1. INTRODUCTION

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. What we are interested in is not only to know the average but also the distribution of the VTT.

The value of travel time can be measured by considering how people trade money for travel time. By observing choices among alternative combinations of cost and travel time, we can infer information about the relative weighting of cost and time, and from this the VTT can be derived. Most studies of VTT are based on Stated Preference (SP) analysis, due to the ability to control time and cost variables (Gunn 2000).

The state of the art from the literature is based on logit mixture models in preference space. In such models, alternatives are compared on basis of utilities, and the utility of an alternative is often assumed to be linear

\[ U = \alpha c + \beta t + \delta'x + \varepsilon, \]

where \( c \) and \( t \) denote the cost and travel time of the alternative and \( x \) is a vector of explanatory variables. \( \alpha \) and \( \beta \) are (negative) coefficients representing marginal utilities of cost and time, \( \delta \) is a coefficient vector, and \( \varepsilon \) is an error term.

In this model, the VTT equals \( \beta/\alpha \), the ratio of the marginal utilities of cost and travel time. Often \( \alpha \) and \( \beta \) are assumed to be random variables to allow for taste heterogeneity in the population. It is common to assume “nice” distributions such as the normal or lognormal for these parameters. The lognormal distribution is quite hard to apply while the normal distribution does not satisfy the sign restriction that \( \alpha \) and \( \beta \) must be non-positive. With random parameters, the estimate of the mean VTT becomes \( E[\beta/\alpha] \). This quantity is hard to estimate and it is very sensitive to the assumptions made regarding the distributions of \( \alpha \) and \( \beta \), since \( \alpha \) appears in the denominator.

It further complicates matters if the errors are heteroskedastic, i.e. if the variance of \( \varepsilon \) varies across the population or across choices of the same individual. Ignoring heteroskedasticity by fixing the scale of \( \varepsilon \) will induce
correlation between $\alpha$ and $\beta$, making the model even harder to estimate (Train and Weeks, 2004).

In view of these serious disadvantages of the model in eq. (1), it seems quite relevant to consider alternative model formulations, where the VTT is not obtained as a ratio of random parameters, and where the scale of the error term is not confounded with the VTT.

Another issue is which distributional assumptions one should make regarding $\alpha$, $\beta$ or the VTT (when using a logit mixture model, these distributions are called mixing distributions). Bad assumptions and identification problems can lead to the mean VTT being wrong by any order of magnitude (Fosgerau, 2006), and hence the choice of adequate mixing distributions is of crucial importance.

Finally, it is relevant to question some of the premises of standard economic utility theory, which often form the base of the discrete choice models used to evaluate VTT. Studies by Tversky and Kahneman (1979&1991) and many others indicate that people’s choices differ from classical utility maximisation in systematic ways. Particularly, the phenomenon of reference-dependence has been documented in many circumstances. Under reference-dependence the valuation of an attribute depends on a reference. Under the special case of loss aversion, losses are weighed more heavily than equally sized gains relative to the reference.

Many studies of evaluation of non-market goods find support of reference-dependent preferences. It is a common finding that selling price or “willingness-to-accept” (WTA) is larger than buying price or “willingness-to-pay” (WTP). This so-called WTA-WTP gap can be induced by loss aversion. Van de Kaa (2005) applies prospect theory to VTT evaluation and finds significant evidence of reference-dependence.

Even though reference-dependence is often supported by data, it is rarely incorporated in discrete choice models. However, we think there is sufficient evidence in favour of reference-dependent preferences to require that our modelling approach allows for reference-dependence.

This paper describes a simple modelling approach that circumvents the mentioned problems of the model in (1), allows for reference-dependent preferences and an easy test of the fit of a chosen mixing distribution as well as a check on its identification.

Based on Fosgerau (2005) we formulate a choice model directly in terms of VTT, instead of in terms of travel time and cost. This allows us to work directly with the VTT distribution rather than a ratio of distributions. A similar approach is also used in Cameron and James (1987), Cameron (1988), and Train and Weeks (2004), who specify models where choices depend on a separately parameterized willingness-to-pay. Our model formulation is made possible by the simple format of the SP data that we employ.
We specify log VTT as a linear index of covariates plus an additive random component representing unobserved heterogeneity. Using the modelling approach from De Borger & Fosgerau (2006), the VTT enters a reference-dependent utility function, which allows for loss aversion in the perceived values of travel time and cost. In binary choices, respondents choose the alternative with the higher reference-dependent utility (the lower generalised cost), though with the chance of making errors. The errors are assumed to be logistic, which makes our model a special case of a logit mixture model.

We assume that errors are multiplicative relative to the VTT, such that people with low VTT make errors of smaller absolute magnitude than people with high VTT.

This model specification works well, and has several advantages:

With present data, the model formulation yields considerable gains in model fit at low estimation cost.

- Reference-dependence is easily allowed for and can be tested.
- Errors are multiplicative such that the scale of the attributes of the alternatives does not affect choice probabilities.
- We work directly with the VTT distribution instead of having to derive the mean of a ratio of random variables correlated in some unknown (and perhaps spurious) way.
- We are able to test the distributional assumption for the random component of VTT by specifying generalised (flexible) distributions and testing against these.
- We are able to provide a check of the identification of the VTT distribution.

Furthermore, the model allows us to estimate a large number of significant parameters for the parameterisation of the VTT.

Our experimental design restricted the range of VTT trade-offs presented to respondents to the interval [2, 200] DKK $^2$ per hour. However, it turned out that a significant share of respondents had a VTT of at least 200 DKK per hour. Hence the right tail of the VTT distribution is not observed and the VTT distribution is not identified unless one is prepared to make strong assumptions regarding the distribution of the right tail.

Such assumptions cannot be verified and can result in any estimated mean VTT beyond a certain minimum (Fosgerau 2006). Although we are not able to make a strong statement about this, we conjecture that the problem of the missing tail is common to many stated choice experiments, reducing the robustness of results quite considerably.
This is an important reason for choosing the parametrise the VTT. We parametrise log VTT as a linear index of covariates representing observed heterogeneity and an independent random component representing unobserved heterogeneity. This allows us to extend the “observed” support of the independent random component to a wider range. Thus, the identification problem is resolved, but at the cost of assuming independence between the index and the random component.

