An approach to the estimation of the distribution of marginal valuations from discrete choice data

Mogens Fosgerau, mf@dtf.dk
Katrine Hjorth, khj@dtf.dk
Stéphanie Vincent Lyk-Jensen, svl@dtf.dk
Danish Transport Research
Technical University of Denmark

April 30, 2007

Abstract

Models such as the mixed logit are often used to measure the distribution of the marginal value of a good based on discrete choice panel data. There are however serious specification and identification issues that are rarely addressed. The consequences for results may be dramatic. This paper points out the issues and presents an approach to dealing with them that may be applied under some circumstances. The issues and the approach are illustrated using a dataset designed to measure the value of travel time.

KEYWORDS: Discrete choice; valuation; mixed logit
JEL codes: C35, C14, Q51, R41

1 Introduction

A great number of studies in environmental, energy, health and transport economics attempt to measure the value of non-market goods from discrete choice data (Bateman et al., 2002). Many valuation studies use discrete choice data from stated preference or stated choice experiments. In such experiments, respondents are asked to indicate which alternative they prefer out of two or more. Alternatives are described in terms of a small number of characteristics one of which is price. Respondents typically make a series of choices.

A popular model for such data is the mixed logit model (McFadden and Train, 2000; Train, 2003). One often specifies indirect utilities of the alternatives as linear in characteristics plus extreme value error terms. Consider for example the indirect utility for an alternative with cost c and an amount of some good (or bad) t given by

$$U = \alpha c + \beta t + \varepsilon. \tag{1}$$

The parameters α and β represent the marginal utilities of cost and the good and ε is an error term. In this model, the marginal value of the good is $w = \beta/\alpha$, the ratio of marginal utilities.[1] In the mixed logit model, α, β or both are assumed to be random variables to allow for taste heterogeneity in the population. The distribution of α, β is known as a mixing distribution. It is common to assume "nice" distributions such as the normal or lognormal for these random parameters. However, the lognormal distribution is quite hard to apply while the normal distribution takes on both positive and negative values, which is unhelpful when there is an a priori sign restriction on the parameter.

The parameters α, β are a product of the parameters and the scale of the errors ε which is assumed to be a fixed value. If the errors are in fact heteroskedastic, then the scale are random and ignoring this will induce correlation between α and β, making the model even harder to estimate (Train and Weeks, 2004).

[1] In section 3.2 we shall distinguish between willingness to pay, willingness to accept, equivalent gain and equivalent loss, which is why we use terms such as marginal value with care.

With random parameters, the estimate of the mean marginal value becomes $E(\beta/\alpha)$. This quantity is hard to estimate and it is very sensitive to the assumptions made regarding the distributions of α and β, especially since β appears in the denominator.

In view of these disadvantages of the model in (1) it seems quite relevant to consider alternative model formulations, where the marginal value is not obtained as a ratio of random parameters, and where the scale of the error term is not confounded with w. In this paper we formulate a mixed logit model in terms of $\log w = \log(\beta/\alpha)$. Then we have only a unidimensional mixing distribution to worry about. Taking logs imposes a sign restriction on w that is desirable in many cases. By modelling the marginal value directly, the problem of the scale is moved to a less disturbing place.

Another issue is the distributional assumptions regarding α, β or w. Bad assumptions and identification problems can lead to the estimated mean of w being wrong by any order of magnitude (Fosgerau, 2006), and hence the choice of an adequate mixing distribution is of crucial importance. This issue is easier to tackle when there only one dimension of mixing. We shall make use of the results in Fosgerau and Bierlaire (2007) to specify a flexible mixing distribution that nests a desired mixing distribution such as the normal. We will also examine the extent to which the mixing distribution is nonparametrically identified. This is relevant in our application where insufficient range of the data gives us identification problems. We conjecture that such problems are common in applications.

