MPRA
Munich Personal RePEc Archive

Using nonparametrics to specify a model to measure the value of travel time

Fosgerau, Mogens

Technical University of Denmark

2007

Online at https://mpra.ub.uni-muenchen.de/12009/
MPRA Paper No. 12009, posted 08 Dec 2008 09:30 UTC

USING NONPARAMETRICS TO SPECIFY A MODEL TO MEASURE THE VALUE OF TRAVEL TIME *

Mogens Fosgerau, mf@dtf.dk
Danish Transport Research Institute

February 11, 2007

2nd. revised version

Abstract

Using a range of nonparametric methods, the paper examines the specification of a model to evaluate the willingness-to-pay (WTP) for travel time changes from binomial choice data from a simple time-cost trading experiment. The analysis favours a model with random WTP as the only source of randomness over a model with fixed WTP which is linear in time and cost and has an additive random error term. Results further indicate that the distribution of log WTP can be described as a sum of a linear index fixing the location of the log WTP distribution and an independent random variable representing unobserved heterogeneity. This formulation is useful for parametric modelling. The index indicates that the WTP varies systematically with income and other individual characteristics. The WTP varies also with the time difference presented in the experiment which is in contradiction of standard utility theory.

KEYWORDS: Willingness-to-pay, WTP, value of time, nonparametric, semi-parametric, local logit
JEL codes: C14, C25, R41

*I thank Bruno De Borger, Stefan Mabit, Ninette Pilegaard and Andrew Daly for comments. I also thank the referees for a number of comments that have substantially improved the paper. The usual disclaimer applies.

1 INTRODUCTION

The point of this paper is to apply nonparametric techniques to binary choice data in order to identify a model specification that is at the same time simple and able to describe the data well. The data relate to trips by bus and train and arise from a stated choice study designed to elicit the marginal willingness-to-pay (WTP) for travel time. Results emerge that challenge the classical model formulation using linear-in-parameters utility with additive independent error terms. The findings are evidence to the strength of nonparametric techniques. It is therefore a second purpose of the paper to illustrate the application of such techniques.

A large number of contributions have estimated logit models, mixed logit models and other models with the common feature that they embody indexes which are linear combinations of variables such as travel times and costs and parameters that are interpreted as marginal utilities. This kind of model is computationally convenient and consistent with random utility maximisation (RUM). It is however not implied by RUM and it may not fit the data very well. There may be other models which are equally RUM consistent and which match data better.

This paper finds first that the classical formulation of the binary logit model in terms of constant marginal utilities of time and cost is misspecified for the current data. Instead a simple formulation is proposed where the WTP is the only source of randomness. This is RUM consistent and fits the data. Cameron and James (1987) is an early example of a similar model. It circumvents the problem of specifying a distribution for the scale of the model (Train and Weeks 2004). Second, it is found that a model whereby the WTP depends on the absolute value of time difference offered and also on the trip duration gives a good representation of the data. These findings are robust as they emerge within a semiparametric model with weak assumptions on the stochastic terms of the model and under various specifications of the systematic variation in WTP. However, the dependence of the WTP on the absolute value of the time difference is not consistent with the standard assumptions of utility maximisation.

Much research has been devoted to the WTP for travel time changes as they usually constitute the main benefit of transport infrastructure investment (Hensher 2001, Mackie, Jara-Diaz and Fowkes 2001). The microeconomic formulation of the theory of the value of travel time was fundamentally formulated by Becker (1965), Johnson (1966), Oort (1969) and DeSerpa (1971). Jara-Diaz (2000) provides a review. The estimation of the WTP for travel time is reviewed in Hensher (2001) and Gunn (2000). See Fosgerau (2006) for warnings about estimating a mean WTP from discrete data.

Previous applications of nonparametric techniques in a transport context include Fosgerau (2006) who uses nonparametric regression, Horowitz (1993), Fosgerau (2005) and Fosgerau (2006) apply the Klein-Spady estimator. Fosgerau and Bierlaire (2006) suggest a seminonparametric extension of the binary mixed logit model. A particular strength of some of these techniques is that they allow one to visually inspect various distributions and relationships prior to the imposition of any specific functional form. Models estimated using nonparametric techniques may be used in their own right or they can be used as a prelude to guide later parametric modelling. This paper may be viewed in this perspective.

The layout of the paper is as follows. Section 2 provides some microeconomic foundation while section 3 reviews some nonparametric and semiparametric methods. Section 4 provides details about the data and presents the empirical analysis. Section 5 concludes.

2 MICROECONOMIC FOUNDATION

2.1 The setting

We shall be concerned with stated choice data arising from a particularly simple design, with two unlabelled alternatives described only by in-vehicle travel time and cost. This simplicity is very much a virtue in the present context since it allows us to undertake a concise analysis of the specification of an econometric model for such data. With more complicated designs there are many things going on at the same time and it may be difficult if not outright impossible to arrive at similarly firm conclusions. The insights that are obtained here in this simple context may later be tested in more complicated settings.