Further advantages of the parametrisation is that it allows us to control for loss aversion and for the size of travel time change and it makes it possible to weight the sample in order to obtain a representative VTT for the Danish population.

The paper is organised as follows: Section 2 presents the methodology and data are described in section 3. The empirical model and variables are presented in section 4, while section 5 shows the results. Section 6 concludes the paper.

2. METHODOLOGY

2.1. Model formulation

Parametrisation of the VTT
We assume that the VTT is positive, and is composed of a systematic part depending on characteristics \( x \) of the respondent and the choice situation, and a random part, that varies across individuals, but is constant across choices of the same individual. That is, \( VTT = \exp(x'\beta + u) \), where \( u \) is a person-specific random variable with mean zero, and \( x \) and \( u \) are independent.

Thus \( u \) represents unobserved taste heterogeneity among individuals and the distribution of \( u \) determines the distribution of the VTT conditional on \( x \). Since we work in a logit mixture framework, the distribution of \( u \) will be referred to as the mixing distribution. Assumptions on the distribution will be discussed later in this section. In deriving our model we need only assume that \( u \) is absolutely continuous, such that it has a density function \( f \).

Reference-dependent utility function
Under reference-dependent preferences, when people are to choose between travel alternatives characterised by attributes like travel time and cost, they base choices on the perceived values of the attributes, rather than the actual values. The perceived value depends on a reference, which may be the current situation, and on whether the attribute represents a loss or a gain relative to the reference. The relation between perceived and actual values is described by a so-called value function.

We shall apply linear value functions of time and cost attributes. Consider a person receiving an amount \( m \) of some good. The value he assigns to \( m \) depends on whether \( m \) is perceived as a gain (\( S(m) > 0 \)) or a loss (\( S(m) < 0 \)) compared to its reference amount e.g. the initial endowment. The perceived value of \( m \) is given by
\[
V(m) = m \cdot e^{-\eta S(m)},
\]

where \(S(m)\) denotes the sign of \(m\).

We assume that \(\eta > 0\), which implies that for \(m \geq 0\), \(-V(-m)\) is larger than \(V(m)\). Hence the value function in eq. (2) exhibits loss aversion: The monetary equivalent of giving up a certain amount of the good exceeds the price he is willing to pay to obtain the same amount of the good.

In line with De Borger & Fosgerau (2006), we assume that the utility of an alternative with travel time \(t\) and travel cost \(c\) is

\[
U(t, c | t_0, c_0) = V_t(-(c - c_0)) + V_c(-VTT(t-t_0))
\]

where \(t_0, c_0\) are the reference travel time and cost, \(V_t, V_c\) are the value functions for time and cost, and \(VTT\) is the monetary value of travel time. We take the reference values to be those of the current trip around which the choice alternatives are obtained as pivots.

**Choices and choice probabilities**

We shall model binary choices in which one alternative is fast and expensive (alternative 1), and the other is cheaper and slower (alternative 2). The travel times and costs of the alternatives are referred to as \(t_1, c_1\) and \(t_2, c_2\).

In our experiment one of the suggested travel times is always equal to \(t_0\), while one of the costs is always equal to \(c_0\). We therefore consider four main types of choices:

**Table 1: Main choice types**

<table>
<thead>
<tr>
<th>WTP (willingness-to-pay):</th>
<th>(t_1 &lt; t_0 = t_2) and (c_1 &gt; c_0 = c_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTA (willingness-to-accept):</td>
<td>(t_1 = t_0 &lt; t_2) and (c_1 = c_0 &gt; c_2)</td>
</tr>
<tr>
<td>EG (equivalent gain):</td>
<td>(t_1 &lt; t_0 = t_2) and (c_1 = c_0 &gt; c_2)</td>
</tr>
<tr>
<td>EL (equivalent loss):</td>
<td>(t_1 = t_0 &lt; t_2) and (c_1 &gt; c_0 = c_2)</td>
</tr>
</tbody>
</table>

To derive choice probabilities, we introduce the following notation:

Define

\[
t = t_1 + t_2 - 2t_0 \quad \text{and} \quad c = c_1 + c_2 - 2c_0
\]

and let \(\Delta t\) and \(\Delta c\) denote the differences between \(t_1\) and \(t_2\), and \(c_1\) and \(c_2\), respectively. Note that
\[ c = \begin{cases} \ c_1 - c_0 = |\Delta c| & \text{for WTP, EL,} \\ \ c_2 - c_0 = -|\Delta c| & \text{for WTA, EG,} \end{cases} \] (5)

and

\[ t = \begin{cases} \ t_1 - t_0 = -|\Delta t| & \text{for WTP, EG} \\ \ t_2 - t_0 = |\Delta t| & \text{for WTA, EL} \end{cases} \] (6)

Disregarding errors, alternative 2 is preferred when

\[ U(t_1, c_1 | t_0, c_0) < U(t_2, c_2 | t_0, c_0). \] (7)

Using the notation defined above this is equivalent to

\[ VTT < \frac{|c| e^{\eta S(c)}}{|t| e^{\eta S(t)}} \] (8)

We shall formulate our model in terms of \( \log \) VTT and add errors to this model. (Note that this makes the errors multiplicative relative to the perceived values of time and cost, instead of additive.) We rewrite eq. (8) by taking logs, and inserting the expression for the VTT together with observation-specific logistic errors \( \varepsilon \) with location 0 and scale \( \mu \). The errors are assumed to be independent across choices and individuals, as well as independent of \( u \) and \( x \). Hence alternative 2 is chosen when

\[ \beta_0 + \beta^* x + u + \varepsilon < \log \nu + \eta_c S(c) - \eta_t S(t), \] (9)

with \( \nu = |\Delta c| / |\Delta t| \).