The final issue we shall take with the mixed logit model as it is usually implemented is that it is somewhat hard in practice to let the marginal value of the good vary with observables. It is not obvious how to parametrise α, β as background variables may enter either coefficient and complicated functional forms may arise. When the model is formulated in terms of $\log w$ it is straight-forward to parametrise $\log w$. In this paper we will show how parametrising the log marginal value along with an independence assumption can assist with identification of the distribution of w as well as it will allow us to take account of reference-dependent preferences in a simple way (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Bateman et al., 1997; De Borger and Fosgerau, 2006).

We apply our approach to a dataset collected to evaluate the value of travel time. Our example application indicates the relevance of the concerns mentioned above. We see no reason that these issues should be specific to our data. This paper provides a readily usable approach to dealing with them. Our hope for this paper is hence that the methodology that we describe will see some use.

The methodology is presented in section 2. Section 3 presents an application to stated preference data designed to measure the value of travel time. Section 4 concludes.

2 Methodology

In this section we describe a simple modelling approach that allows us to deal with the problems mentioned concerning the model in (1). Section 2.1 sets out the model, section 2.2 discusses extension of the mixing distribution using the method of sieves, while section 2.3 discusses estimation of the mean of w.

2.1 Model formulation

Consider a choice between two alternatives characterised by $(c_i, t_i), i = 1, 2$, where c_i is cost and t_i is a bad. Assume further a linear indirect utility as in (1), but for now excluding the error term, i.e., $U_i = \alpha c_i + \beta t_i$ where $\alpha, \beta < 0$. We do not consider dominated choices and freely reorder alternatives such that alternative 1 is cheaper but worse in terms of t: $t_1 > t_2$ and $c_1 < c_2$. Then alternative 1 is preferred if $U_1 > U_2$ or equivalently if

$$\log\left(\frac{\beta}{\alpha}\right) < \log\left(-\frac{c_2 - c_1}{t_2 - t_1}\right) \tag{2}$$

For use below we define $v = \left(-\frac{c_2-c_1}{t_2-t_1}\right)$, the rate of trade-off between money and time presented in the choice situation. We formulate a choice model directly in terms of the marginal value $w = \beta/\alpha$. By assumption w is positive such that we can decompose w as $\log w = \delta x + u$, where the systematic part represents observed heterogeneity x through the index δx and where u is a random variable representing unobserved heterogeneity. We assume that x and u are independent. By including a constant in x we may assume that u has mean zero. We take u to be constant across observations from the same individual, x may depend on individual characteristics and on the choice situation itself.

We introduce subscript n to distinguish individuals and subscript r to distinguish choice situations. Then the model is that we observe

$$y_{nr} = 1 \Leftrightarrow \delta x_{nr} + u_n + \frac{1}{\lambda}\varepsilon_{nr} < \log v_{nr}, \tag{3}$$

where ε_{nr} are independent standard logistic errors and λ is the scale of the error term. Thus we have formulated a mixed logit model, where the distribution of u is the mixing distribution. The distribution of u determines the distribution of marginal values w conditional on x. The next section discusses the choice of mixing distribution. Note that the scale of the time and cost variables is now irrelevant for the choice probability, since only $\log v$ enters the model.

The present formulation allows us to work directly with the distribution of marginal values rather than with a ratio of random marginal utilities. It is hence better directed towards our object of interest, which is w. Somewhat similar approaches are used in Cameron and James (1987), Cameron (1988), and Train and Weeks (2004), who specify models where choices depend on a directly parameterized willingness-to-pay. Fosgerau (2007) uses nonparametrics to support the present model formulation in favour of model (1).

2.2 Approximating and testing the mixing distribution

We assume that the unobserved heterogeneity u is absolutely continuous, such that it has a density function ϕ. Many applications just assume a convenient density such as normal or lognormal. However, as the consequences of misspecification may be grave, we shall use the methodology in Fosgerau and Bierlaire (2007) to test an assumed density against a quite general alternative.[2] In case the assumed distribution is rejected against the more flexible alternative, we can just use the alternative instead.