Consider a binary route choice for some transport mode between unlabelled alternatives 0 and 1, described by in-vehicle travel times t_0 and t_1 and travel costs c_0 and c_1. In every choice situation there is a fast and expensive alternative and a slow and cheap alternative, such that the choice involves a trade-off between time and money. The fact that alternatives are unlabelled allows us to freely exchange them in order to simplify the analysis. Rearrange the alternatives such that

$$\Delta c = c_1 - c_0 < 0 < \Delta t = t_1 - t_0.$$

In other words, we rearrange such that the first alternative 0 is always relatively fast and expensive, while the second alternative 1 is always relatively slow and cheap.

The choice indicator, denoted by y, is defined with the convention that $y = 1$ if the slow and cheap alternative 1 is chosen, otherwise $y = 0$. So we disregard whether the first or the second alternative is chosen but focus on the issue of interest, namely whether the cheap or the fast alternative is chosen.

Now consider the choice probability $P(y = 1|x)$ as a function of some variables x. These variables include $(\Delta c, \Delta t)$ or transformations of $(\Delta c, \Delta t)$. Any (parametric) model that we may want to estimate is a specification of $P(y = 1|x)$ up to some parameters to be estimated. Certain specifications restrict the shape of this function in certain ways. In the following sections we shall see what restrictions on $P(y = 1|x)$ arise from two simple, but very different, models based on utility maximization. The fundamental properties of these models are described in McConnell (1990).

The first is the classical formulation which directly parametrises the utility difference between alternatives. The second approach is based on the difference between alternatives of the expenditure function which is an equivalent representation of preferences. The difference arises from the way randomness is introduced. Later we shall look at the empirical counterpart of $P(y = 1|x)$ using nonparametric methods and use this information to distinguish between the two models.

2.2 Model A

The first type of model includes the familiar binary logit or probit models. We may start with an assumption of homogenous individuals. If we assume that the marginal utilities are constant across individuals and further that the residual utilities are independent of c_i and t_i, $i = 0, 1$, then we obtain under utility maximisation that $y = 1$ when

$$0 < \alpha \Delta c + \beta \Delta t + \epsilon.$$

Model A specifies that the choice depends on constant marginal utilities of time and cost and an additive independent random error. This is a strong assumption, see McFadden (1998). The error term is the only random element in this model and it is unrelated to the WTP. If ϵ is assumed to be normal or logistic, then the familiar probit or logit models arise. In this paper we do, however, not need to assume any particular distribution for ϵ.

Model A implies that

$$P(y = 1) = P(0 < \alpha \Delta c + \beta \Delta t + \epsilon) = 1 - F_\epsilon(-\alpha \Delta c - \beta \Delta t),$$

where F_ϵ is the distribution function corresponding to ϵ. The shape of the contours in $(\Delta c, \Delta t)$-space where the probability is constant is readily obtained:

$$P(y = 1) = q \Leftrightarrow \Delta c = -F_\epsilon^{-1}(1 - q)/\alpha - (\beta/\alpha)\Delta t.$$

So the constant probability contours for model A are straight lines in $(\Delta c, \Delta t)$-space. The spacing between these lines depends on the distribution of ϵ.

2.3 Model B

We now look at an alternative model, which is equally simple but has radically different implications for the shape of the function $P(y = 1)$. This model postulates that

$$y = 1 \Leftrightarrow w < -\Delta c/\Delta t,$$

where w is the individual random value of travel time. Now w is the only stochastic element of the model. We assume that the w are i.i.d. with distribution function F_w. For the moment we also assume that w is independent of $(\Delta c, \Delta t)$.

Then we can again find the contours in $(\Delta c, \Delta t)$-space where the probability $P(y = 1)$ is constant. We have

$$P(y = 1) = P(w < -\Delta c/\Delta t) = F_w(-\Delta c/\Delta t)$$

such that $P(y = 1) = q$ when

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

So for model B, in contrast to model A, the constant probability contours in $(\Delta c, \Delta t)$-space are straight lines that fan out from the origin.

In the empirical section 4 we shall plot $P(y = 1)$ against $(\Delta c, \Delta t)$. Under model A we will expect to find probability contours to be parallel straight lines. Under model B we will expect these to be straight lines that fan out from the origin.

