# Catching the Tail: Empirical Identification of the Distribution of the Value of Travel Time

Börjesson, Maria and Fosgerau, Mogens and Algers, Staffan

Centre for Transport Studies, KTH, Sweden, Technical University of Denmark

2012

# Catching the Tail: Empirical Identification of the Distribution of the Value of Travel Time

Maria Börjesson[1] Mogens Fosgerau[2] Staffan Algers[1]

Recent methodological advances in discrete choice analysis in combination with certain stated choice experiments have allowed researchers to check empirically the identification of the distribution of latent variables such as the value of travel time (VTT). Lack of identification is likely to be common and the consequences are severe. E.g., the Danish value of time study found the 15% right tail of the VTT distribution to be unidentified, making it impossible to estimate the mean VTT without resorting to strong assumptions with equally strong impact on the resulting estimate. This paper analyses data generated from a similar choice experiment undertaken in Sweden during 2007-2008 in which the range of trade-off values between time and money was significantly increased relative to the Danish experiment. The results show that this change allowed empirical identification of effectively the entire VTT distribution. In addition to informing the design of future choice experiments, the results are also of interest as a validity test of the stated choice methodology. Failure in identifying the right tail of the VTT would have made it difficult to maintain that respondents' behaviour is consistent with utility maximization in the sense intended by the experimenter.

[1] Centre for Transport Studies

Royal Institute of Technology

100 44 Stockholm, Sweden

[2] Department of Transport

Technical University of Denmark

# 1 Introduction

Numerous stated choice studies have found a significant proportion of apparent non-traders. We demonstrate empirically that in our case, a value of time study, non-trading can be virtually eliminated by providing a sufficient range of trade-offs in the stated choice design. This finding is of vital practical importance as the effect on the estimated willingness to pay distribution, and hence on the estimated mean value, often depend strongly on the assumptions regarding apparent non-traders.

More specifically, we consider the empirical identification of the distribution of the value of travel time (VTT) from discrete choice data. Essentially, the required data consist of observations of individual choices between alternatives differing in the time and cost dimensions. With binary choices, one trip is faster but more expensive than the other, such that a price of travel time is implicit in each choice. The implicit price of time is the trade-off value or the offered 'bid'. Rational respondents with a VTT that is lower than the bid will choose the cheaper and slower alternative; otherwise they will choose the other alternative. So in making their choice, respondents reveal whether their VTT is larger or smaller than the bid. Observation of many respondents for the same bid then reveals the share of respondents with VTT less than the bid. But this is just the value of the cumulative distribution of the VTT evaluated at the bid. Data for a range of bids then allow the analyst to trace out the VTT distribution.

From this perspective, it is clear that the data do not reveal the VTT distribution outside the range of bids. Estimates of, e.g., the mean VTT hence have to rely on additional identifying assumptions if the range of bids is not sufficiently large. Such assumptions are hard to verify and the impact on the results can be extreme. This is shown in Fosgerau (2006), who analysed discrete choice data with a maximum bid of 25 EUR/h. About 13% of respondents accepted this bid, indicating that their VTT was larger than 25 EUR/h. In other words, these respondents were non-traders with this experimental design (we include seemingly lexicographic behaviour in the term non-traders). Fosgerau (2006) shows that fitting distributions to these data leads to estimates of the mean VTT that can be arbitrarily high.

The need to apply a stated choice experiment design covering the range of preferences when eliciting VTT distributions has been recognized since many years. Fowkes and Wardman (1988) explicitly suggest inclusion of some choices implying "implausible high or low boundary values of time" in order not to erroneously omit respondents using lexicographic decision rules. This advice is supported by explicit studies of lexicographic behaviour. Killi et al. (2007) find that seeming lexicographic behaviour is due primarily to steep indifference curves in combination with insufficient attribute scale extension. Similar evidence on seeming lexicographic behaviour can be found in other application areas. Cairns and van der Pol (2004) use an adaptive design in a health related experiment to show that non-trading behaviour can be virtually eliminated by adjusting trade-offs presented to respondents in the light of their previous answer. While this procedure introduces endogeneity which must be handled in order not to bias results, their evidence does indicate that non-trading is, also in their case, a genuine expression of preferences. Ryan et al. (2004) find (also in a health related experiment) that individuals who appear to adopt non-compensatory decision making strategies do so because they rate particular attributes very highly and not because they try to simplify the task.

In a recent paper, Hess et al. (2010) discuss the incidence of non-trading and lexicographic behaviour and arrive at the recommendation that such observations should be discarded. This advice is followed by Potoglou et al. (2010) and also Ahern and Tapley (2008) discard non-traders in analysis of stated choice experiments. Abrantes and Wardman (2011) note in a meta-study of 226 VTT studies that discarding non-traders is a common approach in VTT practice. Wardman and Ibáñez (2011) find a considerable amount of non-trading in a UK VTT stated choice data set and note that the model fit improves when these are discarded. They suggest that the lower level of non-trading found in a similar US data could be due to the fact that this data collection was computer assisted and not relying on pen and paper.

The results of our paper contradict the recommendation of Hess et al. and much of the above refereed practice. This study indicates that in the Danish data, also analysed by Hess et al., the lexicographic behaviour is a genuine expression of preferences and it would hence be a mistake to discard such observations. Lancsar and Louviere (2006) comment on the deletion of seemingly non-trading observations by saying "it seems somewhat paradoxical, if not paternal, to design and implement discrete choice experiments because one is interested in consumer preferences, but if the results do not conform to researchers' a priori expectations of how preferences 'should' behave, to then impose one's own preferences on the data by deleting such responses".