Together with the independence between \( u \) and \( x \), we shall need to assume that \( u \) is independent of \( S(t) \) and \( S(c) \) - or in other words: \( u \) is independent of the quadrant.

Now let \( y_{nr} \) denote the chosen alternative of individual \( n \) in situation \( r \). Further, let \( x_{nr} \) be the explanatory variables for individual \( n \) in situation \( r \), and let \( u_n \) be the random part of VTT for this individual.

The probability that \( y_{nr} = 2 \), given the value of \( u_n \) is

\[ P(y_{nr} = 2 | u_n) = \frac{1}{1 + \exp(-\mu(\log v_{nr} + \eta_c S(c) - \eta_t S(t) - \beta_0 - \beta^* x_{nr} - u_n))} \] (10)
Assume that we observe \( R \) choices for each individual, but that not all choices are to be included in the analysis. Let \((\tilde{y}_{n1}, \tilde{y}_{n2}, K, \tilde{y}_{nR})\) be the observed choices of individual \( n \), and let \( l_{nr} \) equal 1 if choice \( r \) for individual \( n \) is to be included, and 0 otherwise. Since the \( \varepsilon_{nr} \)'s are independent, the probability of the choice sequence conditional on \( u_n \) is

\[
P((\tilde{y}_{n1}, \tilde{y}_{n2}, K, \tilde{y}_{nR}) | u_n) = \prod_{r=1}^{R} \left[ 1 \{ \tilde{y}_{nr} = 1 \} \cdot (1 - P(v_{nr} = 2 | u_n)) + 1 \{ \tilde{y}_{nr} = 2 \} \cdot P(v_{nr} = 2 | u_n) \right]
\]

(11)

The likelihood contribution from individual \( n \) becomes

\[
P(\tilde{y}_{n1}, \tilde{y}_{n2}, K, \tilde{y}_{nR}) = \int f(u) P((\tilde{y}_{n1}, \tilde{y}_{n2}, K, \tilde{y}_{nR}) | u) \ du
\]

(12)

and the likelihood function is

\[
L = \prod_{n} \int f(u) P((\tilde{y}_{n1}, \tilde{y}_{n2}, K, \tilde{y}_{nR}) | u) \ du
\]

(13)

The model is estimated using simulated maximum likelihood - the choice probabilities in eq. (12) are approximated by means of simulation.

2.2. Approximating the true mixing distribution

Results from Fosgerau and Bierlaire (2005), who estimate a simple logit mixture model on similar data, suggest a lognormal distribution of VTT. This is equivalent to \( u \) having a normal distribution. We are going to test the assumption of a lognormal VTT distribution against a generalised model, by following the methodology described in Fosgerau and Bierlaire (2005).

They define a generalised mixing distribution as a transformation of the base distribution (the normal distribution in our case); the transformation being a semi-nonparametric (SNP) series using Legendre polynomials as a base for functions on the unit interval. Since this generalised distribution is defined as a transformation of the base distribution, the base distribution is a special case of the generalised and can thus be tested against this.

The generalised distribution constitutes a very flexible approximation of the true (unknown) mixing distribution. The approximation improves when including higher order Legendre polynomials (SNP terms). Fosgerau and Bierlaire (2005) show in an example that when testing the base distribution against the generalised, 1 SNP term may be too little to reject a false null, and recommend using at least 2 terms. To be safe, we use 3 SNP terms.

We test the assumption of lognormal VTT by a likelihood ratio test of the reduction from the model with SNP terms to the model with lognormal VTT. If the reduction is not accepted, we shall adopt the generalised distribution instead.

©Association for European Transport and contributors 2006
2.3. Estimating the mean of the VTT

Identification issues
Calculation of the mean VTT requires that we know the VTT distribution. Even though we are able to check the distributional assumptions as described above, this only informs us about the VTT distribution over the range where we have data. The question is whether we observe enough of the VTT distribution from the data to estimate its mean.

Say we have estimated a cumulative distribution function $F$ over some interval $[a ; b]$. Can we then determine the mean of some random variable with this distribution? If $F(b)<1$, then no: There is a positive probability mass to the right of $b$, and we don’t know anything about $F$ outside $[a ; b]$. This means that we may assume anything for $F$ outside $[a ; b]$. Hence the mean can be arbitrarily large, depending on the shape of the assumed distribution to the right of $b$.

However, if $F(a)=0$ and $F(b)=1$ we observe $F$ over its entire support, and all assumptions regarding $F$ can be verified by the data. Hence a sufficient condition for identifying the mean VTT is that we observe its distribution over the entire support.

Because $u$ is assumed to be independent of $x$ and the quadrants, identifying the VTT distribution is equivalent to identifying $u$’s distribution. Hence our concern is whether the distribution of $u$ is identified from the data.

Note that what we observe from data is not the distribution of VTT, but of the VTT with error. From a given choice, what we observe is whether (by eq. (9)) $u + \epsilon$ is greater than or smaller than $\log v + \eta_c S(c) - \eta_t S(t) - \beta_0 - \beta' x$, where $v = |\Delta c | / | \Delta t |$. Thus, $u$ is not observed separately and naturally we cannot check if it is observed over its entire support.

Note that the expression $\eta_c S(c) - \eta_t S(t) - \beta_0 - \beta' x$ contributes by extending the range over which we observe the distribution of $u + \epsilon$. Had we not included covariates, then only the variation in $\log v$ would have been able to give us identification of the distribution of $u + \epsilon$, which is not sufficient with present data (Fosgerau 2006). Since the range of $\log v$ is not sufficiently large, it is necessary to introduce covariates along with the assumption of independence between the covariates and $u, \epsilon$.

