beta + \Delta \mathbf{z}_q + \Sigma^{1/2} \mathbf{v}_q \\ &= \beta + \Delta \mathbf{z}_q + \boldsymbol{\eta}_q \end{aligned} \quad (2)$$

or

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

⁴ This point in part is added to ensure that readers do not overly focus on a specific distribution, detracting from the main objective.

⁵ This section draws on material in Greene *et al.* (2005).

⁶ Strictly, as explained in Section 3, these ‘individual-specific’ parameters are draws from a conditional distribution of the sub-sample of observations that have the same choice alternative.

where β_{qk} is the random parameter for the k^{th} attribute faced by individual q . The term $\beta + \Delta \mathbf{z}_q$ accomodates heterogeneity in the mean of the distribution of the random parameters. The random vector η_q endows the random parameter with its stochastic properties. In isolating the model components, we define \mathbf{v}_q to be a vector of uncorrelated random variables with known variances and denote the matrix of known variances of the random draws as \mathbf{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 $\Sigma^{1/2}$. In order to allow the random parameters to be correlated, we introduce the lower triangular matrix Γ , and extend the model to:

$$\beta_q = \beta + \Delta \mathbf{z}_q + \Gamma \Sigma^{1/2} \mathbf{v}_q. \quad (3)$$

Since the unknown scaling of the random components is already provided by the terms σ_k , the diagonal elements of Γ are normalized at one. The uncorrelated parameters model assumed in (2) then arises by assuming $\Gamma = \mathbf{I}$. The conditional variance of β_q is now

$$\text{Var}[\beta_q | \mathbf{z}_q] = \Gamma \Sigma^{1/2} \mathbf{W} \Sigma^{1/2} \Gamma'.$$

For the individual variance terms, denoting the k^{th} row of Γ as γ_k , we have

$$\text{Var}[\beta_{qk} | \mathbf{z}_q] = \gamma_k \Sigma^{1/2} \mathbf{W} \Sigma^{1/2} \gamma_k' = \sum_{i=1}^k \gamma_{ki}^2 (\sigma_i w_i)^2 \text{ where } \gamma_{kk} = 1 \quad (4)$$

and

$$\text{Cov}[\beta_k, \beta_m | \mathbf{z}_q] = \gamma_k \Sigma^{1/2} \mathbf{W} \Sigma^{1/2} \gamma_m' = \sum_{i=1}^k \gamma_{ki} \gamma_{mi} (\sigma_i w_i)^2 \quad (5)$$

where $\gamma_{ki} = 0$ if $i > k$ and γ_k is the k^{th} row of Γ .

The distribution of β_{qk} over individuals depends in general on underlying structural parameters $(\beta_k, \delta_k, \sigma_k, \gamma_k)$, the observed data, \mathbf{z}_q and the unobserved vector of K random components in the set of utility functions $\eta_q = \Gamma \Sigma^{1/2} \mathbf{v}_q$. The last of these represents a stochastic element that enters the utility functions in addition to the J random elements in ϵ_{iq} . Since β_q may contain alternative specific constants, covariation in η_{qk} induced by $\Gamma \neq \mathbf{I}$ will induce correlation of the random elements in the model across choices. Note that β_q , its component structural parameters $\Omega = (\beta, \Delta, \Gamma, \Sigma, \mathbf{W})$ and the characteristics of the person, \mathbf{z}_q do not vary across the alternatives.

We introduce variance heterogeneity into the model as follows: Let $\Sigma_q^{1/2} = \text{Diag}[\sigma_{q1}, \sigma_{q2}, \dots, \sigma_{qK}]$ where

$$\sigma_{qk} = \sigma_k \times \exp(\theta_k' \mathbf{h}_q) \quad (6)$$

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, Θ , to the model whose k^{th} row is the elements of θ_k . With this explicit scaling, the full model for the variances in our model is now

$$\text{Var}[\beta_q | \Omega, \mathbf{z}_q, \mathbf{h}_q] = \Phi_q = \Gamma \Sigma_q^{1/2} \mathbf{W} \Sigma_q^{1/2} \Gamma'. \quad (7)$$

