Time and income constraints in discrete choice models with an application to mode choice

Jonas, Eliasson

KTH Royal Institute of Technology

November 2000

Time and income constraints in discrete choice models

with an application to mode choice

Jonas Eliasson

Abstract

We discuss how differences in disposable time and money should be incorporated in discrete choice models. Starting from a general framework, we test several indirect utility functions in an application on mode choice, using data from the recent Swedish national travel survey.

Incorporating differences in available income and time improves goodness-of-fit significantly. It also affects the forecasts obtained from the model. Both average elasticities and the implied average values of time depend not only on whether time and income enter the model, but also on the way this is done.

The best results in terms of goodness-of-fit are obtained from a Taylor expansion of a Box-Cox function. We show that it is possible to make the Taylor expansion around different points, and that the choice of expansion point will affect both the goodness-of-fit, the average value of time and the elasticities of the model.

Cite as:

Eliasson, J. (2000) Time and income constraints in discrete choice models, with an application to mode choice. In Eliasson, J.: Transport and Location Analysis, doctoral dissertation, KTH Royal Institute of Technology, Stockholm 2000.

1 Introduction

There are at least three reasons to include effects of differences in available income and time in demand models. The first is that this might improve the performance of the model in the sense of better data fit and improved predictions. The second is that there are questions that cannot be treated otherwise, for example effects of changes in the income distribution or prices on other goods. The third reason is that disregarding individual differences in available time and money and the resulting differences in marginal utilities of time and money can cause severe misjudgments in welfare evaluations, as discussed in for example Eliasson (2000).

The reason that many demand models cannot capture these effects is that they assume a linear indirect utility function, i.e. the marginal utility of income and time is assumed to be constant. Apart from the problem that such functions are not able to reflect effects of differences in available income and time, there is sometimes the additional problem that we may want to evaluate proposed changes where the changes in time and income are so large that the assumption of constant marginal utility of time and money is untenable. Considering effects of available time and money and possible nonlinearities in the corresponding marginal utilities is especially important when a model is used not only to predict changes in future demand, but also to reveal a certain population's preferences, for example their monetary value(s) of time.

In this paper, we will discuss a special case of demand models, namely discrete choice models. The paper contains a theoretical part and an empirical part. In the former, we discuss how income and time enter a discrete choice model in a way that is consistent with microeconomic theory. The main points are the discussions about the necessity to assume working time restrictions ex ante, the possibility to include time or time components in a direct utility function and the implications of this for value of time measurements, and the possibility to map the inherently ordinal conditional indirect utility function to a cardinal utility scale in different ways. Further, we discuss Taylor expansion of the conditional indirect utility function, interpretations of this and the possibility to choose expansion point. Two natural expansion points to consider are the gross available income and time and the expected residual income and time. The latter is special in that it is an endogenously determined point, depending on the estimated parameter values. We briefly discuss the econometrics connected with this, and show in the empirical part that it is possible to estimate this type of model.

In the empirical part of the paper, we compare a linear model to a Box-Cox utility function and several Taylor expansions of this function. The expansions perform consistently best. Including time and income in the model improves goodness-of-fit in general. Elasticities and values of time are also affected. What may be surprising is that not only does it matter *whether* income and time are included or not; the exact *way* this is done has significant impact not only on the model's goodness-of-fit, but also its implied elasticities and average value of time. Even the choice of point around which a Taylor expansion is made can make great difference. We also investigate the effects of restricting the non-linear parameters of the Box-Cox function and its expansions. Restricting non-linear parameters is a common practice when one is not able to directly estimate a non-linear function.

Becker (1965) seems to be the first to propose a general theory for the allocation of time and income. His framework was later refined by De Serpa (1971) and Evans (1972). Bruzelius (1979) provides a comprehensive treatment of these and other similar frameworks, with a particular focus on the value of time. González (1997) gives a theoretical review of the value of time literature and its foundations in microeconomics and household production theory. The connection from this theoretical stream of literature to applied discrete choice models is somewhat loose. An early and influential discrete choice application is Train and McFadden (1978), who study mode choice in connection with the choice of the optimal number of working hours. Jara-Díaz and Farah (1987) compare Train and McFadden's model to a model with fixed working hours, and discuss implications for welfare measures. Bates and Roberts (1986) also propose a model with fixed working hours. Truong and Hensher (1985) compare the implications of using Becker's or De Serpa's framework for mode choice models. Bates (1987) corrects a flaw in this paper regarding the possibility to measure the "resource value of time" (which we will discuss at some length in section 5).

The outline of the paper is as follows. Section 2 and 3 defines a general random utility framework for discrete choices, and section 4 discusses the possibility of transforming the conditional indirect utility. Section 5 discusses the inclusion of (travel) times in the direct utility of an alternative, and shows the implications for the value of time. In section 6, we discuss consequences and interpretations of Taylor expansions of the indirect utility function, and show that it is possible to choose expansion point freely. The empirical part of the paper starts in section 7 with presenting the data. In section 8, we define the indirect utility functions we will compare, and the indicators we will compare between the different models (elasticities and values of time). Section 9 presents estimation results, and section 10 concludes.

2 Discrete choice models

Consider the choice of one alternative out of a finite, unordered set of mutually exclusive alternatives. We can think of the choice of travel mode for a given trip, for example. Assume that each individual n gets a utility U_{in} from choosing alternative i ($i \in \{1, \ldots, I\}$), and that this utility can be separated into an observable part V_{in} and a non-observable part ε_{in}. The vectors $\boldsymbol{\varepsilon}_n = \{\varepsilon_{1n}, \ldots, \varepsilon_{In}\}$ are stochastic variables, with independent draws for each individual n[1]. This results in an *additive random utility* (ARU) *model*:

[1] It is not evident how we should treat the stochastic vector in the case when the same individual makes repeated choices in similar situations, say mode choice before and after a fuel price change. Should we assume that the stochastic vector $\boldsymbol{\varepsilon}_n$ is the same at each choice occasion, i.e. is drawn only once for each individual? Or should we assume

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

The non-observable ε_{in}:s are assumed to be known to the individual but not to the modeler. V_{in}, the measurable part of the utility function, is called the *conditional indirect utility function.* It is assumed to be a function of the (observable) characteristics of alternative i and other variables such as socioeconomic characteristics of individual n. The individual is assumed to be utility-maximizing, and thus chooses the alternative with the highest utility U_{in}. Since the modeler by assumption cannot measure the ε_{in} terms, the best we can hope to predict is P_{in}, the probability that individual n chooses alternative i. To do this, the modeler must measure V_{in} as accurately as possible, and then assume something about the distribution of ε_{in}. If the ε_{in}:s are assumed to be independent and identically Gumbel distributed, we obtain the multinomial logit (MNL) model, the most commonly used discrete choice model.

The restrictive assumption that the error terms are both independent and identically distributed can be relaxed in several ways. The most widely applied is the family of generalized extreme value (GEV) models, and in particular the nested logit model. Ben-Akiva and Lerman (1985) provides a comprehensive treatment of MNL and GEV models. Recently, even more flexible error distributions have been proposed and successfully estimated, notably the mixed multinomial logit (MMNL) model (see Train and McFadden, 1996, for its theoretical foundations, and Bhat (1998) for an application to mode and departure time choices). We will not apply any of these more flexible error specifications. However, the theoretical framework does not depend on the choice of error distribution, and most of the empirical findings can be expected to carry over to the more general models.

3 A general discrete choice framework

In order to use the additive random utility framework, we need to specify how the conditional indirect utility V_i should be measured (we will suppress the individual index n for a while). Typically, this is carried out by assuming some parametric function of the characteristics of the alternative, and the parameters are then estimated using observed or stated choices between alternatives. In this and the following sections, we will show what restrictions must be placed on this parametric function. We will do this by deriving it from a general microeconomic choice framework.

Assume that each alternative is associated with a cost c_i, a time t_i and a direct utility Q_i. This direct utility depends on the attributes of the alternative, which we

---

completely independent draws before and after the change? The answer has deep implications for estimation of and welfare evaluation with random utility models. In welfare economics, the standard answer is “yes” to the first question (see Karlström, 1999, for a discussion); in stated choice experiments, the standard answer is instead “yes” to the second one. In the present paper, we will not deal with observations of repeated choices.

collect in a vector $\mathbf{q}_i$. In the case of choice of travel mode, these attributes will include such things as waiting times, the number of changes, availability of seats and so on.

Then, a fairly general framework for this choice is the optimization problem P1 below. We assume that the utility function is separable in the direct utility Q_i and an indirect utility function $u(y,t)$, where y and t are the money and time devoted to other things than travel.

[P1.] $$\max_{y,t,W,\{\delta_i\}} u(y,t) + \sum_i \delta_i Q_i(\mathbf{q}_i) \quad (2)$$

such that

$$y + \Sigma_i\ \delta_i c_i \leq Y^0 + wW \quad (3)$$

$$t + \Sigma_i\ \delta_i t_i\ \leq T^0 - W \quad (4)$$

$$W_{min} \leq W \quad (5)$$

$$W \leq W_{max} \quad (6)$$

$$y,\ t \geq 0 \quad (7)$$

$$\delta_i \in \{0,1\}\ \forall\ i \quad (8)$$

$$\Sigma_i\ \delta_i = 1 \quad (9)$$

δ_i is 1 if and only if alternative i is chosen. T^0 and Y^0 are the available amount of time and unearned income (from savings, pensions or investments, e.g.)[2]. W is working time, and w is the wage rate. W_{min} and W_{max} are the minimum and maximum working hours. We assume that working hours can be chosen freely within these bounds. Including such bounds has important implications for the model specification. We will explore this in some detail.