To know when $u$ is identified, we need a result from Fosgerau and Nielsen (2006). Their setting is quite similar to our model, except that they estimate the distribution of $u$ from the actual values of $v$ rather than the perceived values. However, their results can be easily transferred to our model to obtain the following criterion: The distributions of $u$ and $\epsilon$ are consistently estimated if the support of $u + \epsilon$ is contained in the support of $\log v + \eta_c S(c) - \eta_t S(t) - \beta_0 - \beta' x$; in other words, if we observe $u + \epsilon$ over its entire support.
It is easy to see that this criterion cannot be fulfilled in principle, since the support of $u + \epsilon$ is the entire real line, and we do not have an infinite data set. However it is useful to check how close we are to observing $u + \epsilon$ over its entire support as this will help detect a poorly identified VTT distribution.

From eq. (9) we see that the probability of choosing alternative 2 is the probability that

$$u + \epsilon < \log{v} + \eta_c S(c) - \eta_i S(t) - \beta_o - \beta'x$$

Hence the probability $P_{nr}^2$ of choosing alternative 2 is the distribution function of $u + \epsilon$ evaluated at $\log{v} + \eta_c S(c) - \eta_i S(t) - \beta_o - \beta'x$. Thus observing the distribution over the entire support is equivalent to the range of observed $P_{nr}^2$’s being $[0 ; 1]$. This is easily checked using the predicted values of $\{P_{nr}^2\}_{nr}$.

Note that identification relies heavily on the assumption of independence between $u$ and $x$ and between $u$ and the quadrants: In fact this allows us to assume that information about $u$’s distribution inferred from an observation from a given quadrant with a given $\beta'x$ can be combined with information from observations from a different quadrant and with different $\beta'x$.

**Derivation of the mean of the VTT**

From the definition of the VTT we see that the mean VTT is

$$E(VTT) = E(\exp(\beta_o + \beta'x))E(\exp(u))$$

The first part is estimated as an average over the sample, possibly fixing some of the $x$ to specific values. The assumption that $u \sim N(0, \sigma^2)$ implies that $\exp(u)$ has a lognormal distribution with mean

$$E(\exp(u)) = \exp(\frac{\sigma^2}{2}).$$

In the generalised model the distribution of $u$ is approximated by a flexible transformation of the normal distribution. The density of $u$ is

$$g(u) = q(\Phi(\frac{u}{\sigma})) \cdot \phi(\frac{u}{\sigma}) \cdot \frac{1}{\sigma},$$

where $\phi$ and $\Phi$ are the density and the CDF of the standard normal distribution, and the transformation $q$ is a density on the unit interval approximated by a squared linear combination of Legendre polynomials (SNP terms).
The density of $z = \frac{u}{\sigma}$ is thus

$$h(z) = q(\Phi(z)) \cdot \phi(z)$$  \hspace{1cm} (18)$$

$E(\exp(u))$ can be expressed in terms of this density as

$$E(\exp(u)) = \int_{-\infty}^{\infty} \exp(\sigma z) \cdot q(\Phi(z)) \cdot \phi(z) \, dz$$

$$= \int_{0}^{\infty} \exp(\sigma \Phi^{-1}(s)) \cdot q(s) \, ds$$  \hspace{1cm} (19)$$

where the latter equality is obtained by substituting $s = \Phi(z)$. The integral in eq. (19) can be approximated by simulation.

The mean VTT in eq. (15) is estimated by replacing $\beta$, $\beta_0$, and $\sigma$ with their maximum likelihood estimates, and combining with either eq. (16) or eq. (19), depending on the distributional assumptions on $u$.

### 3. DATA

Data are extracted from the Danish value of time study. The study was commissioned by the Danish Ministry of Transport and data design and collection was conducted by a consortium of TetraPlan, Rand Europe and Gallup. Business trips were excluded from the analysis (See Burge et al (2004) for further details on the SP design). The sample both encompasses interviews conducted via Internet and Face-to-Face Computer Assisted Personal Interviews (CAPI). The present paper shows the results from one SP design, an abstract time-cost exercise for in-vehicle time – and only considers car drivers.

All subjects in the experiment had to choose between two alternatives, described by travel time and travel cost. All choices were designed relative to a recent actual trip subjects had made. 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 generate 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. Subjects were furthermore presented with a dominated choice situation, where one alternative was both faster and cheaper than the other. The quadrant for this choice situation was random.

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 \([2 ; 200]\) Danish Crowns (DKK) per hour, using stratification to ensure that all subjects were presented with both low and high values. 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 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.  

Finally, it should be noticed that alternatives differ only with respect to time and cost, so that issues such as heterogeneous preferences for various transport modes play no role.

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. Respondents who chose the dominated alternative (the one being slower and more expensive) in the check question were excluded. Moreover, we excluded all choice situations with a dominant alternative regardless of the answer – the dominated choices are only used to identify respondents with irrational answers and contain no information of the value of time. The remaining sample of car drivers consists of 1,819 respondents (14,178 observations).

The background variables available from the interviews are socio-demographic characteristics (e.g. age, income, sex, household status etc.) together with details of the actual trip. Missing values of personal and household income are supplemented with income information from Gallup, when available. Note that 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.