With alternatives described in terms of travel time and cost, the maintained theory holds that respondents make utility maximising choices governed at the margin by the marginal rate of substitution between time and money, i.e. the VTT. Non-trading occurs simply when the maximum bid is not sufficiently high and the share of non-traders decreases as the maximum bid is increased. This observation provides an opportunity for testing the validity of choice experiments. If the share of non-traders did not decrease as the maximum bid was increased then it would be hard to maintain that choices are governed solely by the VTT.[1] The present paper analyses new data from an experiment carried out in Sweden during 2008, comprising the car mode as well as long and short distance bus and train modes. The experiment design was essentially the same as that used in Denmark but with the bid range extended up to 50 EUR/h, about twice the maximum in Fosgerau (2006). The results show first that most of the VTT distribution is now identified. By the above discussion, this also indicates that the methodology passes the validity test implicit in the extension of the bid range. Moreover, we observe about 90 percent of the VTT distribution below 25 EUR/h, in approximate coherence with the Danish findings. Hence, the non-trading found in the Danish data seem to be a genuine expression of VTT values which mostly lie in the interval 25-50 EUR/h.[2]

Many VTT studies have estimated the VTT using standard logit models, although we now know that the VTT is very heterogeneous in the population. After simulation techniques made the mixed logit model easier to estimate, attempts to capture the heterogeneity of VTT have become frequent. Mixed models have normally been estimated in preference space (examples are Hess el al. (2005), Cirillo and Axhausen (2006), Brownstone and Small (2005) and Hensher (2006)), estimating marginal utilities of travel time and travel cost. The mean value of time is then computed as the mean of the ratio between

[1] But not impossible. In principle, it may be that there is an empty interval with seeming non-traders all having VTT above this interval. Such gaps in the distribution of VTT are however quite implausible.
[2] The findings of the Norwegian value of time study are consistent with the Swedish and the Danish findings (Ramjerdi et al, 2010). They used bid ranges similar to those in the Swedish study and found the same, low, frequency of non-trading as found in the Swedish study.

the marginal utilities of travel time and travel cost. Still, it is unclear how this ratio relates to the VTT distribution. When using a randomly distributed cost parameter a mean value of time is often not even defined. The model specifications are rarely tested. In particular, they do not report any check of range in their data.

Inspired by Beesley (1965) and Cameron and James (1987), Fosgerau (2006) proposed an empirical model that estimates the VTT directly in log(*bid*) space. By estimation in *log(bid)* space, in contrast to the more traditional models estimated in the preference space, the problem of computing the ratio between two distributions is avoided. Fosgerau (2007) uses nonparametric techniques to show that the *log(bid)* space model describes the Danish data better than a model estimated in the preference space.

The present paper uses the techniques developed in Fosgerau (2006) and Fosgerau (2007). Thus we do not claim novelty of our econometrical techniques. The contribution of this paper is to demonstrate that the missing tail can be captured by increasing the range of the trade-off values.

Nonparametric techniques are first used to estimate the VTT distribution. It would be possible to stop here, as the nonparametric model provides already an estimate of the distribution of VTT. However, for reasons discussed in the paper, notably the presence of reference dependence, it is desirable to estimate also a parametric model in order to control for various factors relating to the design of the choice experiment. Next, we therefore apply nonparametric techniques to the data to support the choice of parametric model. It is found that a model in *log(bid)* space describes the Swedish data better than a model in preference space, which was also the case with the Danish data.

Section 2 describes the theory and model specification for the nonparametric and parametric models that we use. Section 3 describes the data, including the efforts made to extend the coverage over the support of the VTT distribution. In section 4 we present estimation results: we first apply nonparametric regression techniques to explore the properties of our data; we proceed to choose the parametric model; estimate the parametric model and simulate the VTT distribution. Section 5 concludes.

# 2 Methodology

## 2.1 Reference dependence

The stated choice design for the present study involves binary choices between alternatives constructed as variations around the respondent's current trip, which is treated as a reference point. From the reference trip it is possibly to construct four types of choices, corresponding to the quadrants in Figure 1. The willingness to pay (*WTP*) choice is a choice between the reference trip and an alternative that is faster and more expensive. The willingness-to-accept (*WTA*) choice is the exact opposite, comparing the reference trip to an alternative that is slower but less expensive. The equivalent gain (*EG*) choice is a choice between an alternative with the reference time but cheaper and an alternative with the reference cost but faster. The equivalent loss (*EL*) choice is the exact opposite, including an alternative with reference cost but slower and an alternative with reference time but more expensive.

In Figure 1, the straight lines represent indifference curves, and their slopes hence represent the value of time in each quadrant. $WTP < WTA$ follows immediately from declining marginal utility of money (Randall and Stoll, 1980), but the gap between *WTP* and *WTA* that is found in experiments is often larger than what can be explained by the income effect (Horowitz and McConnell, 2002). Past empirical evidence in the context of VTT indicates that the mean VTT measured for these four types of choices can be very different (De Borger and Fosgerau, 2008).

Prospect theory (Tversky and Kahneman, 1991) holds that the slope of the indifference curves through a point depends on the reference from which it is evaluated, and that kinks occur at the reference point. Moreover, losses matter more than equal sized gains – this effect is called loss aversion. Bateman et al. (1997) show that loss aversion implies WTA>WTP. Moreover, the two other measures, $EL$ and $EG$, should be in between, but their relative size cannot be determined a priori. De Borger and Fosgerau (2008) show moreover that $(WTP * WTA)^{1/2} = (EG * EL)^{1/2}$ under certain assumptions. They test this equality empirically, using the Danish VTT data, and find that it holds in a statistical test of high power. Thus, in the Danish data, the differences in VTT between the quadrants in Figure 1 are consistent with loss aversion.

The WTP-WTA gap leads to the question of which value to use as the VTT when the reference point has lost its meaning. Such situations include, e.g., traffic models, welfare economic evaluations etc., see De Borger and Fosgerau (2008). What is sought is a VTT that can be applied independently of any reference point. We assume that such a reference-free value of time exists. The question is then how it can be identified from estimates of the four valuation measures WTP, WTA, EG and EL. Under the assumption that losses are over-weighted as much as gains are underweighted, De Borger and Fosgerau (2008) show that reference-dependence and the reference-free value of time can be derived as the geometrical average of the WTP and WTA choices or equivalently of the EG and EL choices, i.e. the reference-free VTT equals $(WTP * WTA)^{1/2} = (EG * EL)^{1/2}$. The reference-free value of time is depicted in Figure 1 as the slope of the straight line in grey, drawn through the reference point.