3.1 Assuming constraints on working hours

Consider the problem P1 above. If we fix one $\delta_i = 1$ and $\delta_j = 0$ if $j \neq i$, we get a sub-problem consisting of choosing the optimal working hours W conditional on this choice of i. Call this sub-problem P2(i). If we assume that $u(y,t)$ is strictly increasing in y and t and that time and money are essential goods[3], we can write the first-order optimality conditions for P2(i) as

---

[2] It is not evident how "available" time and income should be defined. It is reasonable to subtract times and costs for "fixed" or "compulsory" activities, but for many activities and goods, it is not evident to what extent they are "fixed" or "compulsory"; it depends on what time scale we consider. We will return to this in the empirical section of this paper. For the moment, it is enough to keep in mind that both Y^0 and T^0 will in general be different for each individual.

[3] The first assumption ensures that the income and time constraint will be binding and that λ and μ will be strictly positive, while the second one allows us to drop the nonnegativity constraint. We will assume this throughout the paper.

$$\frac{\partial u}{\partial y} - \lambda = 0 \tag{10}$$

$$\frac{\partial u}{\partial t} - \mu = 0 \tag{11}$$

$$\lambda w - \mu + \theta_1 - \theta_2 = 0 \tag{12}$$

plus the constraints (3)-(6) above. λ, μ, θ_1 and θ_2 are the Lagrange parameters corresponding to constraints (3)-(6). From (10)-(12) it follows that

$$\frac{\partial u}{\partial t} \Big/ \frac{\partial u}{\partial y} = w + \frac{1}{\lambda}\left(\theta_1 - \theta_2\right) \tag{13}$$

The left-hand side of (13) is called the *resource value of time*. Now, it follows from standard optimization theory (the Karush-Kuhn-Tucker conditions for P2(*i*)) that θ_1 and θ_2 are always nonnegative, and that if and only if they are strictly positive, the corresponding constraint is binding[4]. That is, if and only if θ_1 is zero, the individual would not choose to work less even if we relaxed the minimum work hours constraint. Similarly, if θ_2 is zero, the individual would not choose to work more even if we relaxed the maximum working hours constraint. Hence, if the number of working hours has indeed been chosen without any (binding) constraints, the value of time will be equal to the wage rate w. If we have reasons to believe that the constraints on the number of working hours are indeed not binding, then we must restrict the model such that the value of time will be equal to the wage rate.

Since many models estimate the marginal utilities of time and money either directly or implicitly, this may suggest that we could test whether there are binding working hours constraints simply by checking whether the value of time is in fact equal to the wage rate. If the value of time is lower than the wage rate, then the individual works less than she would without the maximum work hours constraints. Conversely, if the value of time is higher than the wage rate, the individual works more than she would without the minimum working hours constraints. It may come as a surprise to realize that such a test is difficult or impossible to construct, for reasons we will explain in a moment.

In the previous literature, it appears that the *necessity* of either obtaining or assuming explicit information about the bounds W_{min} and W_{max} *before* the model is estimated has not been sufficiently stressed. Various authors have different assumptions of working hours constraints. In the theoretical literature, Becker (1965) and Evans (1972) do not incorporate these constraints in their model at all. De Serpa (1971) introduces a minimum working hours constraint. Turning to the

[4] "Binding" here means a somewhat stronger condition than only that the corresponding constraint is fulfilled with equality. When we say that a constraint is binding, we mean that the solution to the relaxed problem, the problem without this constraint, is different from the solution to the original problem. This distinction only matters when the solution to the relaxed problem happens to satisfy the constraint with equality.

more empirically oriented literature, Train and McFadden (1978) (tacitly) assume that working hours are unconstrained. Bates and Roberts (1986) and Jara-Díaz and Farah (1987) assume that working hours are fixed, i.e. $W_{min} = W_{max}$ = the observed number of working hours. The point we wish to make here is that these assumptions must be made *beforehand*, before the model is estimated, and we cannot use the model to directly reject or confirm these assumptions. This is clear from the following argument.

Let $W^*(i)$ be the solution to P2(i), the optimal number of working hours given that alternative i is chosen. We define the *conditional indirect utility* V_i as

$$V_i = u\left(Y^0 + wW^*(i) - c_i, T^0 - W^*(i) - t_i\right) + Q_i\left(\mathbf{q}_i\right) \tag{14}$$

The chosen alternative i^* will be the one with highest value of (V_i + ε_i), as described above. Assuming utility-maximizing individuals, the observed working hours and chosen alternative reveals i^* and $W^*(i^*)$. However, in order to estimate a discrete choice model, we need information about $W^*(i)$ also for the *non-chosen* alternatives. Clearly, this is impossible without assumptions about W_{min} and W_{max}. Thus, we must assume something about W_{min} and W_{max} before estimating the model, for example that the constraints will not be binding (like Train and McFadden, 1978) or that $W_{min} = W_{max}$ (like Jara-Díaz and Farah, 1987). These assumptions cannot be falsified *within the model framework*, since we, if we misspecify the model, will come up with biased parameter estimates which cannot be used for formal statistical tests. This is the reason that the condition that the value of time should be equal to the wage rate is complicated to test. It is simply because our inability to observe working hours for non-chosen alternatives forces us to assume what these would be before the model is estimated.

There is also another problem with testing if the value of time is equal to the wage rate, namely that the value of time is inherently confounded with its *quality adjustment*. We will discuss this in detail later.

In what follows, we will assume that working hours are fixed, i.e. constrained to be equal to the observed amount of working hours. This is because we will use a Swedish travel survey consisting of regional private trips, and there seems to be ample evidence from other sources that the value of time for such trips is not equal to the wage rate; in most cases, it seems to be lower. This fits with the general impression that Swedish working hours are seldom chosen freely to any large extent. It will be convenient to introduce available income and time adjusted for working time and income:

$$Y = Y^0 + wW \tag{15}$$

$$T = T^0 - W \tag{16}$$

4 Transforming the conditional indirect utility

When specifying an ARU model (1), there are two choices that must be made. The first is what the distribution of the stochastic term should be. Considerable amounts of theoretical and empirical work have been spent on this issue, often with the intent to relax the restrictive assumptions of the multinomial logit model that the error terms are independent and identically distributed. The other choice, which seems to have received less attention, is the choice of cardinal utility scale, or put it differently, the mapping from an ordinal utility measure to a cardinal one.

As stated before, we assume that the individual will choose the alternative with the highest conditional indirect utility. But choosing the "highest" utility means that only the ordinality of U_i matters. Since the ordinality of U_i is preserved under any strictly increasing transformation of the U_i:s, we may perform any such transformation. But we can evidently also transform the V_i:s holding the ARU structure (1) during this transformation, which means that we will obtain different choice probabilities for different transformations of V_i. Although the magnitudal *ordering* of the choice probabilities will not change, the *absolute magnitude* of the choice probabilities will change, and so will in general the model's elasticities. This is often a good way to improve the fit of the model.

It is in general a good idea to try various transformations of V_i. For example, problems with heteroskedasticity may be overcome or lessened with such an approach. We will not pursue this idea systematically in this paper, but there is one particular reason that we bring it up, namely that we will be interested in making Taylor expansions of V_i. It can be argued that such Taylor expansions are more reasonably viewed as *transformations* of V_i than *approximations* of it (even if this "transformation" is not always guaranteed to preserve the relative ordering of the alternatives). While it is not intuitively obvious that this might sometimes improve a model, we will see in an application below that this is in fact often the case. We will return to this issue below.

One last note on transformations: the transformation does not need to be the same for all individuals. In their influential paper, Train and McFadden (1978) propose three different indirect utiliy functions, which together with the assumption of freely chosen working hours result in three different conditional indirect utility functions. All of them are on the general form

$$V_i = w^{-\theta}c_i + w^{1-\theta}t_i \tag{17}$$

The first utility function results in a conditional indirect utility function with $\theta = 1$, the second in one with $\theta = 0$, and in the third θ is a free parameter between 0 and 1. What we wish to suggest here is that these three conditional indrect utility functions might just as well be viewed as transformations of one single conditional indrect utility function. Starting with the first suggestion, where $\theta = 0$, the other two are obtained by dividing by w or w^{θ}. In the latter case, the transformation is controlled by estimated parameter θ. Interpreting the three variants of conditional indirect utility functions as stemming from different underlying utility functions is equivalent to interpreting them as transformations of a joint underlying conditional indirect utility function. It is thus meaningless to ask what is the "correct" interpretation. In many cases, though, the "transformation" approach is helpful in order to come up with new, creative transformations, in a way the approach with different underlying utility functions is not.

5 The direct utility function and the value of time

As stated above, we assume that the direct utility of an alternative is measured by a function $Q_i(\mathbf{q}_i)$, where $\mathbf{q}_i$ is a vector of attributes of alternative *i*. In the case of mode choice, $\mathbf{q}_i$ can include such things as the number of changes (for public transit), the presence of steep climbs along the route (for bicycle) and the risk for accidents (for car). In many cases, it is advisable to include an alternative-specific constant in the function, which can be interpreted as the fixed direct utility of the alternative.

What will prove to be important from the point of model specification is the possibility to include *travel times components* in $\mathbf{q}_i$. With a "travel time component" (or simply "time component"), we will mean the time of some particular part of the trip, such as waiting time at the first bus stop, the time spent in car queues or the time spent on board the train. In a more general setting, turning from just travel to general time allocation theory, we can think of the time spent at some particular activity, such as at a cinema or doing the dishes.