The conditional variance of any specific parameter is now

$$\text{Var}[\beta_{qk}|\Omega, \mathbf{z}_q, \mathbf{h}_q] = \sum_{i=1}^k \gamma_{ki}^2 [\sigma_i w_i \exp(\theta_i' \mathbf{h}_q)]^2 \quad (8)$$

where w_k is the known scale factor $\mathbf{W}_{kk}^{1/2}$

We now have a functional form for an attribute in which its preference profile across a sample is represented by a mean and a standard deviation expression of the general form:

$$\beta_{qk} = \pm \exp[\beta_k + \delta_k' \mathbf{z}_q + \sigma_k \exp(\theta_k' \mathbf{h}_q) v_q] \quad (9)$$

where the sign for the entire expression is imposed by the analyst to represent the behaviourally required sign, v_q is an analytical distribution selected by the analyst, and all other terms are defined above. An empirical example for the travel time parameter in which all element of equation (9) are estimated, is given in equation (10).

$$\beta_{tq} = -\exp[-0.423 + 0.377 * \text{income} + 0.279 * \exp(-0.222 * \text{hours worked}) * v_q] \quad (10)$$

The value of travel time savings (VTTS) (in \$ per person hour) is calculated as $60 * \beta_{tq} / \beta_{c_q}$, where the denominator is assumed to be a fixed parameter. We can specify this global condition for either or both of the parameters β_{tq} and β_{c_q} used to derive VTTS. If both the numerator and the denominator are random then the specification would be $E(\beta_{tq} / \beta_{c_q})$ and not $E(\beta_{tq}) / E(\beta_{c_q})$. Confidence limits can be established for this distribution (see Armstrong *et al.* 2001 and Hensher *et al.* 2005) The parameter estimates in the numerator are individual-specific preferences and are not randomly drawn from the distribution, but conditioned on the known choice, as explained in the next section.

3. Individual Willingness to Pay Measures

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 distinction⁷ since we are not able to establish for each individual, their 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,

⁷ Discussion with Ken Train is appreciated.

$$f(\beta_q | \Omega, \mathbf{Y}_q, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q) = \frac{f(\mathbf{Y}_q | \beta_q, \Omega, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q) P(\beta_q | \Omega, \mathbf{z}_q, \mathbf{h}_q)}{f(\mathbf{Y}_q | \Omega, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q)}. \quad (11)$$

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. The second term gives the marginal probability density for the random β_q implied by the assumed distribution of \mathbf{v}_q . The denominator is the unconditional choice probability for the individual. 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 β_q is achieved 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 framework⁸, the mean of the posterior distribution, or the posterior mean. More generally, for a particular function of β_q , $g(\beta_q)$, such as β_q itself, the conditional mean function is

$$\mathbf{E}[g(\beta_q) | \Omega, \mathbf{Y}_q, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q] = \int_{\beta_q} \frac{g(\beta_q) f(\mathbf{Y}_q | \beta_q, \Omega, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q) P(\beta_q | \Omega, \mathbf{z}_q, \mathbf{h}_q)}{f(\mathbf{Y}_q | \Omega, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q)} d\beta_q \quad (12)$$

The integrals 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, Ω , and the observed data, $(\mathbf{Y}_q, \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q)$ a value of β_q is drawn from its distribution based on (2). For example, using this draw, the logit formula (12) for $L_{jtq}(\beta_q)$ is calculated. This process is repeated for many draws, and the mean of the resulting $L_{jtq}(\beta_q)$ ’s is taken as the approximate choice probability giving the simulated probability,

$$\hat{P}(\mathbf{Y}_q | \mathbf{X}_q, \mathbf{z}_q, \Omega) = \frac{1}{R} \sum_{r=1}^R L_{jtq}(\beta_{qr} | \mathbf{X}_q, \mathbf{z}_q, \mathbf{h}_q, \Omega, \eta_{qr}) \quad (13)$$

R is the number of replications (i.e., draws of β_{qr}), β_{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

$$\hat{E}_s[\beta_q | \text{Individual } q] = \frac{(1/R) \sum_{r=1}^R \beta_{q,r} L(\beta_{q,r} | \text{data}_q)}{(1/R) \sum_{r=1}^R L(\beta_{q,r} | \text{data}_q)}. \quad (14)$$

⁸ A referee suggested that we use a Bayesian estimation process. The advantages of Bayesian estimation are controversial. 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.