The idea in brief is the following. Let ϕ be an assumed base density with corresponding distribution Φ and let g, G be the true density and distribution. In this paper we shall take ϕ as the normal density with standard deviation σ, but other densities are equally possible. Then provided the support of g is contained in the support of ϕ it is possible to transform Φ into G using $G = Q \circ \Phi$, where Q is a continuous distribution on the unit interval. Letting the density corresponding to Q be q, we approximate q using the method of sieves. We take $q_K(z|\gamma)$ as a flexible density on the unit interval with K parameters in γ and the property that any well-behaved density on the unit interval may be approximated as K tends to infinity. Following Bierens (2007), Fosgerau and Bierlaire (2007) use Legendre polynomials to approximate the true q, but other sieves may be used as well.

Let $P(y|x, \nu, u)$ be the choice probability conditional on u. Then we may approximate the true choice probability as follows.

$$\begin{aligned}\int g(u)P(y|x,\nu,u)du = & \int q(\Phi(u))\phi(u)P(y|x,\nu,u)du \\ = & \int_0^1 q(z)P(y,x,\nu,\Phi^{-1}(z))dz \\ \approx & \int_0^1 q_K(z|\gamma)P(y,x,\nu,\Phi^{-1}(z))dz\end{aligned}$$

Note that this formulation allows for panel data. This is in contrast to nonparametric approaches which do not take panel data into account and are only able to estimate the distribution of $u + \varepsilon/\lambda$ (Lewbel et al., 2002). The term $\Phi^{-1}(z)$ is just what one would get when this integral is approximated by simulation (Train, 2003). So the only difference from the standard mixed logit is the weight $q_K(z|\gamma)$.

Now the hypothesis that $g = \phi$ may be tested simply by testing whether $q_K = 1$, i.e., whether q_K is the density of the uniform distribution. If this

[2]The approach is implemented in freeware designed for the estimation of discrete choice models (Bierlaire, 2005) and is hence easy to use.

test is accepted, then q_K disappears from the likelihood and we are back to the standard mixed logit model. If not, then it is feasible to work with the more flexible distribution $Q_K(\Phi(\cdot)|\gamma)$.

2.3 Estimating the mean of w

2.3.1 Identification

Calculation of the mean of the marginal value w requires that we know the distribution of w or, given δx, the distribution of u. Even though we are able to check the fit of the distribution of u as described above, this only informs us about the distribution over the range where we have data. What we observe from data is whether $\delta x_{nr} + u_n + \varepsilon_{nr}/\lambda < \log \nu_{nr}$. So if the range of $\delta x_{nr} - \log \nu_{nr}$ is small, then we are not able to identify the distribution of u.

As an illustration we might consider estimating the mean of a random variable with distribution F. Say we have an estimate of F over some interval $[a; b]$. If $F(b) < 1$, then there is a positive probability mass located to the right of b, hence the mean may be arbitrarily high. Changing the assumptions we might make about the unobserved tail may have dramatic impact on the estimated mean. However, if $F(a) = 0$ and $F(b) = 1$ then the whole distribution is observed and the mean is identified.

In the present case, the data inform us about the distribution of $u + \varepsilon/\lambda$. Because of ε, this sum has support on the whole real line, so in principle it is necessary to have data such that $\delta x_{nr} - \log \nu_{nr}$ varies also over the whole real line. In this case, the distribution of u is identified (Fosgerau and Nielsen, 2005).

Note that the index assumption embodied in δx together with the assumption of independence between x and u contributes by extending the range over which we observe the distribution of $u + \varepsilon/\lambda$. Without the covariates in x then only the variation in $\log \nu$ would contribute toward identification of the distribution of u.

With finite data it is not possible to have $\delta x_{nr} - \log \nu_{nr}$ cover the whole real line. It is, however, useful to check how close we are to observing $u + \varepsilon/\lambda$ over its entire support as this will help detect a poorly identified distribution of u. A way to perform this check is simply to compute the range of predicted choice probabilities and to check how close this range is to the unit interval.