In contrast, we normally do not believe that *costs* can be introduced in $\mathbf{q}_i$, since in standard microeconomics, we do not assume that there is any *direct* disutility associated with paying for something. It only constitutes a disutility to the extent that it reduces the possibility of other consumption. This means that if we can estimate the conditional indirect utility V_i, we can obtain the marginal utility of money through

$$\frac{\partial u}{\partial y} = -\frac{\partial V_i}{\partial c_i}\left(Y - c_i, T - t_i\right) \tag{18}$$

Thus, we can in principle directly measure the marginal indirect utility of money (up to a positive multiplicative constant[5]). In passing, we may also note that it follows from (18) that when we specify the V_i:s, the derivatives $\partial V_i/\partial c_i$ must have the same functional form for all *i* (even if they may be evaluated in different points $(Y\text{-}c_i,T\text{-}t_i)$). Consequently, if V_i is linear in the travel cost c_i, $\partial V_i/\partial c_i$ must be equal for all alternatives. We also note that different components of the travel cost, say fuel cost and car tolls, must have the same marginal disutility (the same parameter if we have a linear specification).

Contrary to travel costs, travel time components like waiting times or walking times may very well be a source of direct utility or disutility. For many people, standing in a crowded bus or trying to find a parking space may well constitute a disutility beyond the mere "time loss". Other parts of a trip may not cause a complete loss of time, for example train in-vehicle time, during which other activities can possibly be carried out, or walking, which at least some people are reported to enjoy.

[5] As explained above, we may replace V_i with some monotone transformation $f(V_i)$. Doing this will change the right-hand side of (18) with the positive multiplicative constant df/dV_i.

This means that the marginal utility of a reduction of a certain time component in general consists of two parts, one stemming from the indirect utility function $u(y,t)$ and one from the direct utility function $Q_i(\mathbf{q}_i)$ Formally, assume that the travel time t_i consists of components t^k_i such that $\Sigma_k\ t^k_i = t_i$. Then, using the definition of the conditional indirect utility from (14),

$$\frac{\partial V_i}{\partial t_i^k} = -\frac{\partial u}{\partial t}\left(Y - c_i, T - t_i\right) + \frac{\partial Q_i(\mathbf{q}_i)}{\partial t_i^k} \tag{19}$$

This shows that the marginal utilities of time components do not have to be equal across modes or time components. For example, if V_i is linear in each time component t^k_i, they could all have different parameters.

Combining (18) and (19), we obtain an expression for the value of time component (k,i) τ^k_i:

$$\tau_i^k \equiv \frac{\partial V_i}{\partial t_i^k} \bigg/ \frac{\partial V_i}{\partial c_i} = \frac{\partial u}{\partial t} \bigg/ \frac{\partial u}{\partial y} - \frac{\partial Q_i(\mathbf{q}_i)}{\partial t_i^k} \bigg/ \frac{\partial u}{\partial y} \tag{20}$$

The first term is called the *resource value of time*; we encountered it previously in (13) when discussing constraints on working hours. The second is called the *quality adjustment* of the value of time. De Serpa (1971) seems to be the first to make this distinction and the first to propose a model where the value of time was different for different time components. Evans (1972) clarified the case where some time components can have negative values, i.e. they are desirable. This is hardly applicable in our case, where time components are merely components of travel time, but applies to the case when time is allocated to activities.

Although it was certainly standard practice among modelers at the time to have different parameters for different time components (and different alternatives in particular), Train and McFadden (1978) did not explicitly consider the possibility of different values of time for different time components in their early and influential paper. This may have created some confusion as to whether it was “permissible” to include alternative-specific or component-specific time parameters. Since we can in principle include any time component in the direct utility $Q_i(\mathbf{q}_i)$, it is obvious from (20) that the resource value of time cannot be uniquely determined. Thus, we cannot speak about “the value of time” without specifying what “time” we mean - time waiting at a bus stop, driving a car or riding a bicycle. Truong and Hensher (1985) argued in their discussion of the implications on discrete choice econometrics of the frameworks of Becker (1965) and De Serpa (1971) that the resource value could be measured directly. This was corrected by Bates (1987), who clarified how including time components in the utility function causes the value of time to be different across time components (and/or alternatives).

6 Taylor expanding the conditional indirect utility

We may sometimes want to approximate the conditional indirect utility of time and money with its first-order Taylor approximation. Recalling the definition of the conditional indirect utility (14), we have

$$V_i = u(Y - c_i, T - t_i) + Q_i(\mathbf{q}_i) \approx$$

$$\approx u(Y,T) - \frac{\partial u}{\partial y}(Y,T)c_i - \frac{\partial u}{\partial t}(Y,T)t_i + Q_i(\mathbf{q}_i) \tag{21}$$

The first term cancels out when comparing different alternatives, so it may be dropped from V_i.

Choosing some parametric function $u(y,t)$, we can differentiate it to obtain some parametric form for $\partial u/\partial y$ and $\partial u/\partial t$, and then estimate the right-hand side of (21). But why would we want to do this? One reason could be same as when we argued that different monotone transformation of V_i should be tested. A Taylor expansion could be viewed as "almost" such a transformation, since it will mostly preserve the relative ordering of the alternatives. In the application presented here, it turns out that the Taylor expansions (there are several different variants, as will be explained below) consistently outperforms the underlying nonlinear utility function (a Box-Cox function). Not only are the goodness-of-fit better, but the results are also more robust towards misspecifications or restrictions of the model.

We cannot present any conclusive evidence as to exactly why is the case, but our experiences suggest that it is because we estimate the marginal utilities *directly* when we estimate the Taylor-expanded form, rather than estimating the underlying utility function which then gives the marginal utilities through differentiation. The former approach seems to be more robust than the latter. This could be connected to the fact that the logit model's (and most discrete choice models') point elasticity with respect to cost and time is directly proportional to the marginal utilities of money and time, respectively.

We should point out that if working hours are assumed to be chosen freely, then we should demand that $w\frac{\partial u}{\partial y}(Y,T) = \frac{\partial u}{\partial t}(Y,T)$, where w is the wage rate. This is because the derivatives in (20) are approximations to the true marginal utilities, for which we know that this condition holds (see eq. 13). Jara-Díaz and Farah (1987) observe that if working hours are assumed to be fixed and the indirect utility function is $u(y,t) = \sqrt{yt}$, then the ratio $\frac{\partial u}{\partial y}(Y,T) \Big/ \frac{\partial u}{\partial t}(Y,T)$ is equal to $Y/T = (Y^0 + wW)/(T^0 - W)$ (in our notation), which they call the expenditure rate. This can be contrasted to the wage rate w, which is by definition the working income divided by working hours. Note, though, that this only holds for the particular indirect utility function they suggest.

6.1 Choosing expansion point

In (21), we expanded the indirect utility function around (Y,T). Looking at the left-hand side, this is a natural suggestion. But we may in fact choose any arbitrary expansion point. This is easily shown by the following argument. Say that we believe that evaluating the derivatives $\partial u/\partial y$ and $\partial u/\partial t$ in another point, say (y^0,t^0), would be a better approximation of the true marginal utilities. Omitting the direct utility function, we have

$$V_i = u(Y - c_i, T - t_i) = u(Y + y^0 - y^0 - c_i, T + t^0 - t^0 - t_i) \approx$$
$$\approx u(y^0, t^0) - \frac{\partial u}{\partial y}(y^0, t^0)(Y - y^0 - c_i) - \frac{\partial u}{\partial t}(y^0, t^0)(T - t^0 - t_i) \tag{22}$$

The approximation will be "good" if the terms (Y-y^0-c_i) and (T-t^0-t_i) are "small". Comparing different alternatives, the term

$$u(y^0, t^0) - \frac{\partial u}{\partial y}(y^0, t^0)(Y - y^0) - \frac{\partial u}{\partial t}(y^0, t^0)(T - t^0) \tag{23}$$

cancels out, leaving only

$$V_i = -\frac{\partial u}{\partial y}(y^0, t^0)c_i - \frac{\partial u}{\partial t}(y^0, t^0)t_i \tag{24}$$

Apparently, we are free to choose any y^0 and t^0 that we believe will give the best approximation of the true marginal utilities.

In some circumstances, a natural candidate would be the *expected residual* income and time, defined by

$$y = Y - \sum_i P_i c_i \tag{25}$$

$$t = T - \sum_i P_i t_i \tag{26}$$

One situation where this could be a natural choice is if the discrete choice is repeated regularly, and if each choice occasion is independent in the sense that we assume that the error vector ε is redrawn each occasion. Another, more pragmatic argument is that the terms (c_i - Σ_i $P_i c_i$) and (t_i - Σ_i $P_i t_i$) can be expected to be smaller, on average, than the costs c_i and times t_i.

What is interesting with this idea, but also presents a little problem, is that the choice probabilities in (25)-(26) depend on y and t, which also appear in the left-hand side. We cannot obtain closed form expressions, but have to solve for y and t in the equation system (27)-(30):

$$y = Y - \sum_i P_i c_i \tag{27}$$

$$t = T - \sum_i P_i t_i \tag{28}$$

$$V_i = -\frac{\partial u}{\partial y}(y,t)c_i - \frac{\partial u}{\partial t}(y,t)t_i + Q_i(\mathbf{q}_i) \tag{29}$$

$$P_i = \frac{e^{V_i}}{\sum_i e^{V_i}} \tag{30}$$

To evaluate the choice probabilities, the system has to be solved for each individual. If the data set is very large it can present a significant computational burden when the parameters of V_i are estimated. Apart from possibly long running times, however, there are no principal difficulties to estimate the model. For each evaluation of the log-likelihood function, the system is solved once for each individual, and the log-likelihood function can then be maximized using standard methods. Later in this paper, we will estimate a model of this type with 5761 observations and 12 parameters, and this is by no means the limit with today's computer power.