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 $\ln(SP)$ 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.

An application appears below.

4. Empirical Example

To illustrate the method of a globally constrained signed WTP function and its behavioural implications, we draw on a data set collected in Sydney in 2000 as part of a larger study (detailed in Hensher 2001) on valuing travel time savings for a number of privately funded tollroad projects. We use a subset of the data related to car commuting drivers. The centrepiece of the survey is a stated choice (SC) experiment. The attributes included in the choice experiments are given in Table 1.

The attributes of the stated choice (SC) alternatives are based on the values for the current trip¹⁰. In the design of the choice experiment, important considerations that needed to be accounted for were:

1. The toll should range from \$0 to \$16.
2. A longer trip should involve higher toll alternatives.
3. For a current trip without a toll, SC alternatives involving a toll should mostly be faster than the current trip.
4. We assume that the faster the trip, the higher the toll; the lower the running costs, the lower the free-flow time; and the lower the slowed down time the lower the uncertainty¹¹.

The overall design has 32 profiles that are blocked in two versions of 16 choice sets. Each choice set presents two stated choice alternatives and the current trip.

Table 1: Attributes in the Stated Choice Experiment

Variables	Number of levels	Range
Free flow travel time	4	% variation from the current
Slowed down travel time	4	% variation from the current
Uncertainty in travel time	4	% variation from the current
Running costs	4	% variation from the current
Toll Cost	4	0 to \$16, with levels determined by a nesting structure (see Hensher 2001)

5. Results

Seven models are compared: the standard multinomial logit (MNL – model 1) and six mixed logit models. Models two and three respectively use a locally unconstrained and constrained Rayleigh distribution without the two terms for heterogeneity around the mean and standard deviation; and models four and five respectively use a globally unconstrained and constrained Rayleigh distribution with the two terms for

¹⁰ Running costs have been specified as 10 litres/100kms. Fuel is priced at 97c/litre.

¹¹ To address issue number four, four nests were built into the design.

heterogeneity¹². In addition we report two models based on a lognormal (model six) and an unconstrained normal distribution (model seven). The focus of this paper is not on the selection of a distribution, but on illustrating the behavioural implications of imposing a global constraint, regardless of the distribution.

Given the focus on signed distributions, we have simplified the models by treating all components of time as having the same marginal disutility¹³ and likewise for each component of cost¹⁴. We allowed the number of intelligent draws (standard Halton) to vary from 50 up to 500, and assessed a number of socioeconomic variables as candidate sources of explanation of heterogeneity around the mean and standard deviation of the selected random parameter for total travel time. 300 draws produced statistically similar results to higher draws, and so we report all models estimated on 300 intelligent draws. The Rayleigh distribution probability function is given in (15).

$$P(r) = \frac{re^{-r^2/2s^2}}{s^2} \quad (15)$$

for $r \in [0, \infty)$. The [moments](#) about 0 are given by

$$\begin{aligned} \mu_m' &\equiv \int_0^{\infty} r^m P(r) dr = s^{-2} \int_0^{\infty} r^{m+1} e^{-r^2/2s^2} dr \\ &\equiv s^{-2} I_{m+1} \left(\frac{1}{2s^2} \right), \end{aligned}$$

where $I(x)$ is a [Gaussian integral](#) (Papoulis 1984, p. 148). The Rayleigh variable¹⁵ is a special case of the Weibull density¹⁶, with parameters 2 and $\frac{s}{2}$ where s is the desired scale parameter in the Rayleigh distribution. The mean is centred as $s^* \sqrt{\frac{\pi}{2}}$ and the standard deviation is $\sqrt{\frac{4-\pi}{2}} s^2$. This distribution has a long tail, but empirically appears much less extreme than the lognormal.