2.3.2 Derivation of the mean of w

From the definition of w and the assumption of independence between x and u, the mean of w is

$$E(w) = E(\exp(\delta x))E(\exp(u)) \tag{4}$$

The first part may be estimated as an average over the sample. If $u \sim N(0, \sigma^2)$ then $\exp(u)$ has a lognormal distribution with mean $E(\exp(u)) = \exp(\sigma^2/2)$.

In the generalised model the distribution of u is approximated by a flexible transformation of the normal distribution. Recall that we denote by ϕ and Φ the density and distribution of a normal random variable with mean 0 and standard deviation σ. The density of u is

$$g(u) = q(\Phi(u)\phi(u) \tag{5}$$

where the transformation q is a density on the unit interval. Using $E(\exp(u)) = \int_0^1 \exp(\Phi^{-1}(s))q(s)ds$, it is possible to approximate $E(\exp(u))$ by simulation.

3 Application

We illustrate the use of the ideas presented in the preceding section on a discrete choice dataset collected to evaluate the value of travel time. The value of travel time (VTT) is an essential notion in transport economics as the time savings evaluated by the VTT often constitute the major part of user benefits for an infrastructure investment. Many countries have launched VTT studies and official sets of VTT are provided in most Western countries.

In principle it is possible to study the distribution of the VTT from revealed preference data. However, the time and cost of trips are generally correlated. Therefore most studies of the VTT are based on stated choice experiments, where the time and cost variables may be controlled by the researcher (Gunn, 2000).

Section 3.1 introduces the dataset, section 3.2 specifies the models to be estimated and presents estimation results, while section 3.3 is concerned with the estimation of the mean VTT.

3.1 Data

Data are extracted from the Danish value of time study (Fosgerau et al., 2007). We use data from one stated choice design, an abstract time-cost exercise for in-vehicle time - and consider only car drivers.

All subjects in the experiment had to choose between two alternatives, described by travel time and travel cost. Alternatives differ only with respect to time and cost, so that issues such as heterogeneous preferences for various transport modes play no role. The travel time and cost was recorded for a recent actual trip subjects had made, these variables are labelled t_0 and c_0. We use observations with trip durations greater than 10 minutes, since for shorter durations it is hard to generate meaningful faster alternatives. We interpret the recent trip as the reference situation and generated choice situations by varying travel time and cost around the reference.

Each subject was presented with eight non-dominated choices, two in each of the four choice quadrants as shown in figure 1. Thus, each respondent had two equivalent gain (EG) type choices where one alternative was faster than the reference while the other alternative was cheaper than the reference. Equivalent loss (EL) choices are the mirror image of this. Similarly, respondents where presented with willingness to pay choices, where one alternative was the reference while the other alternative was faster but more expensive, and with willingness to accept choices, where again one alternative was the reference while the other alternative was slower but cheaper.[3]

[3]Subjects were also presented with a dominated choice situation, where one alternative was both faster and cheaper than the other. The quadrant for this choice situation was