If the data set or the number of parameters is very large, however, we can use the following idea. Instead of solving for y and t from the equation system (27)-(30), we replace the choice probabilities in (27)-(28) with choice probabilities from another model - preferably one that in some sense resembles the one in (29)-(30). A natural suggestion is the Taylor expansion around (Y,T). In all, we get this model, where the equations are evaluated top-down:

$$V_i^0 = -\frac{\partial u}{\partial y}(Y,T)c_i - \frac{\partial u}{\partial t}(Y,T)t_i + Q_i(\mathbf{q}_i) \tag{31}$$

$$P_i^0 = \frac{e^{V_i^0}}{\sum_i e^{V_i^0}} \tag{32}$$

$$y' = Y - \sum_i P_i^0 c_i \tag{33}$$

$$t' = T - \sum_i P_i^0 t_i \tag{34}$$

$$V_i' = -\frac{\partial u}{\partial y}(y',t')c_i - \frac{\partial u}{\partial t}(y',t')t_i + Q_i(\mathbf{q}_i) \tag{35}$$

$$P_i' = \frac{e^{V_i'}}{\sum_i e^{V_i'}} \tag{36}$$

We could expect P_i' in (36) to be fairly close to P_i in (30).

7 The data

The data set was taken from the Swedish National Travel Survey 1994-1997, and consisted of 5761 observations once bad observations were dropped for various reasons such as missing income or missing travel characteristics. Six modes were available: car, car passenger, bus, train, walk and bicycle. Only interzonal trips were considered, dropping intrazonal trips since we had no times or costs for these trips. Alternatives which would require more time or money than was available to the individual were set as unavailable. This mainly affected walk and bicycle. The data set is described in tables 1-5.

Mode	#Chosen	#Available
Car	3549	5740
Car passenger	808	5740
Bus	453	3749
Train	84	686
Walk	68	5088
Bicycle	799	5558

Table 1. Mode distribution

Trip type	# observations
Work	2558
School	148
Service	144
Healthcare	80
Childcare	49
Visit	1098
Leisure	591
Give a ride	226
Daily shopping	449
Other shopping	287
Other	131

Table 2. Trip types

Day of week	# observations
Monday	911
Tuesday	871
Wednesday	922
Thursday	898
Friday	964
Saturday	647
Sunday	548

Table 3. Observations per day of the week

Two different income concepts were tried. The first was gross income, i.e. total income after tax and various supports such as child and unemployment support. Individual income was taken to be half the available household income if the household consisted of two adults. The second was the net income after fixed, household-specific expenses. These expenses were taken from the Swedish Consumer Agency's guidelines (Konsumentverket, 1997), and were based on the number of adults, the number of children and their ages, whether the members of the households had lunch at home, at school or at work and whether the adults were working, students or unemployed. The expenses included costs for food, clothing, newspapers, TV, furniture and so on, and were calculated to cover basic needs for a household, without any "luxury" consumption.

All of the models were estimated once using gross income and once using net income after fixed expenses. Surprisingly, the former income concept gave better goodness-of-fit in all cases, even if the differences were fairly small in most cases. We are a little reluctant to draw any definitive conclusions from this, since the income data was of a rather poor quality. At any rate, this testing convinced us to use the gross income concept in the models reported here. The income distribution is shown in table 4.

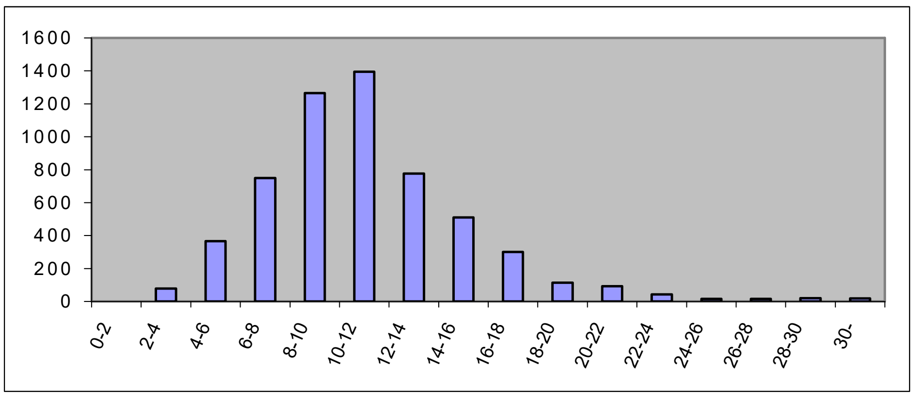

Table 4. Income distribution (1000's kr/month)

Available time was based on job type (16 hours per day minus 9 hours for full-time workers, 7 hours for students and 4.5 hours for part-time workers), with a minor adjustment for parents with children (minus 1 hour). These times were based on surveys on household time use, and in some cases direct tests. It turns out, for example, that the reduction of the available time due to children increases goodness-of-fit significantly.

8 Comparing indirect utility functions

In this section, we will estimate and compare several different specifications of the indirect utility function $u(y,t)$. The outline of this section is as follows. First, we specify the functional form of the indirect utility function $u(y,t)$ and the direct utility function $Q_i(\mathbf{q}_i)$. Then we discuss different measures that can be compared across models, such as the implied value of time. Finally, we present estimation results for several specifications of $u(y,t)$ and a discussion of the results.

8.1 Specifications of the indirect utility function

The most common specification in practical applications is the linear utility function. This is the first indirect utility function we will estimate. The estimated linear utility function will serve as a benchmark, against which we can compare the more general functional forms.

$$\begin{aligned} &u(y,t) = \alpha y + \beta t \Rightarrow V_i = \alpha(Y - c_i) + \beta(T - t_i) + Q_i(q_i) \Rightarrow \\ &V_i = -\alpha c_i - \beta t_i + Q_i(q_i) \end{aligned} \tag{37}$$

The last implication is because Y and T will cancel out when we compare different alternatives, so Y and T can be omitted from V_i. Although not without virtues (mainly its robustness to misspecifications and that it is guaranteed to render a well-behaved log-likelihood function), the linear indirect utility function obviously cannot capture effects of differences in time and income.

The next functional form we will estimate is the Box-Cox function, which contains the linear utility function as a special case when $\theta_1 = \theta_2 = 1$.

$$u(y,t) = \alpha \frac{y^{\theta_1} - 1}{\theta_1} + \beta \frac{t^{\theta_2} - 1}{\theta_2} \Rightarrow$$

$$V_i = \alpha \frac{(Y - c_i)^{\theta_1} - 1}{\theta_1} + \beta \frac{(T - t_i)^{\theta_2} - 1}{\theta_2} + Q_i(\mathbf{q}_i) \tag{38}$$

If θ_1 or θ_2 is zero, the corresponding term is defined to be $\ln y$ and $\ln t$, respectively. This makes $u(y,t)$ a continuously differentiable function for $(y,t) > 0$ for all θ_1 and θ_2. There are in principle no restrictions on the exponents θ_1 and θ_2 (for example, θ_1 or θ_2 may be negative), although we would expect decreasing marginal utility of time and income, which we have if $\theta_1, \theta_2 \leq 1$. The marginal utility of money will be positive as long as α is positive, and conversely for the marginal utility of time.

Note that it is the residual income and time y and t that enter the (indirect) utility function, rather than the cost or time directly. It is rather common to see times and costs being transformed rather than the residual time and income. While this of course would be perfectly correct if the costs and times are interpreted as quality variables, it is not consistent with standard microeconomic theory to do this if times and costs are interpreted as “loss of resources” rather than “inconvenience”, i.e. enter an indirect utility function rather than entering a direct utility function. Especially in the case of travel cost, transforming costs rather than residual income seems dubious.

The last three models we will estimate are based on the first-order Taylor expansion of the Box-Cox function. We will expand the function around three points discussed in section 6.1. The first point is the gross available after-work income Y and time T. Omitting terms independent of i, the Taylor expansion of (38) becomes

$$V_i = -\frac{\alpha}{Y^{1-\theta_1}} c_i - \frac{\beta}{T^{1-\theta_2}} t_i + Q_i(\mathbf{q}_i) \tag{39}$$

The second point we will expand around is the expected residual income y and time t, defined by

$$y = Y - \sum_i P_i c_i \tag{40}$$

$$t = T - \sum_i P_i t_i \tag{41}$$

We remind that y and t are endogenously determined, since the choice probabilities in the right-hand side depend on y and t. The third point we will expand around is an approximation of the expected residual income and time (y',t'), defined by

$$y' = Y - \sum_i P_i^0 c_i \tag{42}$$

$$t' = T - \sum_i P_i^0 t_i \tag{43}$$

The choice probabilities $P^0{}_i$ come from the model where we expand around (Y,T).

There are still only a handful of reported applications of using Box-Cox transformations in logit models. A good example is Mandel et al. (1994), although in this application, the transformation is applied directly to the time and costs of each alternative. There appears to be no applications applying Box-Cox transformations of residual time or income in discrete choice models, and no estimations of the Taylor expansions of it.

Turning to the direct utility function, we assume that it is linear in a number of quality variables and includes an alternative-specific constant:

$$Q_i(\mathbf{q}_i) = \gamma^0{}_i + \Sigma_k \gamma^k{}_i q^k{}_i \tag{44}$$

Here, the $q^k{}_i$:s are various characteristics of the mode. Several quality variables were tested for inclusion. In particular, we tried including all time components. By stepwise elimination in the linear model, the variables not significantly different from zero were eliminated. To achieve comparability between models, we will include the same quality variables in all models, even if some of them will not be significant in all models. This means that the other models are at a slight disadvantage when compared to the linear one, since this is the one we have "optimized" through our choice of quality variables.