Recall that $\Delta c < 0 < \Delta t$. We may take logs to find the constant probability contours in $(log(-\Delta c), log(\Delta t))$-space defined by the equation

$$log(-\Delta c) = log(F_w^{-1}(q)) + log(\Delta t).$$

Hence with model B, the constant probability contours are parallel straight lines with slope 1 in $(log(-\Delta c), log(\Delta t))$-space and with spacing depending on the distribution of w. Define $v = -\Delta c/\Delta t$ as the boundary value of time presented in the choice such that $log(v) = log(-\Delta c) - log(\Delta t)$. Then the constant probability contours will be slope zero straight lines in $(log(v), log(\Delta t))$-space. Moreover,

$$P(y = 1) = P(w < v) = P(log(w) < log(v)).$$

Consider now some other variable x. Plotting $P(y = 1)$ against $log(v)$ and x will yield slope zero straight lines when w is independent of x. This will not be the case if w is not independent of x, which gives a way of detecting such dependence. In the case when $log(w) = \gamma x + u$, then constant probability contours will be parallel straight lines. We shall make use of this property in the empirical section 4.

2.4 Some remarks

2.4.1 Remarks on the two model formulations

The two models A and B are equivalent in a deterministic setting without random elements. Omitting the random error term and fixing (α, β), we simply specify the utility difference as $\alpha\Delta c + \beta\Delta t$. Then, defining $w = \beta/\alpha$, we have

$$0 < \alpha\Delta c + \beta\Delta t \Leftrightarrow w < v$$

The difference between the two models then just lies in the way randomness is introduced.

The mixed logit model (McFadden and Train 2000) in its standard formulation in terms of marginal utilities is a hybrid of models A and B, where the ϵ of model A is assumed to be logistic and (α, β) is assumed to random as in model B following some prespecified joint distribution.

2.4.2 Remarks on WTP and the travel time difference

In the econometric analysis below, we shall see how the WTP depends on covariates and we will focus particularly on Δt and trip duration t.

In general, it is possible for the WTP for a time change to vary with Δt under utility maximisation. This occurs when indifference curves (in (t, c)-space) are non-linear, as the slope of an arc between two points on an indifference curve depends on the distance between them. The slope is the value of time that leaves the individual indifferent in such a pairwise comparison. Fixing one point, the slope will increase with the absolute value of Δt on one side and decrease on the other when indifference curves are convex. Thus the WTP depends on the travel time in both alternatives in a comparison, but the WTP is not monotonous in the absolute value of the travel time difference.

Prospect theory (Kahneman and Tversky 1979) describes the process of editing which might plausibly produce an increasing relationship between the WTP and Δt since small values of Δt may tend to be ignored by respondents. Under prospect theory, reference dependence may also induce such an effect (Tversky and Kahneman 1991, Bateman, Munro, Rhodes, Starmer and Sugden 1997, Borger and Fosgerau 2006). Reference dependence is ruled out under classical utility theory, while it could be possible to include the editing effect through some kind of decision cost. We shall return briefly to these issues in the conclusion but they are not the main the interest of this paper.

Finally, the WTP will in general depend on trip duration under the classical assumptions of RUM. However, trip duration is clearly an endogenous variable since it is chosen by the individual partly based on his WTP. It is therefore not immediately possible to say more than that the two are correlated. Causality may go either way.

3 A REVIEW OF SOME NONPARAMETRIC TECHNIQUES

This paper makes extensive use of nonparametric techniques. Since these methods are still fairly new in transport research, I will review the techniques that I employ in more detail than usual in order to provide a feel for how they work.

For more details one may wish to consult the sources. Härdle (1990), Horowitz (1998), Pagan and Ullah (1999) and Yatchew (2003) are good general references to the vast literature on nonparametric and semiparametric econometrics. The term nonparametric generally relates to models that leave functional form and distributional assumptions unspecified. Semiparametric models are between the nonparametric and the parametric and mix nonparametrics with some parametric element. Some of the techniques reviewed in this section are actually semiparametric.

Consider first the general situation where we have a dependent binary variable y and a vector of independent variables $x \in \mathbb{R}^d$. To model such data we only need to describe the unknown function $P(x) = P(y = 1|x)$, since $P(y = 0|x) = 1 - P(x)$. We seek to describe $P(x)$ from data with minimal assumptions imposed. Denote the data by $\{y_n, x_n\}_{n \leq N}$

3.1 Local constant regression

Even though y is binary, it is useful to define an error term by writing $y = P(x) + \eta$ such that η is the residual in a regression of y on $P(x)$. Then $E(\eta|x) = 0$ since $E(y|x) = P(x)$. This is sufficient to estimate the function $P()$ of x by a so called local constant regression.

Suppose we have many observations of y at some particular point x_0. Then $P(x_0)$ may be estimated just as the average of these y's. In general, when the variables in x are continuous, there will not be many observations available with the same x. We can instead estimate $P(x_0)$ as the weighted average of y in a small neighbourhood of x_0. This amounts to fitting a constant locally, hence the term local constant regression.

A kernel K and a bandwidth h are introduced to define this neighbourhood. The kernel is a general function used to weight observations. This paper uses either a triangular kernel based on a triangular density or a standard normal density kernel. For one-dimensional x, the triangular kernel is defined as $K(x) = (1 - |x|)1\{|x| \leq 1\}$.[1] For more dimensions one may just take the product of such one-dimensional K's. Then K is positive, symmetric, has maximum at zero and tends to zero as $\|x\|$ increases. The same is true when K is taken to be the product of standard normal densities.