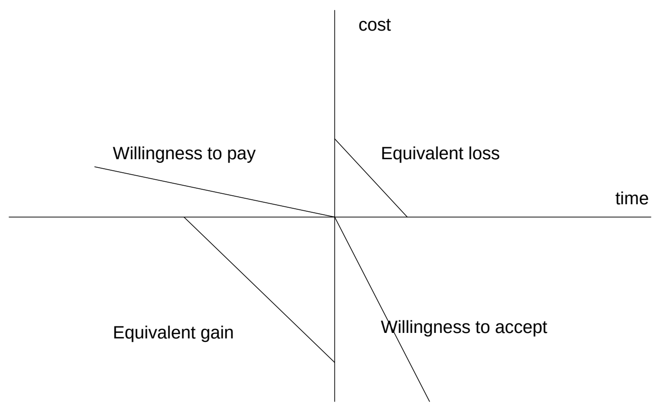

Figure 1: Choice quadrants

The eight choice situations were generated in the following way. First, eight choices were assigned to quadrants at random: two to each quadrant in random sequence. Second, two absolute travel time differences were drawn from a set, depending on the reference travel time, in such a way that respondents with short reference trips were only offered small time differences. Thus there is no asymmetry in the size of the time differences up and down. Both travel time differences were applied to the two situations assigned to each of the four quadrants. Third, eight trade-off values of time were drawn at random from the interval [2 ; 200] DKK per hour[4], using stratification to ensure that all subjects were presented with both low and high values. The trade-off values correspond to v in the econometric model. The absolute cost difference was then found for each choice situation by multiplying the absolute time difference by the trade-off value of time. Fourth, the sign of the cost and time differences relative to the

random. Respondents who chose the dominated alternative (the one being slower and more expensive) in the check question were excluded.

[4] 7.5 DKK = 1 EUR

Table 1: Descriptive statistics

Variable	Min	Mean	Max
y	1.00	1.61	2.00
$\log \nu$	−3.00	−0.54	1.21
c	−1.00	$-1.20 \cdot 10^{-3}$	1.00
t	−1.00	$1.35 \cdot 10^{-3}$	1.00
$\log \Delta t$	1.10	1.92	4.09
$\log c_0$	0.00	3.41	6.75
$\log t_0$	2.40	3.61	5.48
age	16.00	49.87	89.00
$\text{age}^2/100$	2.56	27.02	79.21
female	0.00	0.44	1.00
$\log(\text{income}) - 12$	−1.33	−0.04	1.06
low income	0.00	0.07	1.00
miss. income	0.00	0.07	1.00

reference were determined from the quadrant. The differences were added to the reference to get the numbers that were presented to respondents on screen. Travel costs were rounded to the nearest 0.5 DKK.[5]

Unrealistic answers from the respondents concerning travel distance, main mode journey time, travel cost, calculated speed, share of travel time due to congestion or travel group size led to exclusion of respondents. The remaining sample of car drivers consists of 1,819 respondents (14,178 observations).

The background variables available from the interview are sociodemographic characteristics (e.g., age, income, sex) together with details of the actual trip. Subjects stated their gross annual income, grouped into intervals of 100,000 DKK up to 1 million DKK. We have computed net annual income by applying national tax rates to interval midpoints. Table 1 provides some descriptive statistics for the estimation sample.

[5] In some cases, rounding caused the cost difference to be zero. These observations are omitted from the analysis.

3.2 Model specification and estimation results

We estimate two models, one with a minimal set of covariates (M1) and one with a range of covariates (M2). Each model occurs in two versions (M1,M1x and M2,M2x). In the first version u is assumed to be normal, in the second version the transformation of the density described in section 2.2 is applied.

Estimation is carried out in Biogeme (Bierlaire, 2003; Bierlaire, 2005), which allows for explicit estimation of the error scale λ, as well as for the generalised mixing density described above. We use 500 Halton draws to simulate the likelihood and note that this is sufficient to achieve stable results. Parameter estimates are presented in table 2 while table 3 shows the estimated likelihoods. We shall discuss each model in turn.

M1 comprises first a mean and standard deviation for the normal u. These parameters are very significant.