8.2 Comparing the different models

The most interesting parameters to investigate are the exponents θ_1 and θ_2. If we cannot reject the hypothesis that one or both is equal to 1, then the differences in available income and/or time have no detectable effect.

Obviously, we will want to compare how well the different specifications perform in terms of data fit. Comparing log-likelihood values is the natural way to do this.

However, comparing log-likelihood and parameter values does not reveal to what extent different models behave differently when it comes to predictions. To get a feeling for this, we will compare the average elasticity of the car choice probability with respect to car cost and car travel time. Individual *n*:s cost elasticity is

$$\frac{\partial P_{in}}{\partial c_{in}} \frac{c_{in}}{P_{in}} = -\frac{1}{c_{in}}\left(1 - P_{in}\right)\frac{\partial u}{\partial y} \tag{45}$$

where the marginal utility of money $\partial u/\partial y$ is evaluated conditional on alternative i. (In the current application, $\partial u/\partial y$ only vary across alternatives in the Box-Cox model.) Setting i = "car", i.e. studying the car share's elasticity with respect to car cost, is rather arbitrary. What is interesting in this context is mainly that the average elasticity is a weighted average of the marginal utilities of money, since the c_{in}:s and P_{in}:s will stay the same across models. We choose car largely because it has high choice probabilities and most individuals have this choice available (see table 1). The elasticity with respect to travel (in-vehicle) time is

$$\frac{\partial P_{in}}{\partial t_{in}} \frac{t_{in}}{P_{in}} = -\frac{1}{t_{in}}\left(1 - P_{in}\right)\left(\frac{\partial u}{\partial t} + \gamma_i^k\right) \tag{46}$$

The mode subindex i is still "car", and the time component superindex k refers to "in-vehicle time".

We will calculate the average elasticity, the elasticities for the 5% with the highest car choice probability and the 5% with the lowest car choice probability, excluding those with zero car choice probability.

The next measure we will compare is the average value of time and its standard deviation. Since the value of time is in general different both across individuals, time components and modes, we must decide how the averaging should be carried out.

We will do this by taking a weighted average, where the weight of each time component is equal to the relative length of the component, and the weight of each mode is equal to its choice probability. Let μ^k_{in} be the marginal utility of a reduction of time component k for mode i and individual n (i.e. $\mu^k_{in} = -\partial u/\partial t^k_{in}$), t^k_{in} the corresponding time component (in the data set the model has been estimated on), λ_{in} her marginal utility of money (i.e. $\lambda_{in} = -\partial u/\partial c_{in}$) conditional on choosing mode i and P_{in} the predicted probability that individual n chooses mode i. Then the average value of time τ is calculated in this manner:

$$\tau_{in} = \frac{1}{\lambda_{in}} \frac{\sum_k \mu^k_{in} t^k_{in}}{\sum_k t^k_{in}} \qquad \text{(average over time components)} \tag{47}$$

$$\tau_n = \sum_i P_{in}\tau_{in} \qquad \text{(average over alternatives)} \tag{48}$$

$$\tau = \frac{1}{N}\sum_n \tau_n \qquad \text{(average over individuals)} \tag{49}$$

For the linear model (37) together with the linear direct utility function (44), it follows that $\lambda_{in} = \alpha$ and that $\mu^k_{in} = \beta - \gamma^k_i$. μ^k_{in} will thus not vary across individuals for the linear model, but may do so across modes and time components in case $\gamma^k_i \neq 0$. For the Box-Cox model from (38), μ^k_{in} and λ_{in} becomes

$$\mu^k_{in} = \frac{\partial u}{\partial t}\left(Y_n - c_{in}, T_n - t_{in}\right) - \gamma^k_i = \frac{\beta}{\left(T_n - t_{in}\right)^{1-\theta_2}} - \gamma^k_i \tag{50}$$

$$\lambda_{in} = \frac{\partial u}{\partial y}\left(Y_n - c_{in}, T_n - t_{in}\right) = \frac{\alpha}{\left(Y_n - c_{in}\right)^{1-\theta_1}} \tag{51}$$

and will obviously vary across individuals, modes and time components. Finally, for the Taylor expansion models, (39) and its variants, μ^k_{in} and λ_{in} becomes

$$\mu^k_{in} = \frac{\partial u}{\partial t}\left(Y_n, T_n\right) - \gamma^k_i = \frac{\beta}{T_n^{1-\theta_2}} - \gamma^k_i \tag{52}$$

$$\lambda_{in} = \frac{\partial u}{\partial y}\left(Y_n, T_n\right) = \frac{\alpha}{Y_n^{1-\theta_1}} \tag{53}$$

with obvious modifications for the expansions around (y,t) and (y',t').

The Box-Cox model differs from the other models in that we estimate the marginal utilities μ^k_{in} and λ_{in} implicitly, in the sense that we need to differentiate the estimated conditional indirect utility V_i to obtain them. For the other models, μ^k_{in} and λ_{in} are present in V_i explicitly.

9 Estimation results

Table 5 presents estimation results. The following abbreviations are used for the models:

linear	linear indirect utility function (eq. 37)
Box-Cox	Box-Cox indirect utility function (eq. 38)
TBC(Y,T)	Taylor expansion of the Box-Cox indirect utility function around gross income and time (Y,T) (eq. 39)
TBC(y,t)	Taylor expansion of the Box-Cox indirect utility function around expected residual income and time (y,t) (eq. 40-41)
TBC(y',t')	Taylor expansion of the Box-Cox indirect utility function around an approximation of the expected residual income and time (y',t') (eq. 42-43)

"Car const.", "bus const." etc. are the alternative-specific constants (γ^0_i in eq. 44). The bicycle constant is normalized to 0. "car/bus/train in-veh." are in-vehicle travel times, i.e. the travel time net of auxiliary trips (in the bus and train case) such as walking times to bus stop(s). "bus 1st wait" is the (average) waiting time at the first bus stop. "train aux." is the trip made to get to the first train stop or from the last, which can include walking times or going by bus to the train station. All times are in minutes, and all costs are in Swedish crowns (SEK). (1 US$ is about 8 SEK). "Log-likel." is the final log-likelihood value. "Av. VoT" is the average value of time in SEK/h and "StdD VoT" its standard deviation (across the population, not its statistical standard error). "ε_c high/aver/low" is the average car cost elasticity and the elasticity for the 5% with highest and lowest car choice probabilities (excluding zero choice probabilities). "ε_t high/aver/low" are the car time elasticities.

Most of the time components turn out to have the same marginal (dis)utility, since only a few of the time components have parameters significantly different from zero when tried for inclusion among the "quality" variables. It is interesting that the parameters for train and bus in-vehicle time are positive. This is natural if we assume that the time on-board one of these modes can be used for activities such as reading, and thus does not constitute a "complete resource loss".

The income exponent θ_1 had convergence problems in all models. The log-likelihood function appeared to be very flat in the range $-0.8 < \theta_1 < -0.7$, causing the estimates for α and θ_1 to be unreliable, reflected in large estimated standard errors. To achieve stable values for the rest of the parameters, we fixed $\theta_1 = -0.7$ and estimated the rest of the parameters conditional on this θ_1 value. Fortunately but not surprisingly, the other parameter estimates (except for α) remained essentially the same regardless of what value we chose for θ_1 (in the indicated range).

The exponent θ_2 is significantly different from 1 in all cases, showing that the differences in available time do indeed have a significant effect, as discussed above. Although the income exponent θ_1 was difficult to estimate, it was significantly different from 1. For example, the lowest log-likelihood value for the Box-Cox model was achieved for $\theta_1 = -0.84$ with an estimated standard error of 0.36.

Consequently, the goodness of fit improves when we introduce income and time in the model, and a likelihood ratio test also shows that the log-likelihood improvement is significant. The best fit is achieved by the three Taylor-expasion models.

	linear	*t-stat*	Box-Cox	*t-stat*[6]	TBC (Y,T)	*t-stat*[6]	TBC (y,t)	*t-stat*[6]	TBC (y',t')	*t-stat*[6]
alfa	0.00630	*3.0*	1.94594	*3.7*	4.13780	*5.6*	2.72831	*4.7*	3.80035	*5.2*
beta	0.01874	*31.4*	5.00747	*40.6*	10.21840	*29.4*	8.65732	*28.4*	8.79058	*28.3*
θ_1	1		-0.7		-0.7		-0.7		- 0.7	
θ_2	1		0.8822	*0.040*	0.5474	*0.053*	0.6261	*0.051*	0.6084	*0.046*
car const.	0.82846	*17.5*	1.08295	*23.8*	0.88170	*18.2*	0.89701	*18.8*	0.93286	*19.4*
pass. const.	-0.85872	*-14.8*	-0.57913	*-10.4*	-0.81800	*-14.0*	-0.79044	*-13.7*	-0.76282	*-13.2*
bus const.	0.14642	*1.4*	0.33879	*3.5*	0.39260	*3.9*	0.30198	*3.1*	0.39133	*3.9*
train const.	0.08420	*0.3*	0.29438	*1.2*	0.41820	*1.7*	0.32921	*1.4*	0.64537	*2.7*
walk const.	-1.74637	*-13.6*	-1.94224	*-15.3*	-1.77090	*-13.8*	-1.82316	*-14.3*	-1.82773	*-14.3*
car in-veh.	-0.00311	*-2.7*	-0.00278	*-2.6*	-0.00200	*-1.8*	-0.00251	*-2.3*	-0.00219	*-2.0*
bus in-veh.	0.01174	*7.4*	0.00641	*4.1*	0.00790	*5.0*	0.00941	*5.8*	0.00784	*5.1*
train in-veh.	0.00708	*2.0*	0.00597	*1.9*	0.00530	*1.6*	0.00837	*2.4*	0.00297	*0.9*
bus 1st wait.	-0.00899	*-3.0*	-0.02132	*-7.2*	-0.01240	*-4.0*	-0.01538	*-5.0*	-0.01510	*-5.0*
train aux.	0.01130	*2.7*	-0.00145	*-0.4*	0.00310	*0.8*	0.00226	*0.5*	0.00352	*0.9*
Log-likel.	-5788.3		-5770.6		-5736.6		-5746.19		-5737.2	
Av. VoT	193.23		209.85		130.55		191.55		122.81	
StdD VoT	9.0		138.2		82.9		127.19		82.1	
ε_c high	-0.046		0.064		-0.111		-0.045		-0.141	
ε_c aver.	-0.056		0.075		-0.109		-0.086		-0.136	
ε_c low	-0.157		0.175		-0.229		-0.371		-0.249	
ε_t high	-0.221		0.165		-0.229		-0.236		-0.240	
ε_t aver.	-0.252		0.189		-0.254		-0.258		-0.265	
ε_t low	-0.862		0.637		-0.873		-0.948		-0.899	