The bandwidth h is used to determine the size of the neighbourhood over which to average. Define weights around the point x_0 by $K_h(x - x_0) = K(\frac{x-x_0}{h})$. Now K_h is large near x_0 and diminishes as $\|x - x_0\|$ increases. $K_h = 0$, when x is far away from x_0.

It is possible to estimate P at some point x_0 by forming a weighted average of y_n around x_0

$$\hat{P}(x_0) = \frac{\sum_{n \leq N} y_n K_h(x_n - x_0)}{\sum_{n \leq N} K_h(x_n - x_0)} \tag{1}$$

The weighting ensures that y's near x_0 receive the largest weight. If the (coordinatewise) distance between an x_n and x_0 is greater than the bandwidth h, then x_n receives weight zero from the triangular kernel. The term in the denominator ensures that the weights sum to one.

We can get a view of the shape of P by computing this average for a range of values of x_0. As the sample size tends to infinity we can let the bandwidth h tend to 0, such that the bias from averaging disappears and $\hat{P}$ tends to P. The selection of h is discussed in section 3.5.

[1] The indicator function $1\{\}$ is 1 if the condition is true and 0 otherwise.

3.2 The Klein-Spady estimator

The main virtue of the local constant constant regression is its simplicity. The estimate of P at any point has a closed form expression and is fast to compute. This is utilised by the Klein and Spady (1993) estimator.

Suppose we are willing to assume that $y = 1\{\gamma x > \epsilon\}$. This index assumption turns the model into a semiparametric model. It is a quite strong restriction on the shape of the function P. The model assumes that $x \in \mathbb{R}^d$ can be collapsed into a single dimension via the index γx. In fact, since

$$P(x) = P(y = 1|x) = P(\gamma x > \epsilon) = F_\epsilon(\gamma x)$$

such that

$$P(x_1) = P(x_2) \Leftrightarrow \gamma(x_1 - x_2) = 0,$$

only changes of x in the direction of γ will affect P. Changes of x in directions that are orthogonal to γ will not affect P.

If F_ϵ was known, then γ could be estimated by maximising the log-likelihood function

$$L(\gamma) = \sum_n y_n log(F_\epsilon(\gamma x_n)) + (1 - y_n) log(1 - F_\epsilon(\gamma x_n)) \tag{2}$$

The Klein-Spady estimator simply replaces $F_\epsilon(\gamma x_n)$ in (2) by a nonparametric estimate using local constant regression as in (1). For a given γ we can regress $\{y_m\}_{m \neq n}$ on $\{\gamma x_m\}_{m \neq n}$ at the point γx_n to obtain an estimate $\hat{F}_\epsilon(\gamma x_n)$ of $F_\epsilon(\gamma x_n)$. This estimate has a closed form that can be plugged directly into (2). The result is a likelihood function that can be maximised using standard optimisation routines.

Klein and Spady (1993) present Monte Carlo evidence indicating that there may be only modest efficiency losses with their estimator relative to maximum likelihood estimation when the distribution of ϵ is known. In general, of course, the distribution of ϵ is not known.

All discrete choice models must be normalised since the scale of the parameters is not identified separately from the variance of the error term. It is customary to introduce the normalisation by fixing the variance of the error term to some value. This is impractical when the error term is not specified. The normalisation is instead imposed by fixing the parameter for one of the explanatory variables to 1. This variable should be continuous with a strong impact on the dependent variable.

3.3 The local logit model

A (local) constant is often not a very good approximation to an unknown function to be estimated. The poorer this approximation is, the more the bandwidth must be decreased in order to avoid bias. This leads to more uncertain estimates as fewer observations are then used for averaging. Moreover, local constant regression does not utilise the information that y can only take values 0 and 1.

This situation may be improved on by fitting a local model for a binary dependent variable rather than just a constant. This local model will fit at least as well as a constant and so more observations can be used for the estimate of each $P(x_0)$ without increasing bias. Hence the optimal bandwidth is larger for an appropriate local model.

This is the idea of the local logit model of Fan, Heckman and Wand (1995). It uses the same local weights defined by a kernel and a bandwidth, but now a local logit model is estimated at the point x_0. This is just a plain binary logit model where again the observations are weighted such that observations near x_0 receive large weight while observations far away receive small weight. Consider a local approximation to $P()$ at the point x_0 using $P(x) \approx F(a + b(x - x_0))$ where now F is the logistic distribution. Note that $P(x_0) = F(a)$ and hence $F(a)$ will be our estimate of $P(x_0)$. Maximise

$$L(x_0, a, b) = \sum_n K_h(x_n - x_0)[y_n log(F(a + b(x_n - x_0))) + (1 - y_n) log(1 - F(a + b(x_n - x_0)))]$$

with respect to (a, b) at each point x_0. This means that the estimates for (a, b) depend on x_0. When the bandwidth is large these estimates become more alike and when the bandwidth is infinite the model is just the ordinary logit model which has the same value of (a, b) for all datapoints.