Bateman et al. (1997) and many others[6] indicate that people's choices differ from classical utility maximisation in systematic ways, such that the four valuation measures corresponding to the quadrants in figure 1 are generally different, with $WTP < (EG, EL) < WTA$. De Borger and Fosgerau (2006) develop the theory in Tversky and Kahneman (1991) to find that the differences between the four valuation measures can be captured by a constant level and two variables in the index δx that capture loss aversion in the time and cost dimension, respectively. In our experiment one of the suggested travel times is always equal to the reference t_0, while one of the costs is always equal to the reference c_0. Define $t = t_1 + t_2 - 2t_0$ and $c = c_1+c_2-2c_0$ and note that the sign of t and c indicates the quadrant as shown in table 5. We included the signs of t and c in the index with corresponding parameters labelled η_t and η_c in table 2. These parameters capture loss aversion and we expect $\eta_c < 0 < \eta_t$ (De Borger and Fosgerau, 2006). We find that the loss aversion parameters have the expected signs and that they are strongly significant. Their relative size indicates more loss aversion in the time dimension than in the cost dimension. Finally, the scale of the

[6] E.g., Hess et al. (2007) find evidence in stated choice data of asymmetric preferences around the reference.

Table 2: Model Estimates

	M1		M1x		M2		M2x	
	Estimate	Std.Err.	Estimate	Std.Err.	Estimate	Std.Err.	Estimate	Std.Err.
mean c	−1.188	0.046	−1.113	0.298	−0.385	0.449	−0.708	0.473
σ	1.551	0.053	1.585	0.339	1.193	0.042	1.387	0.224
η_c	−0.205	0.021	−0.205	0.022	−0.185	0.019	−0.185	0.019
η_t	0.406	0.023	0.408	0.023	0.366	0.020	0.366	0.020
λ	1.087	0.030	1.086	0.030	1.212	0.033	1.217	0.033
$\delta_{\Delta t}$					0.400	0.036	0.413	0.035
$\delta_{\log c_0}$					0.670	0.074	0.664	0.072
$\delta_{\log t_0}$					−0.845	0.108	−0.828	0.103
δ_{age}					0.003	0.017	0.001	0.016
δ_{age^2}					−0.029	0.017	−0.027	0.017
δ_{female}					−0.184	0.074	−0.156	0.071
$\delta_{\log income}$					0.702	0.107	0.692	0.101
$\delta_{low\ income}$					0.363	0.202	0.359	0.197
$\delta_{miss\ income}$					−0.062	0.138	−0.112	0.131
γ_1			−0.024	0.107			0.136	0.091
γ_2			−0.044	0.091			−0.114	0.069
γ_3			−0.023	0.052			−0.084	0.061
γ_4			0.200	0.084			0.211	0.083

Table 3: Model resume

	Log Likelihood	Parameters	Obs	Individuals
M1	−7145.13	5	13311	1709
M1x	−7131.59	9	13311	1709
M2	−6903.93	14	13311	1709
M2x	−6888.20	18	13311	1709

Table 4: Likelihood ratio tests

Unrestricted	Restricted	Log Likelihood difference	dof	p-value
M1x	M1	13.54	4	$1.9 \cdot 10^{-5}$
M2x	M2	15.73	4	$2.5 \cdot 10^{-6}$
M2	M1	241.20	9	$3.4 \cdot 10^{-98}$
M2x	M1x	243.39	9	$3.9 \cdot 10^{-99}$

error term in λ is close to 1.

In **M1x** we generalise the distribution of u with four additional parameters γ.[7] This extension is strongly significant as can be seen from tables 3 and 4. Hence we reject normality of u. In this case, the mean and standard deviation of the underlying normal distribution changes only a little when the mixing distribution becomes flexible, but there is no a priori reason the change could not be larger. The loss aversion parameters and the scale of the error term are unaffected.

[7] Further extension is not significant.

Table 5: Quadrants

WTP	$t < 0$, $c > 0$
EG	$t < 0$, $c < 0$
EL	$t > 0$, $c > 0$
WTA	$t > 0$, $c < 0$

M2 extends on M1 by including a number of variables in the index δx. Taken as a whole this extension is very significant.

- The first variable is the absolute time difference between alternatives. This variable allows for the effect that w increases with the size of the time difference. This effect strongly significant, it is not consistent with classical utility maximisation but it is consistent with reference-dependent preferences (Fosgerau, 2007; De Borger and Fosgerau, 2006).