**Figure 1: The four quadrants. The reference-free value of time is drawn as a dashed line.**

## 2.2 Some nonparametric techniques

Nonparametric techniques enable us to explore the properties of data while imposing only minimal assumptions. This paper uses local constant regression and local logit regression. Local constant regression is used to estimate the VTT distribution. Local logit regression is used to plot the response surface describing the expected choice as a function of independent variables. It is then possible to check whether the data are consistent with the response surface implied by various model specifications.

*Local constant regression*

Local constant regression estimates a nonparametric function in one dimension as a weighted average of a dependent variable $y$ in the neighbourhood of the independent variable $x_0$:

$$E(y|x_0) = \sum_n k_n y_n,$$

where $k_n$ is a local weight around the point $x_0$:

$$k_n = \frac{K(\frac{x_n - x_0}{h})}{\sum_m K(\frac{x_m - x_0}{h})} \quad (1)$$

The bandwidth $h$ determines the size of the neighbourhood over which to average. In this paper, $K$ is taken to be a standard normal density kernel or a triangular density kernel.

Consider a binary route choice for some transport mode between unlabelled alternatives, described by the in-vehicle travel times $t_0$ and $t_1$ and the travel costs $c_0$ and $c_1$. Denote the individual specific VTT as $W$. Let $V = -(c_0 - c_1)/(t_0 - t_1) = -\Delta c/\Delta t$ be the trade-off price of travel time, or *'bid'* implicit in the choice situation. The data have the property that $V>0$ for all observations. Let the choice indicator $y$ be defined with the convention that $y = 1$ if the slow and cheap alternative is chosen and $y = 0$ otherwise.

Even though $y$ is binary, we may write $y = E(y|V) + \eta$, where η is the residual in a regression of $y$ on $V$. Then by definition $\mathrm{E}(\eta|\mathrm{V}) = 0$. Moreover, $\mathrm{E}(y|V) = \mathrm{P}(\mathrm{W} < V) = \mathrm{F_W}(\mathrm{V})$ such that the cumulative distribution of $W$ at points $V$ may be estimated by regressing $y$ on $V$. Such a regression may be carried out using, e.g., local constant regression such that no particular functional form is imposed on $F_W(V)$.

*Local logit*

Local constant regression fits a local constant to the data. Local logit regression improves efficiency by utilising the information that $y$ can only take values 0 and 1. The higher efficiency is more important when more independent variables are used. At points $x_n$ near a point $x_0$ it is assumed that $y$ follows a local logit model:

$$P(y = 1|x_n) = \frac{1}{1+e^{\alpha+\gamma(x_n-x_0)}} \ ,$$

where $x$ is a vector of independent variables and $\alpha$ and $\gamma$ are parameters to be estimated. Parameters are estimated at each point $x_0$ in a grid by maximizing the local likelihood:

$$L(\alpha,\beta) = \sum_n k_n[(y_n P(y_n = 1|x_n) + (1 - y_n)P(y_n = 0|x_n))]$$

where $h$ is the bandwidth used to define the neighbourhood of $x_0$ and $k_n$ is the local weight defined as above by (1).

Given the parameter estimates at $x_0$, we can estimate $\hat{P}(y = 1|x_0) = \frac{1}{1+e^{\hat{\alpha}}}$, where $\hat{P}$ and $\hat{\alpha}$ denote the nonparametric estimates of $P$ and $\alpha$ at $x_0$.

To specify a model defined in bid space, assume that there is a distribution of VTT in the population. Each individual chooses the slow alternative if his VTT is smaller than the bid $V$, which leads to the model:

$$P(y = 1|V) = P(W < V) = F_W(V) \qquad (2)$$

We denote this model 1. In this model we assume that all random variation between observations arises from variation in the VTT, $W$. A second model 2 is defined in preference space, including a time and a cost parameter:

$$P(y = 1|\Delta c, \Delta t) = P(0 < a\Delta c + b\Delta t + \epsilon) = 1 - F_\epsilon(-a\Delta c - b\Delta t), \qquad (3)$$

where $F_\epsilon$ is the cumulative distribution function of ϵ. ϵ is assumed to be independent of $(\Delta c, \Delta t)$ and $a$ and $b$ are assumed to be fixed. Thus, this model assumes that none of the random variation is attributed to variation in VTT, but only to random error. The difference between model 1 and model 2 lies completely in how the randomness is interpreted and introduced in the models, i.e. in the assumptions of the properties of the data. To determine which assumptions about the random error components that are most appropriate for a given data set it is informative to visualize the response surfaces.

To visualize the response surface implied by our data, we may estimate the expected choice $E(y|\Delta t, \Delta c)$ as an unparametrized function of $\Delta t$ and $\Delta c$ using local logit estimation and present the result as constant probability contour lines in $(\Delta t,\Delta c)$-space. If model 1 is the true model, the constant probability contour lines would be straight lines, fanning out from the origin. We see that from deriving the function for the constant probability contours using (2):

$$P(y=1)=q \iff -\Delta c = F_W^{-1}(q)\,\Delta t, \tag{4}$$

where $q$ is a given probability to reject the bid. If model 2 is the true model, the constant probability contour lines would be parallel in the $(\Delta t,\Delta c)$-space. This is apparent if we derive the functions for the constant probability contour (3).

$$P(y=1)=q \iff \Delta c = -\frac{F_\epsilon^{-1}(1-q)}{a} - \left(\frac{b}{a}\right)\Delta t \tag{5}$$

A substantial number of studies have found that the VTT increases with the size of the travel time savings in stated choice experiments (see for instance Bates and Whelan (2001), Hultkrantz and Mortazavi (2001) and Cantillo et al. (2006)). To further explore the properties of the value of time distribution and the role of $\Delta t$, a local logit regression is also performed in $(log(\Delta t),log(V))$-space. Rewriting (4), the constant probability lines of model 1 we have:

$$\frac{\Delta c}{\Delta t} = F_W^{-1}(q).$$

Hence, if model 1 is the true model, the constant probability contour lines would be straight and horizontal. The spacing between the lines reveals the shape of the value of time distribution function. Rewriting (5), the constant probability lines would follow the function:

$$\frac{\Delta c}{\Delta t} = -\frac{F_\epsilon^{-1}(1-q)}{a\Delta t} - \left(\frac{b}{a}\right).$$

Hence, if model 2 is the true model, constant probability contour lines would be curved.