Table 5. Estimation results

It is remarkable how different the values of time and the cost and time elasticities are between the different models. For example, the average car cost elasticity is twice as high for the TBC models than for the linear and Box-Cox model. Perhaps even more remarkable is that the different TBC models give so different results in terms of value of time and elasticities, considering that they only differ through the choice of expansion point.

In order to provide some intuition for the elasticities, we plot how the car choice probability varies with varying car cost in table 9, and with car time in table 10, using sample enumeration. The car cost and car time for each individual is multiplied by a constant which varies from 0 to 3, i.e. from free/instantaneous car

[6] For θ_2, the standard error is reported instead of the t-statistic.

travel to three times today's car cost/travel time. The upper curve is the average choice probability for the 5% with highest car choice probability. The middle curve is the average choice probability. The lower curve is the car choice probability for the 5% with the lowest car choice probability. Individuals with no car availability are excluded.

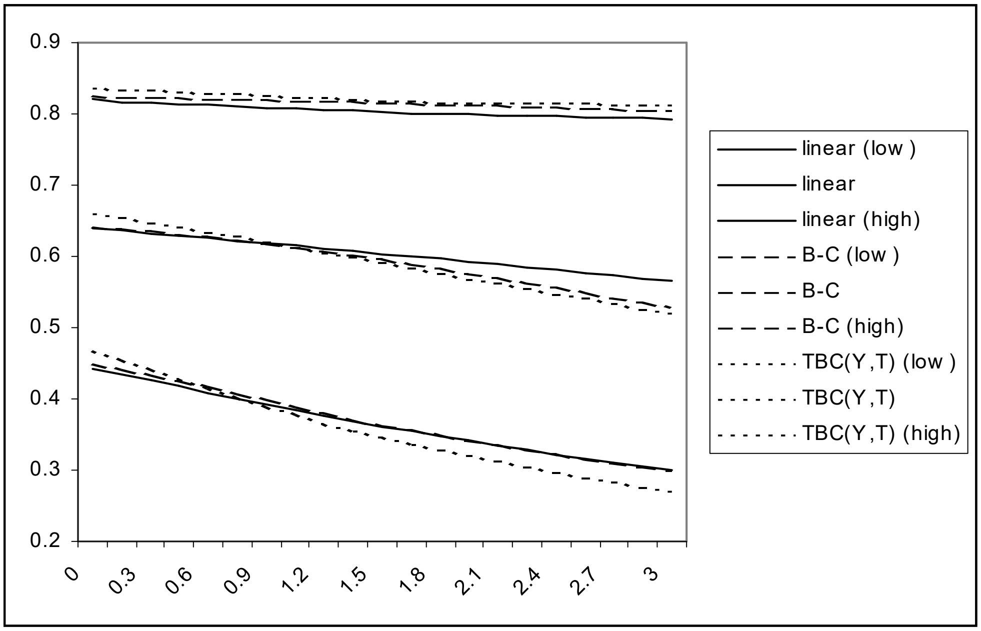

Table 6. Car choice probability with increasing car cost

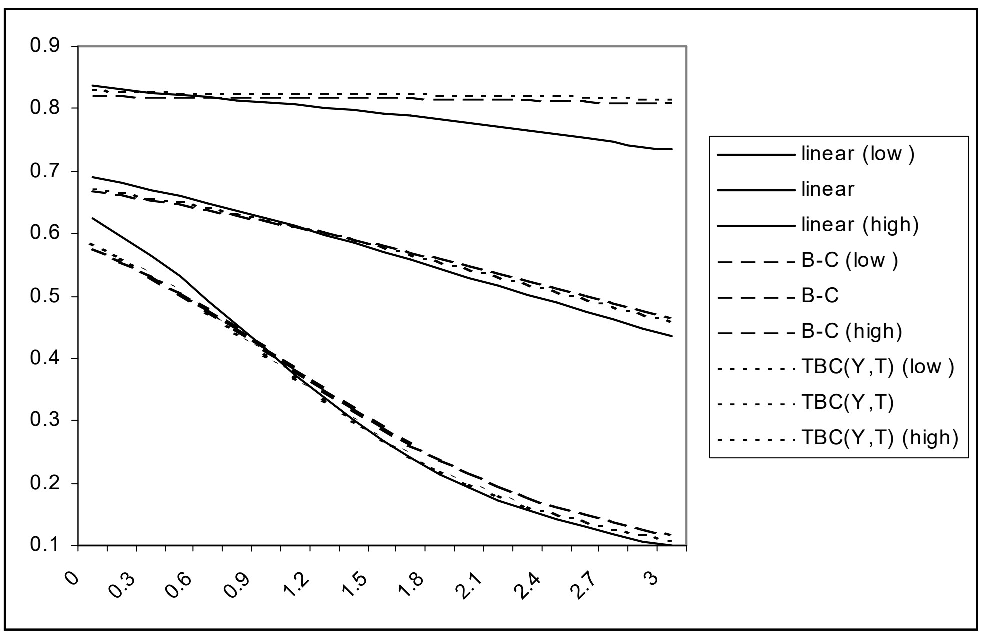

Table 7. Car choice probability with increasing car time

All models except the linear imply that the resource value of time is different across individuals. The average value of time (including quality adjustments) is slightly different across individuals also for the linear model, since it depends on the mode choice probabilities, which are different for different individuals. Table 8 and 9 shows the implied value of time distribution for the Box-Cox model and the Taylor expansion around (Y,T).

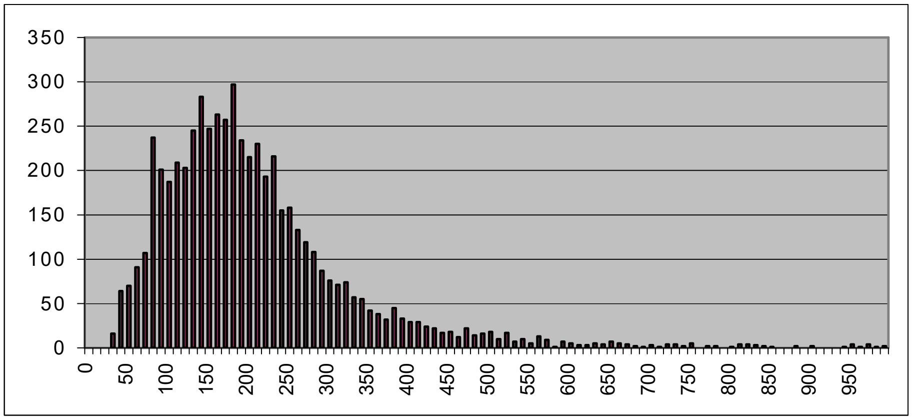

Table 8. Value of time distribution implied by the Box-Cox model (kr/h)

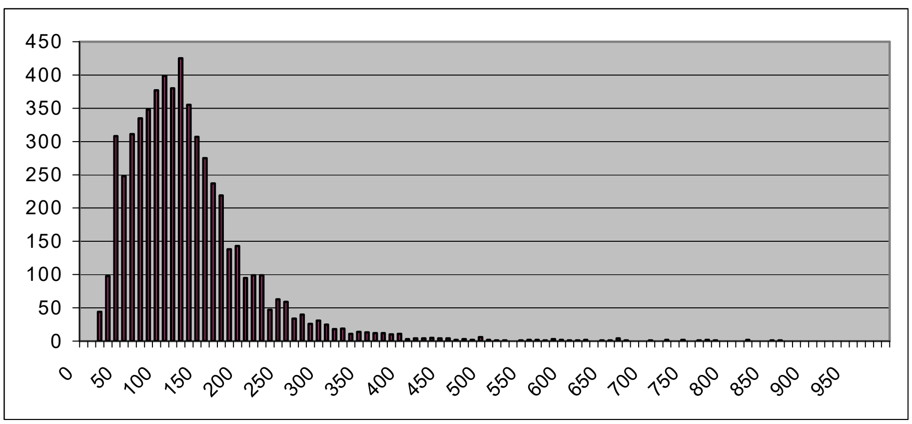

Table 9. Value of time distribution implied by TBC(Y,T) (kr/h)

9.1 Restricting the exponents

The preferable estimation approach is of course to estimate θ_1 and θ_2 together with the other parameters. However, not all estimation programs are able to estimate non-linear parameters. In this case, a few different fixed θ's can be tested to obtain some rough estimate. One disadvantage is that we will not be able to estimate the standard errors of the estimates. On the other hand, if some particular θ values prove to increase the log-likelihood value, using these values is certainly better than using the standard linear model, that is, setting $\theta_1 = \theta_2 = 1$ - which is just as arbitrary as any other θ values.