- The next two variables are the log of reference cost and travel time. The corresponding parameters are again very significant. The estimates indicate that w increases in reference cost and decreases in reference travel time. The effect could be due to self-selection or maybe to the fact that travel times and costs are self-reported.

- Next we have included age and age squared. These variables are not significant at 5 per cent.

- The dummy for gender is significantly different from zero and indicates a lower value of travel time for women.

- Finally, we included log of personal net income together with dummies for the low income group and for missing income. We estimate an income elasticity of the value of travel time of 0.7.

Note that the estimate of σ has decreased relative to M1, indicating that inclusion of observed heterogeneity reduces the role for unobserved heterogeneity.

Finally, **M2x** again generalises the distribution of u with three additional terms and again this extension is very significant. Except for the mean and standard variable of the underlying normal mixing distribution there is only little change from M2 in the parameter estimates.

3.3 The mean value of travel time

Table 6 presents the ranges over the sample of the predicted probabilities of the four models. We have computed the probabilities for the cheaper and

Table 6: Range of estimated probabilities

	Min	Max
M1	0.137	0.911
M1x	0.146	0.909
M2	0.007	0.995
M2x	0.012	0.992

slower alternative 1. We see that large tails are missing in both the low and the high end for models M1 and M1x. The fact that the lowest probability is greater than zero indicates that we have not made the price of time v sufficiently low to make everybody choose the faster, more expensive alternative. This is an issue in models M1 and M1x where we miss the lowest 14 per cent. However, for the determination of $E(w)$ this is a minor issue since we have assumed that w is positive. The fact that the highest probability is less than 1, 9 per cent in models M1 and M1x, is a greater concern as it indicates that we lack information about the high end of the distribution of w. This means that the estimate of $E(w)$ is to a large extent determined by functional form and not by data. Introducing in M2 and M2x a range of variables in the index together with the independence assumption reduces this problem significantly.

We have simulated the distribution of w conditional on x for each individual in the sample. Together with the sample distribution of x this gives us the distribution of w conditional on the sample. Before doing this, however, we have removed the effect of loss aversion by setting the corresponding parameters to zero. Thus we obtain a distribution of marginal values rather than a mixture of WTP, WTA, EG and EL (De Borger and Fosgerau, 2006). We have also fixed the time difference between alternatives Δt to 10 minutes as this difference is also set by design. Table 7 presents some features of the estimated distribution of w for each model.

Consider first M1. The mean value of time is estimated at 60.8 DKK per hour which is a reasonable value considering that the sample average net hourly wage is about 100 DKK per hour. This might have been considered a plausible result had we not now been aware of the problems discussed

Table 7: Estimated VTT distributions

Model	Median	99% quantile	Truncated mean (at 99% quantile)	Mean
M1	18.3	673.7	54.2	60.8
M1x	19.1	1110.7	69.3	79.6
M2	22.9	535.2	53.3	57.5
M2x	23.2	789.0	60.2	67.6

above. We have rejected normality of u and we have seen that the range of data is rather small such that we can put no trust in the estimated left tail of the distribution of w. Relaxing the normality assumption in M1x deals with the first problem, not with the second. Moreover, the more flexible distribution has a very long tail to the left.

A way to resolve the issue would of course be to go back to collect more data, increasing the range of v. When this option is not available, we may instead choose to further parametrise w. We have done this in models M2 and M2x and found that this significantly reduces the problem of the missing tails.

Comparing M2 and M2x we find again that making the distribution of u more flexible reveals a longer left tail although the tendency is not as pronounced as for models M1 and M1x. So if we want to base our estimate of $E(w)$ on data less than on functional form assumptions then we must opt for model M2x.

4 Conclusion

We have presented an approach to the estimation of a distribution of marginal valuations from discrete choice data where the marginal valuation is a latent variable. The approach collects several ideas: formulation of the model directly in terms of the random marginal value rather than in terms of marginal utilities, parametrisation of the marginal value distribution, a method of sieves approach to model unobserved heterogeneity and a simple informal check on identification. These ideas allow us to work with the distribution of the latent variable in an informed way.