The final models are given in Table 2¹⁷. All the parameter estimates are statistically significant and of the expected sign, except for Model 4 and the models for the lognormal and the unconstrained normal. Models 4 and 5 are statistically better than Models 1 to 3 on the usual nested likelihood ratio test. Models 4 and 5 exhibit the same degrees of freedom with the difference being a global constraint (as per equation 10).

¹² While we want to structurally include as many elements of heterogeneity as possible via interacting attributes of the alternatives with characteristics of the respondents (which can be done within the MNL setting), there may still be other sources of heterogeneity that we cannot capture via observed attributes (Train 2003).

¹³ Indeed the mean estimates for free flow and slowed down time were not statistically different, although the standard deviations were.

¹⁴ We aggregated the toll and running cost as total cost and hence treated the marginal disutility of cost as the same for a unit of out-of-pocket running cost (i.e. fuel) and the toll paid.

¹⁵ In the current paper we use a conditional (on choice made) distribution but for an unconditional distribution, the empirical specification used is Rayleigh = $2^* (\text{abs}(\log(mu(0,1))))^{.70710678}$ where mu is the uniform distribution.

¹⁶ The Weibull(b,c) is: $w = b^* (-\log U)^{1/c}$.

¹⁷ To account for simulation variance inherent in the use of random number generators, multiple runs are recommended for random draws (Train 2003, 232). For Halton draws, however this is moot point since, no matter how many times the model is run, you get the same answer. Discussion with Bill Greene is appreciated.

The best model in terms of goodness-of-fit was obtained using the Rayleigh distribution. The overall goodness-of-fit of the model with an unconstrained normal distribution is close to that of the multinomial logit model in which the latter model has allowed for systematic heterogeneity through the interaction of travel time with two contextual effects. For our preferred model 5, the final indirect utility expression from which we calculate the marginal disutility of time and of cost, if we base the calculation on an unconditional distribution (in contrast to the conditional distribution of equation 12), is:

$$V = -0.8788 * \text{toll_route} + \exp[-0.6436 - 0.0035 * \text{time to answer} + 0.8716 * \exp(0.4292 * \text{gender}) * \text{rnr}] - 0.6742 * \text{total cost} \quad (16)$$

where $\beta_{t_q} = \exp[-0.6436 - 0.0035 * \text{time to answer} + 0.8716 * \exp(0.4292 * \text{gender}) * \text{rnr}]$ and $\beta_{c_q} = -0.6742$. The results reported in Table 3 are based on the (preferred) conditional parameter estimates, and hence cannot be identified from substitution into equation 16; instead β_{t_q} is derived from equation 14. The value of travel time savings distribution, for each of the constrained and unconstrained conditional distributions are given in Table 3 with the range (minimum, maximum) and standard deviation calculated from the full distribution of VTTS across the sample¹⁸. The Rayleigh distributions (models 2-5) are graphed in Figure 1.

Table 2: The Final set of Models (300 Halton draws). 7,056 observations (parameter estimates with t-statistics in brackets)