3.4 The Zheng test

Say we have obtained parameter estimates $\hat{\gamma}$ for the model $P(y|\gamma x)$, using in our case the Klein-Spady estimator. We can then calculate

$$\hat{P}(x) = P(\hat{\gamma} x) = P(y = 1|\hat{\gamma} x)$$

from this model as an estimate of $P(x)$. We now wish to test whether this model fits the data, that is test whether $P(x) = \hat{P}(x)$. The Zheng (1996) test is useful in this situation.[2]

[2] A number of similar tests are discussed in Pagan and Ullah (1999).

Define the residuals $\varepsilon = y - \hat{P}(x)$. Then $E(\varepsilon|x) = P(x) - \hat{P}(x)$. If the model is correct, then $E(\varepsilon|x) = 0$. This implies that also

$$\tilde{T} = E[\varepsilon E(\varepsilon|x)p(x)] = 0,$$

where $p(\cdot)$ is the density of x. If, on the other hand, the model is not correct, then

$$\begin{aligned}\tilde{T} &= E[E(\varepsilon|x)^2 p(x)] \\ &= E[(E(y|x) - \hat{P}(x))^2 p(x)] > 0\end{aligned}$$

A standardised sample analogue of $\tilde{T}$ is

$$T = \frac{\sum_{n=1}^{N}\sum_{m=1,m\neq n}^{N} K_h(x_n - x_m)\varepsilon_n\varepsilon_m}{\left[\sum_{n=1}^{N}\sum_{m=1,m\neq n}^{N} 2K_h^2(x_n - x_m)\varepsilon_n^2\varepsilon_m^2\right]^{1/2}}$$

Zheng shows that this statistic is distributed as standard normal under the null hypothesis that the model is correct. Thus we can calculate T and reject our model if $|T|$ is large. Zheng further shows that under the alternative, T diverges to infinity as the sample size increases. This implies that the power of the test to reject false null hypotheses tends to 1 as the sample size increases.

3.5 Bandwidth selection

The choice of kernel is generally less important as results will not be much affected. Selecting an appropriate bandwidth is, however, a genuine issue with no easy solution. If the bandwidth is too large then there is oversmoothing. Too much detail disappears and there will be bias. If the bandwidth is too small, then there will be undersmoothing: that is, overfitting to random fluctuations in data. Thus the choice of bandwidth is about balancing the trade-off between bias and variance. A significant part of the nonparametric literature is devoted to the issue of bandwidth selection but without any single method emerging as a clear winner. Here we shall discuss just the three simplest approaches.

Pagan and Ullah (1999) support the use of so called "eye-balling" when x is at most two-dimensional. It consists of looking at the estimates to select a bandwidth where the estimated function has an appropriate number of features.

Another more automatic approach is to use cross-validation (Härdle 1990, Pagan and Ullah 1999). This is performed using the estimated prediction error $EPE(h) = N^{-1} \sum_n (y_n - m_n(x_n, h))^2$, where the estimate $m_n(x_n, h)$ is computed with bandwidth h, leaving out the n'th observation. A bandwidth can then be selected by minimising $EPE(h)$.

A third approach is to determine an optimal bandwidth from convergence properties of the estimator. This approach relates the bandwidth to the sample size.

3.6 Technical implementation details

For the sake of completeness, this section gives some details on the implementation of the various procedures. The less technically inclined may skip this section. All programming for this paper is carried out in Ox (Doornik 2001).

The local logit model is implemented with a triangular product kernel. Prior to the application of the local logit model, each explanatory variable is rescaled to the unit interval. The bandwidths stated below relate to the rescaled data and are chosen by cross-validation.

The Klein-Spady estimator is implemented with a normal density kernel. The cross-validated bandwidth from the local logit model is first reduced to account for the different kernels (Härdle 1990, Ch. 5.4). Then a further reduction by a factor 0.3 is applied to account for the higher bias of the local constant regression applied by the Klein-Spady estimator.

The Zheng test is implemented using a normal density kernel and the optimal bandwidth suggested by Zheng of $c * N^{-m/5}$, where c is a constant arbitrarily set to 1, N is the sample size and m is the number of dimensions of the domain of the unknown function to be tested.

4 ECONOMETRIC ANALYSIS

With the machinery now in place we can begin the econometric analysis. Recall that the aim is to find a simple model that is able to describe the data well. We will therefore start with something very simple and then gradually elaborate the model. But first the dataset must be presented.