## 2.3 Parametric modelling

The nonparametric models described above could be used to estimate the VTT. However, in this paper we apply a parametric model also, mainly for three reasons. First, we can only take out the effect of variables relating to the design of the experiment, such as the effects of different quadrants, by using a parametric model. Second, as discussed below, full support for the VTT distribution is more easily achieved in a parametric model estimation including covariates, since the covariates create more variation. Third, the parametric model further provides the possibility to separate inter-individual random variation in the VTT from random error affecting choices.

In section 4 it will be shown that the local logit regression results are more consistent with model 1 than with model 2. Therefore we adopt model 1 as a basis for the parametric model, assuming that an individual will choose the slow alternative (y=1) if the VTT is smaller than the trade-off value of time, i.e. if $W<V$. We take logs and add an error term, which leads us to the model[3]:

$$y = 1\{\log W < \log V + \mu\varepsilon\}. \tag{6}$$

The error term ε is taken to be iid standard logistic, such that a logit model results. The parameter $\mu$ is a scale parameter and the VTT is parameterised as:

$$W = \exp(\beta x + \delta), \tag{7}$$

where $\beta$ is a vector of parameters, $x$ is a vector of independent variables and $\delta$ is an individual specific random term. This formulation ensures that $W$ is positive, while the ranges of $\beta$ and $\delta$ are unrestricted. The ease with which independent variables are incorporated is an important advantage of the present model. Note that since this model is defined in the *log bid* space, there is no explicit indirect utility function.

The assumption that $W$ is individual specific and varies randomly in the population takes care of the correlation of the unobserved heterogeneity arising from repeated observations of the same individuals. For convenience, the error $\varepsilon$ is still taken to be iid standard logistic also within individuals such that a logit model results. The model hence separates choice specific logistic errors from the random variation in VTT between individuals.

We estimate a base model in which $\delta$ is taken to follow a normal distribution, such that $W$ is lognormal. To decide if this is an appropriate assumption for the distribution of $\delta$ we also estimate a more flexible model, which nests the base model. In the flexible model the distribution of $\delta$ can vary around the normal distribution using the semi-nonparametric (SNP) technique of Fosgerau and Bierlaire (2007). The fundamental mechanism behind this technique is that any well-behaved function can be approximated arbitrarily closely by polynomials. The Fosgerau and Bierlaire technique uses polynomials to generate densities on the unit interval, using that any distribution can be translated into any other distribution with the same support via a distribution on the unit interval. The technique thus introduces a number of additional parameters, called SNP parameters, corresponding to the coefficients of a certain polynomial, to allow additional flexibility in the mixing distribution. We add three SNP terms in the flexible model, $\gamma_1$-$\gamma_3$. This is done in such a way that the original distribution (in this case the normal distribution) results, if the additional parameters take the value zero. The $\chi^2$ test of parameter restrictions can then be applied to determine if the assumed distribution of $\delta$ can be rejected in favour of the more flexible distribution.

The assumption that $x$ and $\delta$ are independent is crucial, implying that the distribution of $\delta$ is unaffected by a shift in $x$. Using the parametric model specified in (6), we may then define $res = \log(V) - \beta x$. By nonparametrically regressing $y$ on *res* we obtain an estimate of:

$$E(y|res) = P(y = 1|res) = P(\delta - \mu\varepsilon < \log V - \beta x) = P(\delta - \mu\varepsilon < res) = F_{\delta-\mu\varepsilon}(\text{res}),$$

[3] 1{x>y} is defined to take the value 1 if x>y and 0 otherwise.

where $F_{\delta-\mu\varepsilon}$ is the cumulative distribution function of $\delta - \mu\varepsilon$. A necessary condition for the empirical identification of the distribution of $\delta$ is that $F_{\delta+\mu\varepsilon}$ is observed ranging all the way from 0 to 1. By applying local constant regression and regressing $y$ in $res$, an estimate of $F_{\delta-\mu\varepsilon}$ can be plotted as a function of $log(V)$ to investigate if $F_{\delta-\mu\varepsilon}$ can be identified. This is more easily achieved using a model including covariates, since $res$ then varies over a larger range than $\log(V)$ alone.

# 3 Data Collection

## 3.1 Survey

The data used in this study originate from a survey carried out in 2008. The survey comprised car, long and short distance train and bus modes. We show only results for car trips. The results for the other modes were similar. For the car mode, a sample of respondents was drawn from the population register. They were contacted by letter and asked to participate using the internet questionnaire. Non-respondents were contacted by telephone and asked to participate either via the internet or on the telephone. In the questionnaire, respondents were asked to list all car trips on a pre-specified day, from which one trip was randomly selected. Selection probabilities were higher for long distance trips.

Out of the 6000 respondents in the initial sample 3717 responded, implying a total response rate of 62 percent. Out of the 3717 respondents, 2240 did not fulfil the requirements for participation in the survey since they had not made a car trip longer than 5 km as driver on the prescribed survey day. This means that 1477 participated and were in scope for the survey. Of these, 1222 persons responded to the entire questionnaire by internet and 255 persons responded by telephone. Since persons who had not made any valid car trips most likely felt less obliged to respond than others, the relevant response rate may be underestimated.