Our example application has shown that the identification and specification of the latent variable distribution are potentially problematic. The problems are not unique to our model formulation. It is rather the case that our model formulation makes these problems more visible than they would be in a discrete choice model formulated in terms of marginal utilities. We think it not unlikely that identification and specification problems occur in many applications of the mixed logit model.

The approach that we propose was clearly useful in our application and lead us to the specification of a model with a better identified and flexible latent variable distribution. Moreover, we were able to estimate quite many significant effects on the marginal valuation, something which is often hard to achieve. The main difficulty in applying the approach is implementing the method of sieves approximation to an arbitrary mixing distribution. This approximation is implemented in generally available freeware.

References

Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., zdemiroglu, E., Pearce, D. W., Sugden, R. and Swanson, J. (2002) *Economic Valuation with Stated Preference Techniques* Edward Elgar Cheltenham,UK.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997) A Test of the Theory of Reference-Dependent Preferences *The Quarterly Journal of Economics* **112**(2), 479–505.

Bierens, H. J. (2007) Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results *Econometric Theory* **Forthcoming**.

Bierlaire, M. (2003) *BIOGEME: a free package for the estimation of discrete choice models* Proceedings of the 3rd Swiss Transport Research Conference Monte Verit, Ascona, Switzerland.

Bierlaire, M. (2005) An introduction to Biogeme *www.epfl.ch/biogeme* .

Cameron, T. A. (1988) A New Paradigm for Valuing Non-Market Goods using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression *Journal of Environmental Economics and Management* **15**, 335–379.

Cameron, T. A. and James, M. D. (1987) Efficient Estimation Methods for "Closed-Ended" Contingent Valuation Surveys *Review of Economics and Statistics* **69**(2), 269–276.

De Borger, B. and Fosgerau, M. (2006) Another test of the theory of reference-dependent preferences: the trade-off between money and time.

Fosgerau, M. (2006) Investigating the distribution of the value of travel time savings *Transportation Research Part B: Methodological* **40**(8), 688–707.

Fosgerau, M. (2007) Using nonparametrics to specify a model to measure the value of travel time *Transportation Research Part A: Policy and Practice* **Forthcoming**.

Fosgerau, M. and Bierlaire, M. (2007) A practical test for the choice of mixing distribution in discrete choice models *Transportation Research Part B: Methodological* **Forthcoming**.

Fosgerau, M., Hjort, K. and Lyk-Jensen, S. V. (2007) *The Danish Value of Time Study - Final Report* www.dtf.dk.

Fosgerau, M. and Nielsen, S. F. (2005) Deconvoluting preferences and errors: a semi-nonparametric model for binomial data *Econometric Society European Meeting 2006* .

Gunn, H. (2000) An Introduction to the Valuation of Travel-Time Savings and Losses *Handbook of Transport Modelling* Elsevier Science Ltd. chapter 26.

Hess, S., Rose, J. and Hensher, D. (2007) Asymmetrical Preference Formation in Willingness to Pay Estimates in Discrete Choice Models *Working Paper* **ITLS-WP-06-12**.

Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk *Econometrica* **47**(2), 263–292.

Lewbel, A., Linton, O. and McFadden, D. (2002) Estimating features of a distribution from binomial data. unpublished work.

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

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

Train, K. and Weeks, M. (2004) Discrete Choice Models in Preference Space and Willingness-to-Pay Space *in* A. Alberini and R. Scarpa (eds), *Applications of Simulation Methods in Environmental Resource Economics* Kluwer Academic Publishers.

Tversky, A. and Kahneman, D. (1991) Loss Aversion in Riskless Choice: A Reference-Dependent Model *The Quarterly Journal of Economics* **106**(4), 1039–1061.