4.1 Data

The data origin from the Danish value of time study conducted for the Danish Ministry of Transport and Energy. The questionnaire design is discussed in Burge and Rohr (2004) and in Burge, Rohr, Vuk and Bates (2004). This paper uses binary stated choice data from a simple within-mode experiment, conducted for the respondent's current mode. I have selected respondents using the two main public transport modes, bus and train, for this analysis.

Respondents choose between unlabelled alternatives differing only with respect to in-vehicle travel time and cost. The alternatives are generated relative to a reference trip, which is a recent trip undertaken by the respondent.

Four types of choices are generated as time and cost gains or losses relative to the reference trip. In this way, the (t, c) plane is divided into four quadrants with origin at the reference trip. One quadrant compares the reference trip with a slower but cheaper trip. The opposite quadrant compares the reference with a faster but more expensive trip. Also one quadrant compares a trip with reference time but higher cost with a slower trip with reference cost. The opposite quadrant compares a trip with reference time but lower cost with a faster trip with reference cost.

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 from the interval [2 : 200] Danish Kroner (DKK) per hour[3], 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.

The data have been trimmed prior to analysis by removing the about 5% of observations with the largest journey times and time and cost differences as otherwise the data would have been very thin at these large values. This left 8763 bus and 6688 train observations for analysis. Some summary statistics for the data employed are given in Table 1.

4.2 Local logit in $(\Delta t, \Delta c)$-space

We begin the econometric analysis by estimating a local logit model where the expected choice is viewed as an unspecified function of Δt and Δc. The

[3] 1 EUR = 7.5 DKK.

local logit model is estimated in a grid of 21 $\times$ 21 points over the range of Δt and Δc using a cross-validated bandwidth of .25.

The results for the bus and train modes are shown in Figure 1. Several findings emerge. First, from the estimated regression surfaces we note that the constant probability contours generally have a positive slope, which corresponds to a positive value of time. This is reassuring, but of course not so surprising. Second, the figures suggest there might be a tendency for the slopes to increase as the time difference gets larger. Thus the WTP distribution may not be independent of the time difference.

Most importantly, the constant probability contours are clearly not parallel, as they would be in model A with constant marginal utilities and independent error terms. It rather seems as if the contours fan out from the origin as they would in model B. It is thus apparent that the data are much more consistent with model B than with model A. We therefore use model B as the starting point for our further elaboration of a model specification.

4.3 Local logit in WTP space

A local logit regression is then performed in $(log(v), log(\Delta t))$-space, where v is the boundary value of time. This uses cross-validated bandwidths of 0.5 for bus and 0.4 for train. These bandwidths relate to data rescaled to the unit interval so they are comparable to those above. The fact that the bandwidths are higher suggests that the local model fits the data a lot better in $(log(v), log(\Delta t))$-space. The results from this local logit regression, shown in figure 2, seem to support this view.

The constant probability contours are roughly parallel, indicating that their spacing shows the distribution of the WTP. It is apparent that this picture would arise from a model like model B.

The constant probability contours do however seem to depend on Δt, which contradicts the assumption that w is independent of Δt. Thus, the WTP per minute seems to depend on the absolute value of the time difference. Following the discussion in section 2.4.2 it is clear that the standard utility theoretic model for choices determined only by time and cost cannot accommodate this effect of Δt - a broader theory is required. Dependence of the WTP on Δt is not a new finding, refer for example to Bates and Whelan (2001), Hultkrantz and Mortazavi (2001) and Cantillo, Heydecker and de Dios Ortuzar (2006).

To capture the relationship, we introduce a parameter for $log(\Delta t)$ to the model. Consider the extension

$$\begin{aligned} y &= 1\{log(w) < log(v)\} \\ log(w) &= \gamma log(\Delta t) + u \end{aligned} \tag{3}$$

where u is independent of $(v, \Delta t)$ with unknown distribution. Thus the distribution of u is taken as fixed and the location of $log(w)$ is shifted linearly by $\gamma log(\Delta t)$. Based on figure 2 we expect a positive estimate of γ.

The parameter γ is estimated using Klein-Spady, which can be done without having to specify the distribution of u. The Zheng test statistic is applied to test the restriction of model (3) against the general model $P(y|log(v), log(\Delta t))$, that merely specifies that the choice probability depends on $log(v)$ and $log(\Delta t)$. Results are shown in Table 2. It appears that the dependency on $log(\Delta t)$ is strongly significant and of the expected sign. The restriction that the relationship can be captured by the parameter γ is accepted by the Zheng test. It thus seems that model (3) is an adequate model for the data after allowing for dependence of $log(w)$ on Δt.

4.4 Introducing journey time

In section 2.4.2 it was argued that the WTP may depend on travel time. Therefore the model is expanded by including the variable *jtime* for travel time in the main mode. This variable measures the main mode travel time of the actual current trip, which was reported by the respondents and around which the choice alternatives are pivots. Expand the model in (3) by

$$log(w) = \gamma log(\Delta t) + \eta log(jtime) + u$$

Again, parameters are estimated using Klein-Spady and the linear restriction is tested using the Zheng test. The results are shown in Table 3.