160 respondents were discarded because they had reported an unrealistically low or high speed of the reference trip, because they did not pay the trip themselves (making the cost less relevant) or because they had only chosen the left or right alternative through all SC questions (5 telephone respondents and 19 internet respondents). After cleaning the data, 1317 respondents remained. Table 1 presents some descriptive statistics for the sample.

**Table 1: Descriptive statics.**

| | *Min* | *1 Quartile* | *Median* | *Mean* | *3 Quartile* | *Max* |
|---|---|---|---|---|---|---|
| *Having children < 13 years in household, dummy* | 0 | 0 | 0 | 0.25 | 0 | 1 |
| *Reference travel distance (km)* | 5 | 12 | 30 | 93 | 128 | 1114 |
| *Reference travel time (min)* | 4 | 15 | 30 | 81 | 105 | 840 |
| *After tax monthly income (k€)* | 0.40 | 1.10 | 1.80 | 1.61 | 1.80 | 4.17 |
| *Age* | 18 | 38 | 50 | 49.76 | 61 | 83 |
| *Employed, dummy* | 0 | 0 | 1 | 0.61 | 1 | 1 |
| *Purpose: commute, dummy* | 0 | 0 | 0 | 0.24 | 0 | 1 |
| *Purpose: service, dummy* | 0 | 0 | 0 | 0.49 | 1 | 1 |
| *Purpose: recreation, dummy* | 0 | 0 | 0 | 0.16 | 0 | 1 |

## 3.2 Experimental Design

The stated choice experiment comprises binary choices, where the alternatives are described by the in-vehicle travel times $t_0$ and $t_1$ and the travel costs $c_0$ and $c_1$. Recall that the implicit bid, $V$, offered in each choice equals $-(c_0 - c_1)/(t_0 - t_1)$. A key issue in designing the choice experiment was to ensure that the bid range extends over the support of the VTT distribution. Using a very wide bid range in the stated choice design implies that a larger part of the distribution is supported by the data, if respondents' behaviour is consistent with the theory. On the other hand, a larger bid range comes at the price of low efficiency in the data collection: Asking respondents already rejecting low bids additional questions with higher bids does not add much information about the shape of VTT distribution. Extensive piloting was undertaken to determine a bid range that would cover the VTT distribution to a reasonable extent.

The binary stated choice questions all related to a trip that the respondent recently had undertaken. The design of the stated choice questions is similar to the Danish VTT study, described in Fosgerau et al. (2006). In our case, the design generated eight time differences in the 10 – 30 percent range of the observed travel time, divided into four strata[4]. Two travel time differences were randomly assigned to each of the four quadrants. Eight VTT bids were drawn from 6 VTT strata[5] in the range 0.5 – 50 EUR/h and assigned randomly to each of the eight time differences. The absolute cost difference was then found for each choice situation by multiplying the absolute time difference by the trade-off value of time.

# 4 Results

The first observation is that only 15 out of 1317 drivers or 1.1 percent rejected all bids. The small percentage that accepted all bids may have values of time larger that the highest bid (which is between 40 and 50 EUR /h for 13 out of these 15 drivers), but this result could also be due to random error.

## 4.1 Nonparametric estimation of VTT

We then carry out local constant regression of $y$ on $V$, providing an estimate of $F_W$ in model 1. The regression is carried out for each of the four quadrants: WTP, WTA, EL and EG. We use a normal density kernel with bandwidth 0.05, selected by eyeballing.

**Figure 2: Probability of rejecting the bid as function of the bid (EUR/h) (WTP quadrant).**

**Figure 3: Probability of rejecting the bid as function of the bid (EUR/h) (WTA quadrant).**

Figure 2 to Figure 5 show the estimated VTT distribution for the four types of choices. The solid line represents the mean, whereas the dashed lines represent the asymptotic 95% pointwise confidence bands (low and high limits). The interval will cover the true value with probability 95% in large samples. The dotted interval shows the uniform confidence bands (low and high limits). The uniform confidence bands will contain the entire distribution in 95% of repeated large samples.

---

[4] Strata of time differences (percentages of observed travel time): 2 draws in [10%-15%], 2 draws in [15%-20%], 2 draws in [20%-25%], 2 draws in [25%-30%].

[5] Strata of bids (EUR/h): 1 draw in [0.5-1.5], 1 draw in [1.5-4], 2 draws in [4-10], 2 draws in [10-20], 1 draw in [20-40] ,1 draw in [40-50].

Figure 6 combines the estimates of $F_w$ for the four quadrants in a single graph. First, the estimates of $F_w$, and hence the probability of rejecting a bid, increases with the bid. The estimates of $F_w$ thus behave like cumulative distribution functions. This is reassuring and indicates that respondents' behaviour is consistent with theory. Second, the estimate of $F_w$ is not significantly different from one at the upper end of the bid range, which implies that the upper tail of the distribution may be assumed to be identified. Third, it seems that the WTA distribution roughly first-order stochastically dominates the distributions in the other quadrants, implying that the mean WTA is larger than the mean VTT in the other quadrants. Fourth, all quadrants have a significant share of the mass near zero.

Fifth, 87-93 percent of the mass of the distributions is found below 25 EUR/h. This corresponds approximately to the findings in the Danish data, in which about 85 percent of the VTT distribution was observed below the maximum bid 25 EUR/h. The finding that a substantially larger part of the distribution is revealed by the Swedish data with maximum bid 50 EUR, compared to the Danish data, is reassuring and provides support for the assumption that respondents' choices are governed by utility maximization.