Descriptive statistics of the variables are shown in Table 2 below.
### Table 2: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chosen alternative, $y$</td>
<td>1</td>
<td>1.59</td>
<td>2</td>
</tr>
<tr>
<td>Log $\nu$</td>
<td>-3.00</td>
<td>-0.54</td>
<td>1.21</td>
</tr>
<tr>
<td>$S(c)$</td>
<td>-1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>$S(t)$</td>
<td>-1</td>
<td>0.00</td>
<td>1</td>
</tr>
<tr>
<td>$\min(\mid \Delta t \mid - 20, 0)$</td>
<td>-17</td>
<td>-11.57</td>
<td>0</td>
</tr>
<tr>
<td>$S(c)$ $\min(\mid \Delta t \mid - 20, 0)$</td>
<td>-17</td>
<td>0.02</td>
<td>17</td>
</tr>
<tr>
<td>$S(t)$ $\min(\mid \Delta t \mid - 20, 0)$</td>
<td>-17</td>
<td>0.02</td>
<td>17</td>
</tr>
<tr>
<td>Log anchorVTT</td>
<td>-3.00</td>
<td>-0.55</td>
<td>1.21</td>
</tr>
<tr>
<td>Missing anchor dummy</td>
<td>0</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td>Log $C_0$</td>
<td>0</td>
<td>3.45</td>
<td>6.75</td>
</tr>
<tr>
<td>Log $t_0$</td>
<td>2.40</td>
<td>3.63</td>
<td>5.48</td>
</tr>
<tr>
<td>Share of congestion</td>
<td>0</td>
<td>0.09</td>
<td>0.68</td>
</tr>
<tr>
<td>Employer-paid dummy</td>
<td>0</td>
<td>0.06</td>
<td>1</td>
</tr>
<tr>
<td>Fixed arrival time dummy</td>
<td>0</td>
<td>0.56</td>
<td>1</td>
</tr>
<tr>
<td>Greater Copenhagen Area dummy</td>
<td>0</td>
<td>0.20</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>16</td>
<td>49.71</td>
<td>89</td>
</tr>
<tr>
<td>Age$^2$/100</td>
<td>2.56</td>
<td>26.82</td>
<td>79.21</td>
</tr>
<tr>
<td>Female dummy</td>
<td>0</td>
<td>0.42</td>
<td>1</td>
</tr>
<tr>
<td>Home owner dummy</td>
<td>0</td>
<td>0.79</td>
<td>1</td>
</tr>
<tr>
<td>Student dummy</td>
<td>0</td>
<td>0.05</td>
<td>1</td>
</tr>
<tr>
<td>Log personal income – 12</td>
<td>-1.33</td>
<td>-0.02</td>
<td>1.06</td>
</tr>
<tr>
<td>Missing personal income dummy</td>
<td>0</td>
<td>0.07</td>
<td>1</td>
</tr>
<tr>
<td>Low income group dummy</td>
<td>0</td>
<td>0.07</td>
<td>1</td>
</tr>
<tr>
<td>Log household income – 12, couples with children</td>
<td>-1.33</td>
<td>0.20</td>
<td>1.18</td>
</tr>
<tr>
<td>Missing household income dummy, couples with children</td>
<td>0</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>Log household income – 12, couples without children</td>
<td>-1.33</td>
<td>0.15</td>
<td>1.18</td>
</tr>
<tr>
<td>Missing household income dummy, couples without children</td>
<td>0</td>
<td>0.05</td>
<td>1</td>
</tr>
</tbody>
</table>

### 4. EMPIRICAL MODEL

In this section we briefly describe the covariates included in the model. There are three main groups of covariates: Variables determined by the experimental design (design variables), characteristics of the trip, and socio-economic characteristics of the respondent.

In the model, all travel times are in minutes, and all costs are in DKK. The unit of $\nu$ is DKK per minute.
4.1. Design variables

The most important factor in the experimental design is the size of the travel time difference ($|\Delta t|$). To capture the effect of VTT increasing with $|\Delta t|$ up to a travel time difference of 20 minutes, we include the term $\min(|\Delta t| - 20, 0)$. We allow the effect of $\Delta t$ on VTT to vary across quadrants by adding interaction terms between $\min(|\Delta t| - 20, 0)$ and the quadrant.

Another relevant factor is the potential presence of anchoring effects: In a contingent valuation study of the willingness to pay for public goods, Green et al. (1998) find that WTP is considerably affected by the price suggested to respondents. Their conclusion is that the mean WTP is pulled towards the anchor value - this is referred to as anchoring. Van de Kaa (2005) finds significant evidence of anchoring in his analysis of the data from the Dutch Value of Time Study. It is therefore highly relevant to investigate the presence of similar anchoring effects in our data.

If an anchoring effect is present we should observe that respondents suggested a higher time value in their first choice tend to have larger time values. To control for this effect we include an anchor variable that equals the value of $\log V$ from the first choice situation. We also include a dummy for missing anchor values (this happens when $V$ is zero in the first choice situation, because of rounding of the cost attributes).

4.2. Trip characteristics

We will investigate the contribution to the VTT of a number of characteristics. These variables are briefly described in the following.

**Reference travel time and cost.** The reference is taken as the current trip around which the hypothetical trips are pivoted. We include the log of reference travel time (in minutes) and the log of reference travel cost (in DKK). **Who pays for the trip.** We include a dummy for employer-paid trips and expect this to have a positive effect on VTT, since the marginal disutility of travel cost will be lower for individual who do not pay themselves. **Level of congestion.** We include the share of driving time due to congestion in total travel time and expect the VTT to increase with the level of congestion. Note that the share of congestion time is fixed for each respondent. Therefore we cannot separate the direct effect of congestion on the VTT from a selection effect whereby respondents with high VTT tend to experience more congestion. The latter is a likely effect, as congestion is concentrated around Copenhagen where incomes are higher. **Arrival time flexibility.** We include a dummy for respondents who have to arrive at their destination at a fixed time or with some flexibility. (The base is “The arrival time is of no importance”.)
4.3. Socio-economic characteristics

We will also investigate the contribution to the VTT of a range of socio-economic characteristics.

**Income.** We include the log of after tax personal income (demeaned) and dummies for lowest and highest income groups. The after tax income is computed using average tax rates and assuming no interest payment deductions etc. We also include a dummy for missing income information. For couples we furthermore include the after tax household income together with missing income dummies. **Age.** We expect VTT to increase with age up to a certain point, and then decrease. We therefore include both age (in years) and age squared divided by 100. **Gender.** We include a dummy for females. **Geography.** We include a dummy for the Greater Copenhagen Area. **Home ownership.** We include a dummy for whether the respondent owns his house/residence. **Student.** We include dummy for students.

5. RESULTS

Estimation is carried out in Biogeme (Bierlaire 2005). Biogeme allows for explicit estimation of the error scale $\mu$, as well as for the generalised mixing density described in previous sections. We use 300 Halton draws to simulate each likelihood contribution and note that this is sufficient to achieve stable results. Table 2 shows the estimated parameters. The following section discusses the parameter estimates.