	MNL	Mixed Logit					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Attribute		Locally Unconstrained Rayleigh	Locally Constrained Rayleigh	Globally Unconstrained Rayleigh	Globally Constrained Rayleigh	Globally Constrained Lognormal	Unconstrained Normal
Travel cost (\$)	-0.4317 (-15.70)	-0.4839 (-10.9)	-0.5141 (-12.5)	-0.6557 (-13.9)	-0.6742 (-13.9)	-.7604 (-14.60)	-0.5728 (-13.75)
Toll constant	-1.0531 (-12.42)	-0.9641 (-9.57)	0.9866 (-10.9)	-0.9100 (-8.6)	-0.8788 (-8.1)	-0.8245 (-7.07)	-0.9244 (-8.93)
Travel time (minutes)	-0.1156 (-12.60)	-0.1898 (-5.98)	-0.0347 (-10.1)	-0.0203 (-1.16)	-0.6436 (-5.8)	-2.3001 (-17.26)	-0.0939 (-7.94)
Travel time std deviation		0.0667 (3.53)	0.0347 (10.06)	0.0977 (8.95)	0.8716 (8.2)	1.0257 (14.48)	0.1424 (7.21)
<i>Heterogeneity around the mean:</i>							
Travel Time (mins): Time to Answer each choice task (secs)*	0.00026 (1.97)			0.0004 (1.53)	-0.0035 (-2.0)	-.0030 (-1.66)	0.00033 (1.35)
<i>Heterogeneity around the variance:</i>							
Travel time (mins): gender (1=male)*	0.0372 (6.30)			-0.2631 (-4.3)	-0.4292 (-4.7)	ns	-0.4982 (-3.40)
Log-likelihood at convergence	-1806.4	-1821.57	-1816.82	-1792.55	-1789.70	-1797.9	-1805.05

*Note: In the MNL model, these starred variables are interactions of travel time respectively with 'time to answer a choice set' and gender.

¹⁸ The software Nlogit calculates the parameter estimates for equation (14) for each of the constrained and unconstrained distributions for each respondent. See Hensher et al. (2005, p 681).

Table 3: The VTTS for each Model (\$ per person hour)

	MNL	Mixed Logit					
VTTS	Model 1*	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
		Locally Unconstrained Rayleigh**	Locally Constrained Rayleigh	Unconstrained Rayleigh	Globally Constrained Rayleigh	Globally Constrained Lognormal	Unconstrained Normal**
Mean	11.66	8.47	11.43	10.91	10.67	20.73	9.16
Standard Deviation	2.46	4.79	1.82	5.53	6.14	15.00	5.29
Minimum	4.27	-3.78	7.41	1.01	2.31	7.63	-2.82
Maximum	15.99	18.91	21.88	36.51	33.49	117.17	32.36

*For MNL, the standard deviation and range is created from variations in gender and time to answer each choice set.

** In Model 7, 1.1 percentage of sample have negative VTTS. In Model 2, the percentage is 0.487. There appears to be no special features of the observations to establish any behavioural correlates with negative values.

The locally constrained Rayleigh (Model 3) in which we set the standard deviation parameter equal to two standard deviations from the mean¹⁹ artificially constrains the distribution and produces a narrowing of the range of VTTS. In contrast, allowing the entire distribution to be unconstrained (Model 4), produces the widest range, approaching zero²⁰ and, as is often in other studies, moving into the negative domain. In contrast to the unsigned specification of model 4, model 5 is globally constrained to ensure that the contribution of all sources of marginal (dis)utility to travel time do not produce a sign change that would result in some negative VTTS, *regardless of what analytical distribution is chosen*.

Figure 1 shows the influence of allowing for heterogeneity around the mean and standard deviation parameter estimates for travel time for the suite of models assuming a Rayleigh distribution. In particular we see significant movements down and up compared to model 3 as we move to the outer domains of the distributions. This confirms the importance of accounting for the heterogeneity beyond that revealed by the estimation of the standard deviation of the marginal (dis)utility of travel time. Reliance on an average may not be too bad under model 3 (given a mean of \$11.43 and a standard deviation of \$1.82), but this is risky for models 4 and 5 where the standard deviation VTTS are both over 50% of the mean VTTS.

¹⁹ There is any number of arbitrary restrictions.

²⁰ The objective of this paper is not to show negative VTTS in a distribution, but to alert researchers and users of WTP distributions of the very real risk of producing a set of negative values at the low end of a distribution.