The parameters γ for $log(\Delta t)$ are still positive and significant. The parameter η for journey time is not significantly different from zero for bus while it is strongly significant and positive for train. Train trips are generally much longer than bus trips. The Zheng test accepts the restrictions of the model against the general alternative.

Local logit regressions of y on $(log(jtime), log(v)-log(\Delta t))$ are performed in figure 3 to show the dependency of choices on the journey time. The slopes estimated in Table 3 are also in evidence in the figure. Independence of u and the index implies that the constant probability contours should be parallel in Figure 3, which they seem to be. Thus the assumption of independence seems to be a fair approximation to the data.

4.5 Introducing more covariates

The conclusions of the previous section 4.4 are checked by adding a number of variables to the model. Expand the definition of $log(w)$ to

$$log(w) = \delta x + u \tag{4}$$

where now the variables in the vector x include $log(\Delta t)$, $jtime$ and a number of other variables. Descriptive statistics for the variables are provided in Table 4.

The variable Sex is 1 for females and 0 otherwise; $income$ is net personal annual income; $inc1$ is a dummy for the lowest income group (less than 100,000 DKK/year); $incNA$ is a dummy for missing income information; $Commute$ is 1 when the travel purpose is commuting; $Passengers$ is 1 when there is at least one accompanying person on the trip. Note that $log(income)$ and age have been demeaned before input to the estimation procedure.

The parameters in Table 5 are estimated using Klein-Spady. The results make intuitive sense. The coefficients for $log(\Delta t)$ and $log(jtime)$ are much the same as before. Income has a strong influence on the location of the WTP distribution with a significantly positive coefficient. The age terms indicate decreasing WTP with age. Women have lower WTP than men with similar values of the coefficients for all modes. The travel purpose dummy for commuters is not significant and the presence of accompanying persons has no detectable influence on WTP.

This final model is tested by the Zheng test for the hypothesis that $E(y|log(v), \delta x) = E(y|log(v) - \delta x)$ with δ fixed at the estimates in Table 5. This hypothesis is accepted indicating that conditional on the definition of the index δx, we can accept the model in (4), whereby log WTP is equal to the linear index plus an independent error.

5 CONCLUSION

Even though the mixed logit model can, in principle, approximate any RUM discrete choice model by using enough mixing (McFadden and Train 2000), this can easily be very demanding in terms of data and the complexity of the model. It is thus still worthwhile to look for good model specifications. Our results indicate quite clearly that the present data are rather more consistent with a random WTP determining choices than with the classical deterministic utility difference and an independent error term. So it is likely that a model taking the random WTP as a fundamental feature will be simpler and less demanding on data.

For the present data we have formulated a simple model that focuses directly on the quantity of interest: the WTP. The problems of estimating the WTP as the ratio of two possibly random parameters are avoided by this model.

Furthermore, we have found that the model whereby the log WTP is described as a linear index plus an additive independent residual describes the data well. This is a very useful result for parametric modelling. Based on this result one could specify

$$y_t = 1\{log(w_t) < log(v_t) + \epsilon_t\}$$
$$log(w_t) = \delta x_t + u$$

where subscript t indexes the choice situation to allow for panel data, and ϵ_t are choice specific logistic errors. The WTP is given a log-linear formulation and contains a random component u. Such a model has in fact been estimated in Fosgerau and Bierlaire (2006) where u has a normal distribution and the model is shown to greatly outperform the classical formulation. They accept the normality assumption for u in tests against general alternatives.

The model can be estimated with commercially (and freely) available software. Indeed, one may just specify one "utility function" to be ϵ_{1t} and the other to be

$$-\mu log(v_t) + \delta x_t + \sigma u + \epsilon_{2t},$$

where u is for example a random normal variable, σ is a standard deviation to be estimated and μ takes care of the scale of the error terms ϵ, such that the estimate of $log(w)$ becomes

$$log(w_t) = \frac{\delta}{\mu} x_t + \frac{\sigma}{\mu} u.$$

Returning to the results of this paper, the local logit model was used to visualise the relationship between independent variables and choice probabilities. This technique is not hard to use and one can imagine many other circumstances where it would be useful for inspecting the data prior to the development of complicated models. In general it seems very worthwhile to start a modelling exercise using methods that do not impose many assumptions before moving on to more models of more explicit form.

Also the possibility to check a functional form against a general alternative should be noted. I chose the Zheng test among several candidates since it is easy to implement and theoretically powerful. This kind of test offers the possibility to check a discrete choice model against a more general alternative.

The log WTP formulation enabled the estimation of a number of determinants of individual WTP, something which otherwise could have been hard to achieve.