5.1. Quadrant effects and covariates

The estimated quadrant effects are very significant, and have the expected sign indicating loss aversion. $\eta_t$ is about twice the size of $\eta_c$, which means that loss aversion is stronger for time than for travel cost. This relationship is in line with common findings of reference-dependence.

The difference between $\eta_c$ and $\eta_t$ defines the relative ordering of the perceived value of $v$ in the four quadrants. Since $0 < \eta_t < \eta_c$, for a given value of $v$ the perceived value is highest in the WTP quadrant, and lowest in the WTA quadrant. The perceived value in the EG quadrant is lower than in the EL quadrant, with the actual value in between. Hence respondents perceive an offered price of time as higher when asked to pay for time, and lower when asked to give up time. If not controlling for this effect, the researcher would observe higher VTT values in the WTA quadrant than in the WTP quadrant.

VTT increases with the size of the change in travel time, such that an additional minute added to a time change less than 20 minutes increases VTT by 5.6%. Similar effects (though not necessarily of the same scale) have been observed in the Dutch and British Value of Time Study (e.g., Gunn 2001 and Van de Kaa 2005). Note however, that the sign of the effect is not consistent with utility theory or with

©Association for European Transport and contributors 2006
prospect theory, as they predict the opposite sign, i.e. diminishing marginal sensitivity. The effect here may be interpreted as an editing effect (Kahneman & Tversky 1979), related to pre-processing of the choice situation by respondent prior to evaluation. In plain words it means that small time savings tend to be disregarded since it is not worth the mental effort to find out what they are worth in money.

We find no sign of an anchoring effect (both anchor terms are very insignificant). This is probably due to the design of the survey, as time values are not explicitly stated. This is a very convenient finding, as it would have been difficult to defend an estimate of the VTT if it was dependent on the SP design.

As expected, the VTT increases with the level of congestion. An increase in the congested share of driving time of 25 percentage points increases VTT with 26%, while an increase of 50 percentage points increases VTT with 60%. Also as expected, respondents on employer-paid trips have higher VTT (almost twice as high) than as respondents who pay for themselves.

The income variables are the most prominent among the socio-economic characteristics. The income elasticity is around 40% both for personal and household income, and the terms are very significant. Please note that this cannot be taken as a general income elasticity, since it is conditional on the other covariates which also depend on income: particularly, the congestion share, the Greater Copenhagen Area dummy, home owner status, and employment status.

Other things being equal, being a student decreases the VTT by almost 50%, and women’s VTT is about 15% lower than men’s. For respondents living in the Greater Copenhagen Area the VTT is generally 35% higher than for the rest of the population. The home ownership dummy and the missing income dummies are not significant. The age variables do not have the expected sign, as VTT decreases with age in the range 16-89 years. We expect this is because age is correlated with other variables as income and home ownership.
Table 3: Estimated parameters of model with 3 SNP terms

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Std.error</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>2.2307</td>
<td>0.5513</td>
<td>***</td>
</tr>
<tr>
<td>$\eta_c$</td>
<td>0.2617</td>
<td>0.0409</td>
<td>***</td>
</tr>
<tr>
<td>$\eta_t$</td>
<td>0.4931</td>
<td>0.0425</td>
<td>***</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.8821</td>
<td>0.2739</td>
<td>***</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1.1989</td>
<td>0.0319</td>
<td>***</td>
</tr>
<tr>
<td>$\beta$’s of design variables:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min(</td>
<td>Δt</td>
<td>- 20, 0)</td>
<td>0.0546</td>
</tr>
<tr>
<td>S(c) min(</td>
<td>Δt</td>
<td>- 20, 0)</td>
<td>-0.0067</td>
</tr>
<tr>
<td>S(t) min(</td>
<td>Δt</td>
<td>- 20, 0)</td>
<td>-0.0103</td>
</tr>
<tr>
<td>Anchor VTT</td>
<td>0.0031</td>
<td>0.0376</td>
<td></td>
</tr>
<tr>
<td>Missing anchor</td>
<td>0.2075</td>
<td>0.2191</td>
<td></td>
</tr>
<tr>
<td>$\beta$’s of trip characteristics:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of reference cost ( $c_0$ )</td>
<td>0.6726</td>
<td>0.0720</td>
<td>***</td>
</tr>
<tr>
<td>Log of reference time ( $t_0$ )</td>
<td>-0.8219</td>
<td>0.1039</td>
<td>***</td>
</tr>
<tr>
<td>Share of congestion</td>
<td>0.9159</td>
<td>0.2568</td>
<td>***</td>
</tr>
<tr>
<td>Employer-paid trip</td>
<td>0.6437</td>
<td>0.1491</td>
<td>***</td>
</tr>
<tr>
<td>Fixed arrival time</td>
<td>0.2572</td>
<td>0.0684</td>
<td>***</td>
</tr>
<tr>
<td>$\beta$’s of socio-economic variables:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater Copenhagen Area</td>
<td>0.3035</td>
<td>0.0862</td>
<td>***</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0297</td>
<td>0.0179</td>
<td>*</td>
</tr>
<tr>
<td>Age^2/100</td>
<td>0.0038</td>
<td>0.0178</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.1662</td>
<td>0.0704</td>
<td>**</td>
</tr>
<tr>
<td>Home owner</td>
<td>0.1248</td>
<td>0.0888</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>-0.6058</td>
<td>0.1887</td>
<td>***</td>
</tr>
<tr>
<td>Log personal income - 12</td>
<td>0.4089</td>
<td>0.1155</td>
<td>***</td>
</tr>
<tr>
<td>Missing personal income</td>
<td>0.0329</td>
<td>0.1977</td>
<td></td>
</tr>
<tr>
<td>Low income group</td>
<td>0.3069</td>
<td>0.1985</td>
<td></td>
</tr>
<tr>
<td>Log household income – 12, couples with children</td>
<td>0.3974</td>
<td>0.1297</td>
<td>***</td>
</tr>
<tr>
<td>Missing household income, couples with children</td>
<td>0.1858</td>
<td>0.2695</td>
<td>***</td>
</tr>
<tr>
<td>Log household income – 12, couples without children</td>
<td>0.3882</td>
<td>0.1286</td>
<td>***</td>
</tr>
<tr>
<td>Missing household income, couples without children</td>
<td>0.0261</td>
<td>0.2141</td>
<td>***</td>
</tr>
<tr>
<td>Parameter of SNP term 1</td>
<td>-0.4184</td>
<td>0.0855</td>
<td>***</td>
</tr>
<tr>
<td>Parameter of SNP term 2</td>
<td>-0.0871</td>
<td>0.1183</td>
<td>***</td>
</tr>
<tr>
<td>Parameter of SNP term 3</td>
<td>0.2284</td>
<td>0.0738</td>
<td>***</td>
</tr>
</tbody>
</table>