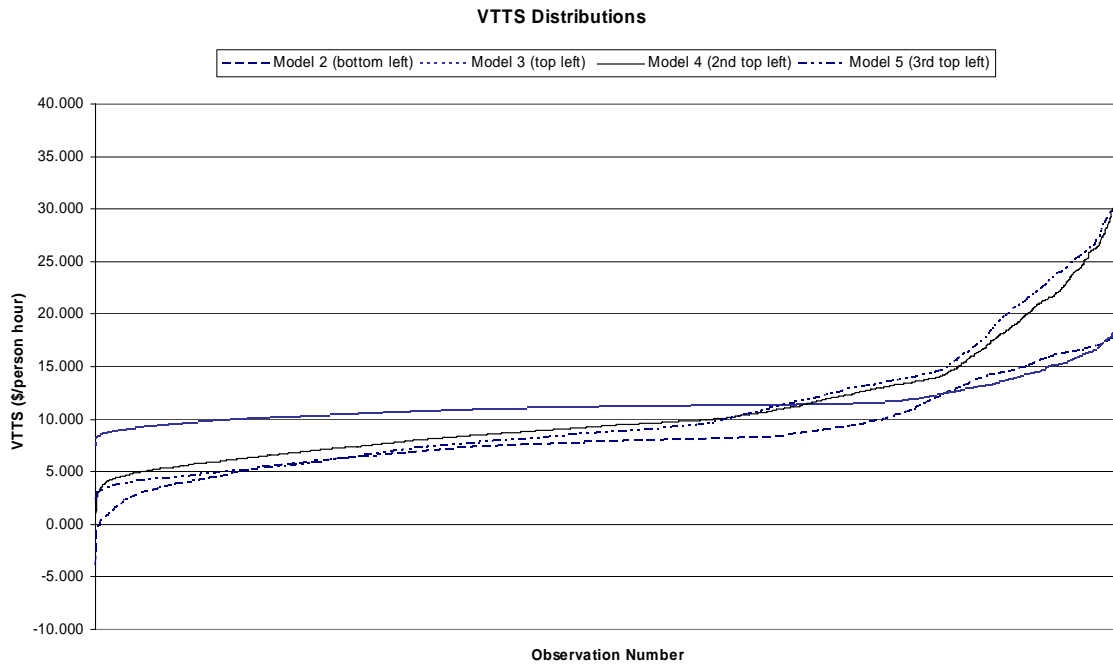


Figure 1: The VTTs distributions for each of the mixed logit specifications with Rayleigh

Looking at the sources of heterogeneity, gender has a statistically significant influence on the variance estimate. In Model 5, our preferred model, being male reduces the amount of heterogeneity in the distribution of travel time marginal (dis)utility. A source of heterogeneity around the mean is the time a respondent took to assess and make a choice in each stated choice screen, for a fixed design. All other things being equal, the mean estimate marginal (dis)utility increases the longer a screen is assessed²¹. Haaijer *et al.* (2000) show that information on time taken to make choice decisions, referred to as response latencies in the marketing literature, can be useful in obtaining better measurement of marginal utilities from observed choices.

We also ran models with a lognormal distribution (Model 6) and an unconstrained normal distribution (Model 7) to contrast the preferred globally constrained model using a Rayleigh distribution. The VTTs distributions are shown in Figure 2. True to form, the lognormal produced a very long tail, with VTTs as high as \$117.7 per person hour in contrast to \$33.49 for the global Rayleigh. The unconstrained normal is also reported to illustrate the presence of very low VTTs, including negative values, albeit only 1.1% in this example. The upper limit of VTTs for both the unconstrained normal and the globally constrained Rayleigh is almost identical, although the mean of the normal is lower due to the unconstrained distribution.

²¹ The reasons why a screen may take longer or quicker to assess and make a choice is of great importance in choice analysis, since it may signal a specific information processing strategy linked to either cognitive constraints or a very specific rational set of processing rules. See Hensher (in press a,b) for more details where it is recognised that relevancy is what matters and not complexity based on the number of items to process. Importantly, this does not imply, as one referee suggested, that "...it would be necessary to design a complex experiment to have much higher estimates and justify the construction of projects..." Rather it tells us that, for a given experiment, there is heterogeneity in the time to process each choice set, for any number of possible reasons.