**Figure 4: Probability of rejecting the bid as function of the bid (EUR/h) (EL quadrant).**

**Figure 5: Probability of rejecting the bid as function of the bid (EUR/h) (EG quadrant).**

**Figure 6: Probability of rejecting the bid as function of the bid (EUR/h) (all quadrants).**

*Local Logit*

The local logit model was estimated in a grid of 21*21 points over the range of Δt and Δc using a bandwidth of 0.4 times the range of each variable and a triangular density kernel. The resulting estimated response surfaces are depicted in Figure 7 and Figure 8, showing the constant probability curves of $\hat{P}(y = 1|\Delta t, \Delta c)$ for the WTP and WTA type-choices.[6]

In both quadrants, there is a tendency that the probability lines fan out from the origin and results are hence more consistent with model 1 than with model 2. The figures show also a tendency that the slope increases for time differences less than 100 minutes in WTA choices, implying that the VTT increases with Δt. This tendency is less marked for WTP choices.

**Figure 7: Local logit plot in (Δt,Δc)-space. WTP.**

**Figure 8: Local logit plot in (Δt,Δc)-space. WTA.**

In order to further investigate the response surface and the role of Δt, a local logit regression is performed in (log(Δt),log(V))-space, using the bandwidth 0.6 times the range of data and a triangular density kernel. Figure 9 and Figure 10 show the plots for the WTP and WTA quadrants. The lines are roughly parallel, indicating again that data are more consistent with model 1 than with model 2. The equal spacing indicates that the distribution of the error term is skewed. The slope increases with log(Δt), showing that VTT increases with the size of the time difference presented in the choice. These findings confirm the results of Fosgerau (2007) from the Danish data.

**Figure 9: Local logit plot in (log(Δt), log(*V*))-space. WTP.**

[6] Plots for the EG and EL quadrants are available on request. The results for these quadrants are similar to those shown.

**Figure 10: Local logit plot in (log(Δt), log(*V*))-space. WTA.**

## 4.2 Parametric model estimation

The previous section indicates that the data reveal the VTT distribution up to a point that allows us to determine the mean VTT reasonably well. The VTT distribution differs between quadrants. As has been discussed, the quadrant effect must be removed to obtain a reference-free VTT (De Borger and Fosgerau, 2008). To exclude the effect of the different quadrants, we specify these effects in a parametric model. We have found that model 1 describes the data better than model 2. Therefore we use the parametric model defined in section 2.

We estimate the base model specified by (6) and (7) and a flexible model which are specified the same way except that the semi-nonparametric terms, $\gamma_1$- $\gamma_3$, induce flexibility around the normal distribution $\delta$ as explained in section 2.3. The flexible model is estimated in order to test the assumption that that δ is normal. In the base model the VTT is parameterised as:

$$W = \exp(\delta + \beta_{EL} + \beta_{EG} + \beta_{WTP} + \beta_1 \log \Delta t + \beta_2 \log \text{cost} + \beta_3 \log \text{time}).$$

$\beta_{EL}$, $\beta_{EG}$ and $\beta_{WTP}$ are dummy parameters for the quadrants, and hence zero if the choice is not within the quadrant. WTA is the base case quadrant and this dummy, $\beta_{WTA}$, is therefore constrained to be always zero. $\beta_1$, $\beta_2$ and $\beta_3$ are parameters to be estimated. To capture the observed property of the data that VTT increases with the size of the travel time difference between the two alternatives in the binary choice we include (*log* Δ*t*). In particular, it allows for the effect that small travel time changes may have a smaller unit value. The specification also includes the log of the reference trip cost (*log cost*) and the log of the reference travel time (*log time*), which are also input variables in the construction of the experimental design. We include the design variables in the model to be able to control for the influence of the design variables on the VTT estimates. Socio-economic variables or trip purpose variables are not included in this model because there is a risk that these are not independent of $\delta$ and thus inclusion of these variables would increase the risk that the assumptions of the parametric model would be violated.

The parametric models are estimated using Biogeme (Bierlaire, 2003; Bierlaire, 2008). The parameter estimates are presented in Table 2.

**Table 2: Results from parametric model estimation.**

| | Base model | | Flexible model | |
|---|---|---|---|---|
| # draws: | 1500 | | 1500 | |
| # parameters: | 9 | | 12 | |
| # observations: | 8877 | | 8877 | |
| # individuals: | 1317 | | 1317 | |
| Final LL: | -3269 | | -3250 | |
| Adjusted Rho$^2$: | 0.47 | | 0.47 | |
| Variable | Value | t-test | Value | t-test |
| Mean δ | 0.179 | 1.71 | -0.926 | -2.38 |
| Std dev of δ | 1.230 | 27.76 | 2.020 | 6.58 |
| *log Δt* | 0.149 | 2.03 | 0.153 | 2.09 |
| *log cost* | 0.057 | 0.49 | 0.068 | 0.65 |
| *log time* | 0.194 | 1.28 | 0.183 | 1.27 |
| EG quadrant, dummy | -0.360 | -5.79 | -0.361 | -5.83 |

| | | | | |
|---|---|---|---|---|
| EL quadrant, dummy | -0.193 | -3.10 | -0.193 | -3.12 |
| WTP quadrant, dummy | -0.482 | -7.61 | -0.481 | -7.61 |
| $\gamma_1$ | | | 0.387 | 2.22 |
| $\gamma_2$ | | | 0.077 | 0.54 |
| $\gamma_3$ | | | -0.436 | -1.96 |
| scale | 1.660 | 13.78 | 1.680 | 13.97 |

The time difference is significantly positive, indicating that the VTT increases as the time difference increases. The standard deviation parameter is significantly positive, indicating that there is heterogeneity in the VTT. The parameters for the design variables travel time and travel cost are not significantly different from zero (although they are jointly significant). The $\chi^2$ test shows that the SNP parameters $\gamma_1$-$\gamma_3$ yield a significant improvement in overall model fit. Thus the normal distribution is rejected in favour of the more flexible distribution.