Note: *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. For the scale $\mu$, the significance indicator is with respect to the test of $\mu = 1$. ©Association for European Transport and contributors 2006
5.2. Distribution of the VTT

In the estimated model the mixing density is given by eq. (17), with the transformation $q$ being approximated by a squared series of Legendre polynomials of degree 1, 2, and 3 and three corresponding estimated SNP terms. If the three SNP terms are all zero, the transformation $q$ is identically equal to one, and the density in eq. (17) becomes the density of $N(0,\sigma^2)$. Hence if the SNP terms are all insignificant, the mixing distribution can be assumed to be normal (i.e. the VTT is lognormal).

Two of the three SNP terms are significantly different from zero at the 1% level, indicating that the VTT distribution is not lognormal. We test this with a likelihood ratio test against a model with no SNP terms (i.e. with lognormal VTT). The log likelihood values of the two models are given in Table 4 below.

The significance probability (p-value) of the test of reduction from the model with 3 SNP terms to the model with lognormal VTT is 0.001%. Hence the hypothesis of lognormal VTT is rejected.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log likelihood</th>
<th>Observations</th>
<th>Individuals</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model with lognormal VTT</td>
<td>-7302.64</td>
<td>14178</td>
<td>1819</td>
<td>28</td>
</tr>
<tr>
<td>Model with 3 SNP terms</td>
<td>-7289.64</td>
<td>14178</td>
<td>1819</td>
<td>31</td>
</tr>
<tr>
<td>Model with 4 SNP terms</td>
<td>-7287.74</td>
<td>14178</td>
<td>1819</td>
<td>32</td>
</tr>
</tbody>
</table>

Since the highest order SNP term in the approximation is significant at the 1% level, it may be that higher order terms are needed to approximate the mixing density. We test this with a likelihood ratio test against a model with 4 SNP terms. The log likelihood of this model is given in Table 4. The reduction from 4 to 3 SNP terms is accepted with a significance probability of 5.1%, and we conclude that 3 SNP terms is sufficient.

The mixing density with 3 SNP terms, is shown in Figure 1. For comparison, Figure 1 also shows the estimated normal mixing density from the model without SNP terms (note that its location has been transformed to correct for the difference in $\beta_0$).
The SNP density has a longer tail to the right and a larger mode than the normal mixing density from the model with no SNP terms.

Figure 1: Approximated mixing density (3 SNP terms)

Given the estimated density of $u$ we perform the support check described in section 2.4.1. We compute the range of the observed quantiles of the distribution of $u + \varepsilon$, i.e. the predicted probabilities $\{\hat{P}_{nr}^2\}_{nr}$ of choosing alternative 2. The probabilities are obtained by integrating the conditional probabilities from eq. (10) with respect to the distribution of $u$. This is done numerically; using random draws from $u$'s distribution. The range of $\{\hat{P}_{nr}^2\}_{nr}$ is [0.0115 ; 0.9883]. Hence we observe the $u + \varepsilon$ distribution between its 1.15% and 98.83% quantiles. We shall not worry about the left tail, since its exact shape is less important to the mean estimate (because the VTT distribution is bounded from below by zero). The unobserved right tail only covers a little more than 1% of the probability mass of the $u + \varepsilon$ distribution, which must be considered good. However, identification could be improved by expanding the range of $\nu$, and this should be taken into account when designing similar SP experiments in the future.
6. CALCULATION OF THE MEAN VALUE OF TRAVEL TIME

6.1. Deciding on covariates

For computation of the mean VTT, we must decide on the values of the covariates that we will apply. Mostly these are the same as those in the sample. But those variables that are given by the stated choice design deserve special consideration.

Based on Fosgerau & De Borger (2006) we fix the loss aversion dummies at 0. Thus the values that we compute are averages of the WTP and the WTA and also of the equivalent gain and the equivalent loss.

The main issue is what travel time saving to apply. For time changes in the range 3 to 20 minutes, the unit value of travel time increases with the size of the time change. Hence a 10-minutes time saving is worth more than two 5-minutes time savings. This is not unexpected, as the same pattern was found in the Dutch and British Value of Travel Time studies (e.g., Gunn (2001 and Van de Kaa 2005). When evaluating transport investments, however, a lower unit VTT for small time changes is not appropriate, and a constant unit value should be assigned to all time savings (Zhang et al. 2004). Hence the mean VTT we wish to calculate must not depend on the size of the time change. This means we have to choose a level of time change for which to evaluate the mean VTT.

This choice is a crucial point in the mean calculation: Since the effect of the size of the time change is quite large, different choices of time saving sizes will produce very different mean VTT values. In accordance with the British Value of Travel Time study (Mackie et al. 2001 and 2003), we shall base our choice on the assumption that the observed lower unit VTT for small time changes is not a “true feature” of the value of travel time, but is caused by the artificial nature of the experimental design, i.e. from editing in the sense of Kahneman & Tversky (1979).