A conclusion for the present data is that the distribution of WTP depends on the absolute value of the time difference. This is not a new finding. Using similar data, Hultkrantz and Mortazavi (2001) find also that small travel time changes are valued less than large. The effect has also been found in the Netherlands and the UK (Bates and Whelan 2001). This raises the issue of how to interpret this finding, particularly if the interest is to measure a single WTP for use in project evaluation. The subject of small travel time changes in project evaluation is discussed in Welch and Williams (1997) and Mackie et al. (2001). Hultkrantz & Mortazawi argue mainly in favour of an explanation in terms of decision costs whereby the effort in deciding whether a given time saving is worth the cost may outweigh the potential gain. This is similar to the editing of prospect theory (Kahneman and Tversky 1979). Cantillo et al. (2006) consider the similar idea of perception thresholds, which also produce a lower value for small travel time changes. The form of the dependency of the WTP distribution on the time difference is useful, since the time difference has been found to only affect the location of the distribution of log WTP. The distribution of log WTP is otherwise unchanged.

References

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

Bates, J. and Whelan, G. (2001) Size and sign of Time Savings *ITS Working Paper* (561).

Becker, S. G. (1965) A theory of the allocation of time *Economic Journal* **75**(299), 493–517.

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

Burge, P. and Rohr, C. (2004) *DATIV: SP Design: Proposed approach for pilot survey* Tetra-Plan in cooperation with RAND Europe and Gallup A/S.

Burge, P., Rohr, C., Vuk, G. and Bates, J. (2004) Review of international experience in VOT study design.

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

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* **40**(9), 807–825.

DeSerpa, A. C. (1971) A Theory of the Economics of Time *The Economic Journal* **81**, 828–845.

Doornik, J. A. (2001) *Ox: An Object-Oriented Matrix Language* Timberlake Consultants Press London.

Fan, J., Heckman, N. E. and Wand, M. P. (1995) Local Polynomial Regression for Generalized Linear Models and Quasi-Likelihood functions *Journal of the American Statistical Association* **90**(420), 141–150.

Fosgerau, M. (2005) Unit income elasticity of the value of travel time savings.

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

Fosgerau, M. and Bierlaire, M. (2006) A practical test for the choice of mixing distribution in discrete choice models *Transportation Research Part B: Methodological* **Forthcoming**.

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

Hensher, D. (2001) Measurement of the Valuation of Travel Time Savings *Journal of Transport Economics and Policy* **35**(1), 71–98.

Horowitz, J. (1993) Semiparametric estimation of a work trip mode choice model *Journal of Econometrics* **58**, 49–70.

Horowitz, J. (1998) *Semiparametric Methods in Econometrics* Lecture Notes in Statistics Springer New York.

Härdle, W. (1990) *Applied Nonparametric Regression* Vol. 19 of *Econometric Society Monograph Series* Cambridge University Press.

Hultkrantz, L. and Mortazavi, R. (2001) Anomalies in the Value of Travel-Time Changes *Journal of Transport Economics and Policy* **35**(2), 285–300.

Jara-Diaz, S. R. (2000) Allocation and valuation of travel time savings *in* D. Hensher and K. Button (eds), *Handbooks in Transport, Vol. 1: Transport Modelling* Pergamon Press Oxford pp. 303–319.

Johnson, M. B. (1966) Travel Time and the Price of Leisure *Western Economic Journal* **4**, 135–145.

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

Klein, R. W. and Spady, R. H. (1993) An Efficient Semiparametric Estimator for Binary Response Models *Econometrica* **61**(2), 387–421.

Mackie, P. J., Jara-Diaz, S. R. and Fowkes, A. S. (2001) The value of travel time savings in evaluation *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 91–106.

McConnell, K. E. (1990) Models for referendum data: The structure of discrete choice models for contingent valuation *Journal of Environmental Economics and Management* **18**(1), 19–34.

McFadden, D. (1998) Measuring Willingness-To-Pay for Transportation Improvements *in* T. Gärling, T. Laitila and K. Westin (eds), *Theoretical Foundations of Travel Choice Modeling* 1 edn Pergamon Oxford chapter 15, pp. 339–364.

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

Oort, O. (1969) The evaluation of travelling time *Journal of Transport Economics and Policy* **3**, 279–286.

Pagan, A. and Ullah, A. (1999) *Nonparametric Econometrics* Cambridge University Press.

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

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

Welch, M. and Williams, H. (1997) The sensitivity of Transport Investments Benefits to the Evaluation of Small Travel-Time Savings *Journal of Transport Economics and Policy* **XXXI**(3), 231–254.

Yatchew, A. (2003) *Semiparametric Regression for the Applied Econometrician* Themes in modern econometrics Cambridge University Press.

Zheng, J. X. (1996) A consistent test of functional form via nonparametric estimation techniques *Journal of Econometrics* **75**(2), 263–289.