The quadrant dummies are significantly different from zero, indicating that there are indeed differences between quadrants. The dummy parameters imply that $WTP < EL$, $EG < WTA$, which is consistent with theory. Moreover, $½(\beta_{WTP}) = ½(\beta_{EL} + \beta_{EG})$ cannot be rejected in a statistical test with t-statistic 0.65 for the base model and 0.37 in the flexible model, implying that $(WTP * WTA)^{1/2} = (EG * EL)^{1/2}$ holds. The differences in VTT between the quadrants may thus be caused by loss aversion and we may thus rely on the theory set out by De Borger and Fosgerau (2008) to derive a reference-free VTT.

As described in section 2.3, we may apply local constant regression to estimate the CDF of the random elements specified in the parametric model, $F_{\delta-\mu\varepsilon}(res)$, where $res = \log V - \beta x$. To identify the VTT distribution without having to make assumptions of a specific distribution, we need to observe $F_{\delta-\mu\varepsilon}(res)$. To estimate $F_{\delta-\mu\varepsilon}(res)$ we regress $y$ on $res$, which gives us

$$P(\mu\varepsilon + \delta < \log V - \beta x) = P(\mu\varepsilon + \delta < res) = F_{\delta-\mu\varepsilon}(\mathrm{res})\ .$$

Figure 11 shows the estimate of $F_{\delta-\mu\varepsilon}(res)$ for the base model. This CDF reaches 0.9989 at the end of the range of $res$. Figure 12 plots the density of $res$, showing that the observations are in the range (-2.7,4.2). Specifically, according to Figure 12 there are lots of data above 2.5, where $\hat{F}_{\delta-\mu\varepsilon}(res)$ reaches 95%, and we may thus say that we can effectively identify the entire distribution of VTT.

**Figure 11: Nonparametric regression of choices y on residuals** $\mathrm{res} = \log(\mathrm{V}) - \beta'\mathrm{x}$ .

**Figure 12: Density of** $\log(\mathrm{V}) - \beta'\mathrm{x}$ .

We could stop here and compute the mean VTT from the estimate of $F_{\delta-\mu\varepsilon}(res)$. Another option is to compute the VTT by simulation using the parametric model and a sample of individuals. The advantage of using the parametric model for simulating the VTT is that the noise in the data, i.e. the error varying randomly within individuals, is separated from the random variation in the VTT between individuals. We can also take out the effect of quadrants to obtain the reference-free VTT.

Figure 13 and Figure 14 show the simulated value of time distribution for all individuals in the estimation sample. We have used a value of $\Delta T = 20\ min$ for all respondents, while the reference travel time and cost are as observed. The reference-free VTT is calculated as the geometric average: $(WTP * WTA * EG * EL)^{1/4}$, where for instance $WTP = \exp(\delta + \beta_{WTP} + \beta_1 \log \Delta\mathrm{t} + \beta_2 \log \mathrm{cost} +$

$\beta_3 \log \text{time}$) and $WTA = \exp(\delta + \beta_1 \log \Delta t + \beta_2 \log \text{cost} + \beta_3 \log \text{time})$. The figures show the VTT distributions simulated with the base model and with the flexible distribution. In Figure 13 the distributions are plotted in the bid range 0.1-20 EUR. In Figure 14 the same distributions are plotted in the range 0.1-100 EUR. Table 3 shows the mean values of the simulated VTT distributions.

The plot reveals that the flexible model adjusts the lognormal distribution by putting some mass near zero. This is consistent with the nonparametric regression plots, which also display some mass close to zero. Figure 13 and Figure 14 further reveal that the flexible model predicts a flatter right tail, beyond the range in which we have data.

Using the unbounded parametric distributions to compute the mean value of time requires that some assumption is made concerning the VTT distribution above the range in which we have data. Table 3 shows the mean VTT computed using various truncation points. One option is to use the residual analysis to give an indication of a truncation point. At the maximum observed *res*=4.2 (see Figure 12), the CDF in Figure 11 reaches 0.9989. This corresponds to a bid *V* of 131.9 EUR/h for the mean value of $\beta x$ = -1.11, which could be used as a truncation point of the VTT distribution. Changing the truncation point to 130 EUR/h from, for instance, 150 EUR /h has a negligible impact on the mean VTT. Moreover, in the range of truncation points of 130-200 EUR /h, the mean reference-free VTT remains stable in both models, but more so in the base model.

**Figure 13: VTT distributions of the two models (bid range - €20).**

**Figure 14: VTT distributions of the two models (bid range -€100).**

The mean VTT of the uncensored distribution is higher in the flexible model than in the base model, which indicates that the semi-nonparametric method modifies the normal distribution primarily outside the range where data supports the VTT distribution. This is consistent with the finding that the tail of the flexible distribution is flatter than that of the base model. In the range where the data does not support the VTT distribution the model estimation provides, however, no guidance concerning the distribution of the VTT. This underscores the importance of truncation in practical VTT estimation. The choice of truncation point seems, however, not to have a significant impact on the VTT results.

**Table 3: Mean VTT (EUR/h) for the base model and the flexible model.**

| 1 | Base model | Flexible model |
|---|---|---|
| Not censored | 7.1 | 7.8 |
| Censored at 130 EUR /h | 7.0 | 7.0 |
| Censored at 150 EUR /h | 7.0 | 7.1 |
| Censored at 200 EUR /h | 7.0 | 7.2 |

# 5 Conclusions

The possibility to identify the VTT and reduce lexicographic behaviour is of practical importance for all types of valuation studies, and value of time studies in particular, using stated choice data. To observe the entire VTT distribution, and compute its mean value without imposing strong assumptions, the range of bids in the data must extend over the support of the VTT distribution. Some earlier VTT studies have not met this condition. Fosgerau (2006) shows that in the Danish VTT data, a mean VTT cannot be derived, because the data did not support the full VTT distribution. However, increasing the bid range to 0.5-50 EUR/h in the present stated choice design proved to be sufficient to reveal the VTT distribution up to a point that enabled us to derive the mean VTT with reasonable confidence. It is thus possible to identify the VTT distribution in practice and to estimate a mean VTT. Moreover, we have shown that it is possible to keep the incidence of seeming non-trading or lexico-graphical behaviour small; only 1-2 percent of the respondents chose the cheapest alternative throughout the choice experiment. This result suggests that high incidence rates of lexicographical behaviour found in earlier stated choice data, not only in the Danish data, is caused by insufficient variation in the designs rather than non-trading in contradiction of utility maximization.