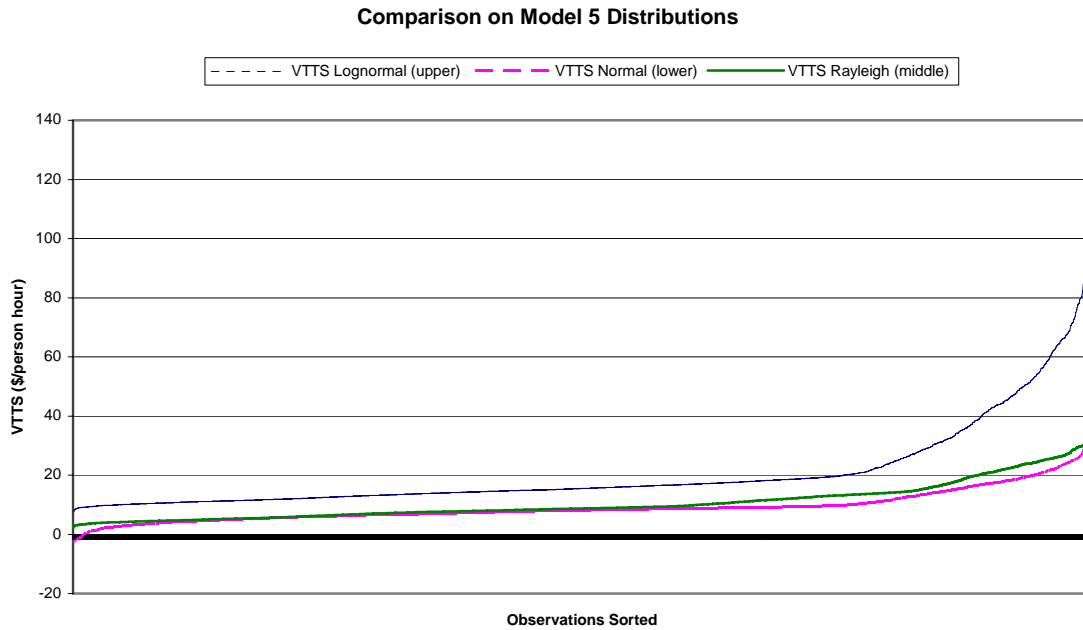


Figure 2: Comparison of VTTS distributions for globally constrained Rayleigh, lognormal and unconstrained normal.

6. Conclusions: Caveats and Warnings

The method proposed herein resolves the problem of behaviourally implausible sign *changes* when a general specification of a marginal disutility expression is included in a choice model *for any analytical distribution*. However this success has caveats. It does not guarantee that the single-sign domain will deliver plausible behavioural outputs.

The empirical evidence herein happens to produce an appealing spread of values with no bunching at the extremes, especially the zero point, which we have found in other empirical exercises with constraints on sign. Even the upper end of values does not exhibit the very long tail attributable to a *very small number* of observations that is common in lognormal applications. This may or may not be study specific.

The current state of practice imposes suitable constraints on the relationship between the mean and the standard deviation of an analytical distribution in order to guarantee the positive WTP and a behaviourally plausible profile in the positive domain under a number of analytical distributions, especially the constrained triangular (see Hensher and Greene 2003). As discrete choice models continue to evolve in sophistication, the opportunity to delve even deeper into causes of behavioural response becomes irresistible²². The approach developed in this paper signals a way forward to ensure that distributions imposed on specific attributes can comply with at least one important behavioural condition, that of the sign over the entire domain of its distribution.

²² In a recent paper, Greene and Hensher (2004) have generalised mixed logit to account for all the deep parameterisation set out herein but in addition have incorporated re-parameterisation of the residual unobserved heterogeneity after allowing for candidate random parameters. This additional set of deep parameters are alternative-specific and are further dimensionality within which to investigate generalised specification of the marginal disutility of specific attributes. See also Ben Akiva *et al.* (2001).