Sometimes, there may be other reasons to avoid estimating non-linear parameters, and instead choose some particular functional form in advance. This may be due to very large data sets or time constraints, or that there are theoretical or empirical results from other sources that suggests some particular functional form. With a judicious choice of function, one can obtain an indirect utility function that is able to capture effects of income and time constraints but is still reasonably easy to estimate. Common choices are logarithms and square roots. These happen to be special cases of the Box-Cox function, corresponding to $\theta_1 = \theta_2 = 0$ and $\theta_1 = \theta_2 = 0.5$.

With this background, it is interesting to study how much the models change in terms of data fit, elasticities and values of time if we restrict the exponents to these values. If the change is negligible, then this indicates that it might not be worth the extra work to go from logarithms or square roots to the general Box-Cox functions.

Just as above, we will study both the underlying indirect utility function and its first-order Taylor expansion around three points: the gross income and time (Y,T), the expected residual income and time $(y,t) = (Y\text{-}\Sigma_i\, P_i c_i, T\text{-}\Sigma_i\, P_i t_i)$ and an approximation to this $(y',t') = (Y\text{-}\Sigma_i\, P^0{}_i c_i, T\text{-}\Sigma_i\, P^0{}_i t_i)$, where the choice probabilities $P^0{}_i$ come from the first Taylor expansion.

Table 10 presents estimation results for a logarithmic indirect utility function

$$
\begin{aligned}
& u(y,t) = \alpha \ln y + \beta \ln t \Rightarrow \\
& V_i = \alpha \ln(Y - c_i) + \beta \ln(T - t_i) + \gamma_i^0 + \sum_k \gamma_i^k q_i^k
\end{aligned}
\tag{54}
$$

and its first order Taylor expansions. The one around (Y,T) becomes, dropping terms independent of i,

$$
V_i = -\frac{\alpha}{Y} c_i - \frac{\beta}{T} t_i + \gamma_i^0 + \sum_k \gamma_i^k q_i^k \tag{55}
$$

and the others are obtained by replacing (Y,T) with (y,t) and (y',t'). Table 12 also repeats the linear model for comparison.

	linear	*t-stat*	Box-Cox	*t-stat*	TBC (Y,T)	*t-stat*	TBC (y,t)	*t-stat*	TBC (y',t')	*t-stat*
alfa	0.0063	*3.0*	1.94594	*3.7*	4.1378	*5.6*	2.72831	*4.7*	3.80035	*5.2*
beta	0.01874	*31.4*	5.00747	*40.6*	10.2184	*29.4*	8.65732	*28.4*	8.79058	*28.3*
car const.	0.82846	*17.5*	1.08295	*23.8*	0.8817	*18.2*	0.89701	*18.8*	0.93286	*19.4*
pass. const.	-0.85872	*-14.8*	-0.57913	*-10.4*	-0.818	*-14*	-0.79044	*-13.7*	-0.76282	*-13.2*
bus const.	0.14642	*1.4*	0.33879	*3.5*	0.3926	*3.9*	0.30198	*3.1*	0.39133	*3.9*
train const.	0.0842	*0.3*	0.29438	*1.2*	0.4182	*1.7*	0.32921	*1.4*	0.64537	*2.7*
walk const.	-1.74637	*-13.6*	-1.94224	*-15.3*	-1.7709	*-13.8*	-1.82316	*-14.3*	-1.82773	*-14.3*
car in-veh.	-0.00311	*-2.7*	-0.00278	*-2.6*	-0.002	*-1.8*	-0.00251	*-2.3*	-0.00219	*-2*
bus in-veh.	0.01174	*7.4*	0.00641	*4.1*	0.0079	*5*	0.00941	*5.8*	0.00784	*5.1*
train in-veh.	0.00708	*2.0*	0.00597	*1.9*	0.0053	*1.6*	0.00837	*2.4*	0.00297	*0.9*
bus 1st wait.	-0.00899	*-3.0*	-0.02132	*-7.2*	-0.0124	*-4*	-0.01538	*-5*	-0.01510	*-5.0*
train aux.	0.0113	*2.7*	-0.00145	*-0.4*	0.0031	*0.8*	0.00226	*0.5*	0.00352	*0.9*
Log-likel.	-5788.30		-5897.90		-5772.60		-5790.40		-5789.60	
Av. VoT	193.23		114.84		130.19		179.82		130.24	
StdD VoT	8.98		77.19		66.47		102.74		74.80	
ε_c high	-0.046		-0.047		-0.076		-0.042		-0.079	
ε_c aver.	-0.056		-0.056		-0.087		-0.069		-0.091	
ε_c low	-0.157		-0.150		-0.226		-0.227		-0.233	
ε_t high	-0.221		-0.221		-0.223		-0.253		-0.227	
ε_t aver.	-0.252		-0.246		-0.253		-0.263		-0.266	
ε_t low	-0.862		-0.811		-0.874		-0.806		-0.956	

Table 10. Estimation results for the logarithmic indirect utility function

Table 11 presents estimation results for a square-root indirect utility function

$$u(y,t) = 2\alpha\sqrt{y} + 2\beta\sqrt{t} \Rightarrow$$
$$V_i = 2\alpha\sqrt{Y - c_i} + 2\beta\sqrt{T - t_i} + \gamma_i^0 + \sum_k \gamma_i^k q_i^k \tag{56}$$

and its first order Taylor expansions. The factor 2 is to make the function directly comparable to the Box-Cox function. The expansion around (Y,T) becomes

$$V_i = -\frac{\alpha}{\sqrt{Y}} c_i - \frac{\beta}{\sqrt{T}} t_i + \gamma_i^0 + \sum_k \gamma_i^k q_i^k \tag{57}$$

and the others are obtained by replacing (Y,T) with (y,t) and (y',t'). Table 11 also repeats the linear model for comparison.

	linear	*t-stat*	Box-Cox	*t-stat*	TBC (Y,T)	*t-stat*	TBC (y,t)	*t-stat*	TBC (y',t')	*t-stat*
alfa	0.0063	*3.0*	0.14225	*4.1*	0.20209	*4.9*	0.16715	*4.6*	0.19528	*4.7*
beta	0.01874	*31.4*	0.36093	*33.7*	0.47573	*30.7*	0.43900	*30.3*	0.44876	*29.8*
car const.	0.82846	*17.5*	0.90691	*19.4*	0.81825	*17.0*	0.82710	*17.3*	0.83802	*17.4*
pass. const.	-0.85872	*-14.8*	-0.77876	*-13.5*	-0.88154	*-15.1*	-0.86734	*-14.9*	-0.86078	*-14.7*
bus const.	0.14642	*1.4*	0.20625	*2.1*	0.25489	*2.5*	0.20057	*2.0*	0.24772	*2.4*
train const.	0.0842	*0.3*	0.15298	*0.6*	0.24375	*1.0*	0.18832	*0.8*	0.37156	*1.5*
walk const.	-1.74637	*-13.6*	-1.80304	*-14.1*	-1.72302	*-13.4*	-1.75199	*-13.7*	-1.74884	*-13.6*
car in-veh.	-0.00311	*-2.7*	-0.00259	*-2.3*	-0.00201	*-1.8*	-0.00224	*-2.0*	-0.00211	*-1.9*
bus in-veh.	0.01174	*7.4*	0.01065	*6.7*	0.01090	*6.8*	0.01214	*7.4*	0.01120	*7.0*
train in-veh.	0.00708	*2.0*	0.00802	*2.4*	0.00675	*2.0*	0.00930	*2.6*	0.00538	*1.6*
bus 1st wait.	-0.00899	*-3.0*	-0.01220	*-4.0*	-0.00791	*-2.6*	-0.00957	*-3.1*	-0.00918	*-3.0*
train aux.	0.0113	*2.7*	0.00737	*1.8*	0.00927	*2.3*	0.00900	*2.1*	0.01005	*2.5*
Log-likel.	-5788.3		-5805.81		-5756.00		-5761.30		-5757.91	
Av. VoT	193.23		126.39		129.97		149.12		130.18	
StdD VoT	8.98		39.85		34.08		42.64		37.26	
ε_c high	-0.046		-0.052		-0.088		-0.054		-0.093	
ε_c aver.	-0.056		-0.057		-0.090		-0.078		-0.096	
ε_c low	-0.157		-0.142		-0.211		-0.227		-0.217	
ε_t high	-0.221		-0.246		-0.221		-0.240		-0.225	
ε_t aver.	-0.252		-0.251		-0.238		-0.262		-0.264	
ε_t low	-0.862		-0.774		-0.817		-0.845		-0.970	

Table 11. Estimation results for the square-root indirect utility function

Restricting the exponents in this way will obviously decrease the log-likelihood values. The interesting question is first how much worse the fit of the models become, and second how the elasticities are affected. After all, these indicate how much the model forecasts will differ.

We see that once again, the Taylor expansions outperform the underlying indirect utility function. This was true when the indirect utility function was a Box-Cox function, and it still holds when we switch to a logarithmic or square root function. Moreover, the restricted versions of the Taylor-expansion models give, on the whole, similar results in terms of elasticities and values of time as their non-restricted counterparts. This is definitely not true when we compare the restricted versions of the Box-Cox function to its general version; the results change substantially.