The level of time change for which we wish to evaluate mean VTT should be large enough to eliminate this “disturbance”. Following the logic of our model, the natural choice would be 20 minutes, since this is by definition the time change over which VTT remains constant. However, it must be recognised that the choice of this threshold depends on the experimental design, as the share of observations with a time change higher than 20 is low – around 7%. Had we had more observations with larger time changes, we might have selected another threshold and the coefficient of the $\Delta t$ term would be different. Hence, it may seem that the level of time change for mean calculation is somewhat arbitrary, but as there is no better argument for choosing another level, we suggest 20 minutes. This approach is similar to the
one followed in the British Value of Travel Time Study (Mackie et al. 2001 and 2003).

Since we are only interested in the VTT for non-business trips, the “employer-paid-trip” dummy is set to zero when calculating the mean VTT.

To compute the mean VTT from eq. (15), we first calculate \( \exp(\beta' x_n + \beta_0) \) for each individual, fixing \( |\Delta t| \) to 20 and the “employer-paid-trip” dummy to zero. These values are averaged over the sample to obtain \( E(\exp(\beta' x + \beta_0)) \). \( E(\exp(u)) \) is computed by simulation, using the estimated SNP coefficients.

In addition, we compute quantiles and truncated means of the VTT distribution. This is done numerically by simulating a large number of VTT values for each individual (each value of \( \exp(\beta' x + \beta_0) \)), weighting these values according to the distribution of \( u \), and calculating the empirical distribution function. The resulting cumulative distribution function of VTT is shown in Figure 2.

**Figure 2: Approximated VTT cumulative distribution function (3SNP)**

6.2. Results
As mentioned previously, to identify the VTT distribution we use the assumption that \( u \) is independent of \( x \) and the choice quadrants. In effect, we are using responses for large \( v \)’s from respondents who are expected (by their covariate vector \( x \) ) to have low VTT, to infer the behaviour of
respondents who (by their covariate vector $x$) are expected to have high VTT outside the range of $v$.

This implies that, even though range of $v$ in the experiment is [2 ; 200] DKK per hour, the independence assumption causes some individuals to have an expected VTT that is much higher than 200 DKK/hour. This is unavoidable, since a significant share of people agree to pay 200 DKK/hour to save time.

As a consequence, we report in Table 5 the mean VTT, as well as different quantiles of the VTT distribution and also the mean of the distribution truncated at 200 DKK per hour and at the 99% quantile.

<table>
<thead>
<tr>
<th>Table 5: Estimated mean VTT</th>
<th>DKK per hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>38</td>
</tr>
<tr>
<td>Mean</td>
<td>268</td>
</tr>
<tr>
<td>99% quantile</td>
<td>4160</td>
</tr>
<tr>
<td>Truncated means</td>
<td></td>
</tr>
<tr>
<td>At 200 DKK</td>
<td>64</td>
</tr>
<tr>
<td>At 99% quantile</td>
<td>147</td>
</tr>
</tbody>
</table>

We notice that the mean is very high and higher than the maximum VTT in the experiment of 200 DKK per hour. This is worrying for two reasons: First, because the estimated VTT is much higher than expected when designing the experiment, and second and more importantly, because it implies that the identification of the VTT distribution relies crucially on the assumption of independence between $x$ and $u$.

A possible further explanation for the high VTT is that for right-skewed distributions the mean increases with the variance. Hence the mean VTT increases with $\sigma$, and with the variation in $\beta^'x$. Thus the mean estimate is very sensitive to choice of covariates and the way they are assumed to affect the VTT.

7. CONCLUSION

The presented modelling approach is a straightforward alternative to the standard model in eq. (1). The direct parametrisation of VTT allows us to handle the VTT distribution directly instead of having to obtain it as a ratio of
two distributions. This not only facilitates the computation of the mean, it also makes it easy to check the identification of the distribution and test it against generalised distributions. We find that generalising the lognormal VTT distribution improves the model fit considerably.

The reference-dependent utility function incorporates preference asymmetry to allow for loss aversion. Our findings indicate significant loss aversion with respect to travel time and cost, respectively. The separate handling of such perception “errors” implies that they are implicitly controlled for when evaluating the mean VTT.

We are able to control for anchoring in a simple way and find no significant effect. Moreover, we are able to obtain significant and sensible estimates of the effect of a large number of determinants of the individual VTT, something which is often hard to achieve.

In these respects we therefore conclude that our proposed approach is successful. It has also revealed some problems for the calculation of the mean VTT.

We have noted that the range of the design time values $\nu$ is not large enough to allow us to observe the distribution of $u + \varepsilon$ over the whole of its support. Therefore we were forced to introduce covariates along with an independence assumption in order to extend the range over which we observe the distribution of $u + \varepsilon$. This resulted however in a large number of individuals having expected VTT above the maximum value of $\nu$ in the SP design. This problem is revealed by the fact that truncation of the individual VTT at the maximum $\nu$ decreases the estimated mean by more than a factor 4.

This problem is not specific to our approach. The fact that the variation in $\nu$ is not sufficient to identify the whole VTT distribution is equally true for the classical model in eq. (1). It is rather the case that our approach makes the problem clearly visible which must be viewed as an advantage.

A possible remedy might be to relax the assumption of independence between the observed and unobserved heterogeneity parts of the individual VTT. Our intuition is that this is unlikely to solve the problem as it brings back the original issue that the range of bid time values in $\nu$ is too limited. For future studies it might alleviate the problem if the range of $\nu$ was extended. The problem could remain if a significant share of respondents continues to agree to pay even very high prices in order to save time.

Another remedy is to give up the ambition of revealing the whole VTT distribution. Then one might be satisfied with knowing a range of quantiles including of course the median. This is much easier to achieve.

Bibliography


Notes
2 1 DKK = 7.5 EUR
3 For more general value functions, see De Borger & Fosgerau (2006).
4 In some cases, rounding caused the cost difference to be zero. These observations are omitted from the analysis.
5 In the experiment $|\Delta|$, takes the values {3, 5, 10, 15, 20, 30, 45, and 60}. For each T in this set, we have estimated a model with the term $\min(|\Delta|-T,0)$. The model with T=20 has the highest likelihood.