The relation between of lexicography and bid range is consistent between the Swedish, Danish and the Norwegian VTT data (Ramjerdi et al, 2010), indicating that the conclusions drawn in terms of possibility of identify the VTT is transferable between studies. While the present study provides guidance about necessary bid ranges and suggests that bid ranges in general should be wider in stated choice studies than what has previously been the practise, any new study will need piloting to ascertain whether the intended bid range is sufficient.

Based on nonparametric local logit regressions, a parametric model in *log(bid)* space was selected. This agrees with the findings of (Fosgerau, 2006). The parametric model was used to estimate the VTT distribution while controlling for effects induced by the stated choice experiment. The lognormal distribution of the VTT was rejected in a test against a more flexible distribution. However, the two distributions led to similar values of the mean VTT.

The analysis in this paper has concentrated on car drivers. A parallel analysis has been carried out for similar data concerning public transport. The results of the analysis are very similar, but the share of respondents choosing the cheapest alternative throughout the choice experiment was less than 1 %. The distribution was approximately lognormal for both long-distance trips and regional public transport trips.

# 6 References


Abrantes, P. A. L. and Wardman, M. R. (2011). Meta-analysis of UK values of travel time: An update. Transportation Research Part A: Policy and Practice, 45(1), 1-17.

Ahern, A. A. and Tapley, N. (2008). The use of stated preference techniques to model modal choices on interurban trips in Ireland. Transportation Research Part A: Policy and Practice, 42(1), 15-27.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., Sugden, R., (1997). A test of the theory of reference-dependent preferences. The Quarterly Journal of Economics 112 (2), 479–505.

Bates, J. and Whelan, G. (2001). "Size and sign of Time Savings", ITS Working Paper no. 561.

Beesley, M. E. (1965). "The Value of Time Spent in Travelling: Some New Evidence", Economica, vol. 32, no. 126, pp. 174-185.

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

Bierlaire, M. (2008). "An introduction to Biogeme", http://biogeme.epfl.ch/.

Brownstone, D., & Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transportation Research Part A: Policy and Practice, 39(4), 279-293.

Börjesson, M., Fosgerau, M., and Algers, S. (2009). The income elasticity of the value of travel time is not one number. Paper presented at the 2009 European Transport Conference, Leiden, The Netherlands.

Cairns, J. and van der Pol, M. (2004). Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments. Social Science and Medicine 58, pp 2211–2218

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, 355-79.

Cameron, T. A. and James, M. D. (1987). "Efficient Estimation Methods for "Closed-Ended" Contingent Valuation Surveys", Review of Economics and Statistics, vol. 69, no. 2, pp. 269-276.

Cantillo, V., Heydecker, B., and de Dios Ortuzar, J. (2006). "A discrete choice model incorporating thresholds for perception in attribute values", Transportation Research Part B: Methodological, vol. 40, no. 9, pp. 807-825.

Cirillo, C., & Axhausen, K. W. (2006). Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research Part A: Policy and Practice, 40(5), 444-457.

De Borger, B. and Fosgerau, M., (2008). A test of the theory of reference-dependent preferences: the trade-off between money and time. Journal of Urban Economics, Volume 64 (1), 101-115.

Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A 41 842–856

Fosgerau, M. and Bierlaire, M. (2007). A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B 41 784–794

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

Fosgerau, M., Hjorth, K., and Vincent Lyk-Jensen S. (2006). An integrated approach to the estimation of the value of travel time. ETC proceedings, http://www.etcproceedings.org/

Fowkes, A.S. and Wardman, M. (1988). The Design of Stated Preference Travel Choice Experiments with Particular Regard to Inter-Personal Taste Variations. Journal of Transport Economics and Policy, 22(1), 27-44

Hensher, D. A. (2006). Towards a practical method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions. Transportation Research Part A: Policy and Practice, 40(10), 829-840.

Hess, S., Bierlaire, M., & Polak, J. W. (February). Estimation of value of travel-time savings using mixed logit models. Transportation Research Part A: Policy and Practice, 39(2-3), 221-236.

Hess, S., Rose, J.M. and Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data Transportation Research Part D 15 (2010) 405–417

Horowitz, J. and McConnell, K.E. (2002). "A Review of WTA/WTP Studies" Journal of Environmental Economics and Management 44 426-447.

Hultkrantz, L. and Mortazavi, R. (2001). "Anomalies in the Value of Travel-Time Changes", Journal of Transport Economics and Policy, vol. 35, no. 2, pp. 285-300.

Killi, M., Nossum, Å. and Veisten, K. (2007). Lexicographic answering in travel choice: Insufficient scale extensions and steep indifference curves? EJTIR, 7, no. 1 (2007), pp. 39-62

Lancsar, E. and Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics. 15: 797–811

Potoglou, D., Robinson, N., Kim, C. W., Burge, P. and Warnes, R. (2010). Quantifying individuals' trade-offs between privacy, liberty and security: The case of rail travel in UK. Transportation Research Part A: Policy and Practice, 44(3), 169-181.

Randall, A. and Stoll, J. R. (1980). "Consumer's Surplus in Commodity Space", American Economic Review, vol. 70, no. 3, pp. 449-455.

Ryan, M., Watson, V. and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Economics 18: 321–336 (2009)

Ramjerdi, F., S Flügel, H. Samstad, M. Killi, (2010) Value of time, safety and environment in passenger transport. TØI-reports. 1053b/2010.

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

Wardman, M. and Ibáñez, N. (2011). The congestion multiplier: Variations in motorists' valuations of travel time with traffic conditions. Transportation Research Part A: Policy and Practice, In Press.