The logarithmic model family in table 10 perform on the whole worse than the square-root model in table 11. Especially the logarithmic model (eq. 54) gives a poor fit and seemingly unreliable (or at least atypical) values of time and elasticities. Compared to the linear model, the Taylor expansions of the logarithmic indirect utility function achieve similar goodness-of-fit in two cases and a bit better in one, while the expansions of the square-root function clearly outperforms the linear model in all three cases.

The unimpressive result of the Taylor-expanded logarithmic function is interesting since this is arguably the most common way in applications to incorporate effects of time and income. Our results here indicate that dividing by the square root of Y and T gives better results than dividing by Y and T. Of course, the preferable way is to divide by $Y^{\theta 1}$ and $T^{\theta 2}$ and then estimate θ_1 and θ_2, if this is possible.

We will provide an illustration of how much all these models differ from each other by plotting their estimated average time and cost elasticities. On the *x*-axis is the time elasticity and on the *y*-axis is the cost elasticity. Each model is represented by a point in the diagram in table 12.

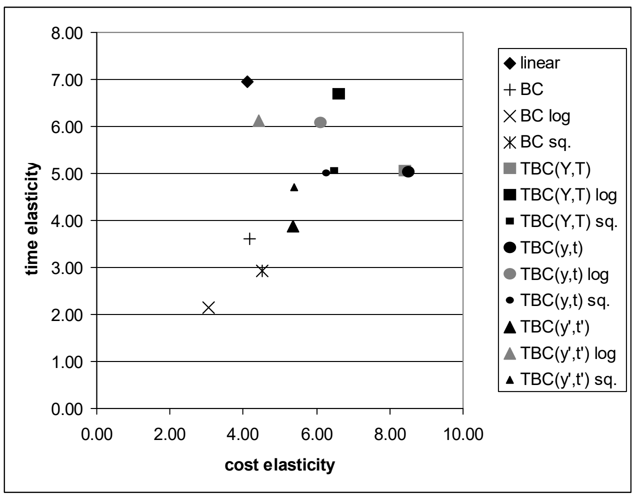

Table 12. Time and cost elasticities scatter plot.

This diagram strengthens the general impression that which model we choose has relatively large impact on the conclusions and the forecasts. Although all the models have similar or at least related functional forms, are estimated on the same data and use the same variables, values of time and elasticities are fairly different. Of course, "fairly different" is a relative measure; there is no doubt that different *samples* could give elasticities that differed from each other at least to the extent we see here. What may be a little surprising is that one single data set estimated with, after all, very similar methodologies exhibit this scattered pattern.

It is difficult to draw any general conclusions about the performance of the different models. However, it seems fair to say that the Box-Cox models, to which we count the logarithmic and square-root models, generally give smaller average elasticities than the other ones, while the linear tends to give higher time elasticities.

Turning to the values of time, these are much more different than the elasticities. This is expected, since this measure involves a quotients between parameter estimates. In table 13, the values of time are depicted. It is clear that they are scattered without any clear concentration. Even though several models give estimates around 130 kr/h, other models estimate the value of time to be around 200 kr/h.

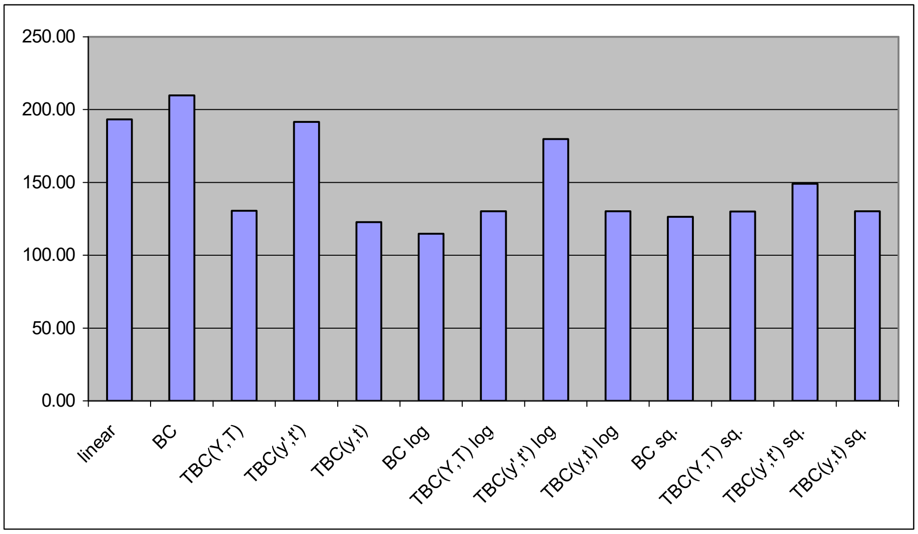

Table 13. Average values of time.

In summary, then, our main conclusions are

- including time and income effects will in general improve the goodness-of-fit of the model
- elasticities and values of time are likely to change, and especially the latter is likely to change dramatically
- the model is affected by exactly in what way time and income is included; even the choice of Taylor expansion point matters
- models which are linear in time and cost for each model, i.e. Taylor expansions of the underlying indirect utility function, tend to give more robust results.

10 Conclusions

There are at least three reasons to use utility functions able to capture income and time budget effects in demand models. The first is that this might improve the performance of the model in the sense of better likelihood values and improved predictions. The second reason is that it enables us to explicitly analyze questions such as changes in the distribution of time or income, or changes in prices on other goods. The third reason is that it improves welfare evaluations: it enables us to assess the impacts of a proposed change on different income groups, and it also enables us to evaluate proposed changes where non-constant marginal utilities of income and/or time are discernible also on the individual level.

In this paper, we have discussed how income and time can enter a discrete choice model in a microeconomically consistent way, emphasizing the necessity of assuming working time restrictions ex ante, that the possibility of including time components in a direct utility function causes the value of time for different time components to be different, and the possibility to transform the conditional indirect utility function. We have also discussed Taylor expanding the conditional indirect utility function and shown that the expansion point can be chosen freely.

In an application, we have shown that including time and income in the model improves the goodness-of-fit, and that elasticities and values of time can differ significantly depending on the way time and income is introduced. In particular, the choice of Taylor expansion point matters. In general, we found that models that were linear in travel cost and travel time, such as Taylor expansions of the conditional indirect utility function, gave better fit and were more robust.

Overall, the way that differences in available incomes and time is taken into account can make a large difference, whether the purpose is to make forecasts or to calculate cost-benefit measures. Omitting such differences is likely to yield biased parameter estimates, elasticities and values of time. The method of choice seems to be a Taylor expansion of a Box-Cox indirect utility function, but it is wise to estimate several models for comparison. In case one does not have access to software able to estimate functions that are non-linear in the parameters, or the size of the data set is prohibitive, choosing exponents ex ante yields acceptable results.

Even in cases when average elasticities or values of time change only slightly, the elasticities for different sub-groups with respect to income or time can change drastically. The results in this paper should be enough to show that it is worthwhile to take such differences into account.

Acknowledgments

Thanks to Lars-Göran Mattsson, Roger Stough and Staffan Algers for valuable comments, and to Staffan Algers also for help with the data. Financial support from the Swedish Transport and Communications Research Board is gratefully acknowledged.

11 References

Bates, J. and M. Roberts (1986) *Value of time research: summary of methodology and findings*. 14th PTRC Summer Annual Meeting, University of Sussex, UK. [as reviewed by González, 1997]

Bates, J. (1987) Measuring travel time values with discrete choice models: a note. *The Economic Journal* **97**, 493-498.

Becker (1965) A theory of the allocation of time. *The Economic Journal* **75**, 493-517.

Ben-Akiva, M. and S. Lerman (1985) *Discrete choice modeling*. MIT Press, Cambridge, Massachusetts.

Bhat, C. (1998) Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice. *Transportation Research B* **32**, 61-75.

Bruzelius, N. (1979) *The value of travel time*. Croom Helm, London, UK.

De Serpa, A. (1971) A theory of the economics of time. *The Economic Journal* **81**, 828-846.

Evans, A. (1972) On the theory of the valuation and allocation of time. *Scottish Journal of Political Economy*, February, 1-17.

Eliasson, J. (2000) The use of average values of time in road pricing. A note on a common misconception. In J. Eliasson: *Transport and Location Analysis*, diss., Dpt. of Infrastructure and Planning, Royal Institute of Technology, Stockholm.

González, R.M. (1997) The value of time: a theoretical review. *Transport Reviews* **17**, 245-266.

Jara-Diaz, S. and M. Farah (1987) Transport demand and users' benefits with fixed income: the goods-leisure trade-off revisited. *Transportation Research B* **21**, 165-170.

Karlström, A. (1999) Hicksian welfare measures in a nonlinear random utility framework. In *Four essays in spatial modelling and welfare analysis*, doctorate thesis, Department of Infrastructure and Planning, Royal Institute of Technology, Stockholm.

Konsumentverket (1997) *Koll på pengarna? Konsumentverkets beräkningar.* (*How much does it cost? Calculations of the Consumer Agency*). Report, Konsumentverket (The Swedish Consumer Agency), Stockholm.

Mandel, B., M. Gaudry and W. Rothergatter (1994) Linear or non-linear utility functions in logit models? The impact on German high-speed rail forecasts. *Transportation Research B* **28**, 91-101.

Train, K. and D. McFadden (1978) The goods/leisure tradeoff and disaggregate work trip mode choice models. *Transportation Research* **12**, 349-353.

Train, K. and D. McFadden (1996) Mixed MNL models for discrete response. Working paper available on WWW: elsa.berkeley.edu/users/mcfadden/. Forthcoming in *Journal of Applied Econometrics*.

Truong, P. and D. Henscher (1985) Measurement of travel time values and opportunity cost from a discrete choice model. *The Economic Journal* **95**, 438-451.