References

- Armstrong, P., Garrido, R. and Ortuzar, J de Dios (2001) Confidence intervals to bound the value of time, *Transportation Research*, 37E, 143-161.
- Ben-Akiva M. E., Bolduc D., and Bradley M. (1993) Estimation of travel choice models with randomly distributed values of time, *Transportation Research Record*, No. 1413, 88-97.
- 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. (1998) Accommodating variations in responsiveness to level-of-service variables in travel mode choice modeling, *Transportation Research* 32A (7), 495-507.
- Bhat, C.R. (2000) Incorporating observed and unobserved heterogeneity in urban work mode choice modeling, *Transportation Science*, 34 (2), 228-238.
- Brownstone, D. and Train, K. (1999) Forecasting new product penetration with flexible substitution patterns, *Journal of Econometrics*, 89, 109-129.
- Fosgerau, M. (2004) Investigating the distribution of the value of travel time savings, Danish Transport Research Institute, Copenhagen, November.
- Godoy, G. and Ortuzar, J. de D. (2005) On the estimation of mixed logit models, Department of Transport Engineering, Pontificia Universidad Catolica de Chile.
- Greene, W.H. and Hensher, D.A. (2004) Generalised mixed logit: separating out heterogeneity revealed through random parameters and heterogeneity through latent kernel effects, *Stern School of Business, New York University*.
- Greene, W.H., Hensher, D.A. and Rose, J. (2005) Accounting for heterogeneity in the variance of unobserved effects in mixed logit models, *Transportation Research B*.
- Haaier, R., Kamakura, W. and Wedel, M. (2000) Response latencies in conjoint choice experiments, *Journal of Marketing Research*, 37 (3), 376-382.
- Hensher, D.A. (2001) The Valuation of Travel Time Savings for Car Drivers: Establishing a Distribution for Sydney for Commuting and Non-Commuting Travel, Report prepared for Transfield Project Development, Sydney, April.
- Hensher, D.A. (in press –a) Revealing differences in behavioral response due to the dimensionality of stated choice designs: an initial assessment, *Environment and Resource Economics*.
- Hensher, D.A. (in press - b) How do Respondents Handle Stated Choice Experiments? - Attribute processing strategies under varying information load, *Journal of Applied Econometrics*.
- Hensher, D.A. and Goodwin, P.B. (2004) Implementation of values of time savings: the extended set of considerations in a tollroad context, *Transport Policy* 11 (2), 171-181.
- Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: the state of practice, *Transportation*, 30 (2), 133-176.

- Hensher, D.A., Rose, J. and Greene, W.H. (2005) *Applied Choice Analysis: A Primer*, Cambridge: Cambridge University Press.
- Hess, S., Bierlaire, M. and Polak, J. (2005) Estimation of value of travel-time savings using mixed logit models, *Transportation Research A* (2/3), 221-237.
- Huber, J. and Train, K. (2001) On the similarity of classical and Bayesian estimates of individual mean partworths, *Marketing Letters*, 12, August, 259-270.
- Meijer, E. and Rouwendal, J. (2000) Measuring welfare effects in models with random coefficients, Department of Econometrics, University of Groningen, July.
- Papoulis, A. (1984) *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, pp. 104 and 148.
- Revelt, D., and Train, K. (1998) Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level, *Review of Economics and Statistics*, LXXX (40), 647-657.
- Rigby, D. and Burton, M. (2004) Modelling indifference and dislike: a bounded Bayesian mixed logit model of the UK market for GM food, paper presented at Agricultural Economics Society 78th Annual Conference, Imperial College, London, 2-4 April.
- Sillano, M. and Ortuzar, D. de D. (2005) Willingness-to-pay estimation with mixed logit models: some new evidence, *Environment and Planning A*, 37, 525-550.
- Small, K., Winston, C. and Yan, J. (2002) Uncovering the distribution of motorists' preferences for travel time and reliability: implications for road pricing, Working Paper, Department of Economics, University of California, Irvine.
- Train, K. and Sonnier, G. (2002) Mixed logit with bounded distributions of part worths
- Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.
- Weisstein. E, W. Rayleigh Distribution in *MathWorld*-A Wolfram Web Resource. <http://mathworld.wolfram.com/RayleighDistribution.html>