

Estimation of value of travel-time savings using Mixed Logit models

Stephane Hess* Michel Bierlaire† John W. Polak‡

September 7, 2004

Abstract

In this paper, we discuss some of the issues that arise with the computation of the implied value of travel-time savings in the case of discrete choice models allowing for random taste heterogeneity. We specifically look at the case of models producing a non-zero probability of positive travel-time coefficients, and discuss the consistency of such estimates with theories of rational economic behaviour. We then describe how the presence of unobserved travel-experience attributes or conjoint activities can bias the estimation of the travel-time coefficient, and can lead to false conclusions with regards to the existence of negative valuations of travel-time savings. We note that while it is important not to interpret such estimates as travel-time coefficients per se, it is nevertheless similarly important to allow such effects to manifest themselves; as such, the use of distributions with fixed bounds is inappropriate. On the other hand, the use of unbounded distributions can lead to further problems, as their shape (especially in the case of symmetrical distributions) can falsely imply the presence of positive estimates. We note that a preferable solution is to use bounded distributions where the bounds are estimated from the data during model calibration. This allows for the effects of data impurities or model misspecifications to manifest themselves, while reducing the risk of bias as a result of the shape of the distribution. To conclude, a brief application is conducted to support the theoretical claims made in the paper.

*Centre for Transport Studies, Imperial College London, stephane.hess@imperial.ac.uk, Tel: +44(0)20 7594 6105, Fax: +44(0)20 7594 6102

†Institute of Mathematics, École Polytechnique Fédérale de Lausanne, michel.bierlaire@epfl.ch, Tel: +41(0)21 693 25 37, Fax: +41(0)21 693 55 70

‡Centre for Transport Studies, Imperial College London, j.polak@imperial.ac.uk, Tel: +44(0)20 7594 6089, Fax: +44(0)20 7594 6102

1 Introduction

Random utility models have been used extensively in the field of transportation research for over thirty years. Initially, virtually all applications were based on the Multinomial Logit (MNL) model (McFadden 1974), which, although easy to implement and estimate, is limited in its scope due to a set of stringent assumptions, notably with regards to the nature of the substitution patterns across alternatives, and the assumption of a complete absence of random taste heterogeneity across decision-makers. The former restriction was eased by the introduction of a family of models known as Generalised Extreme Value (GEV) models, of which the best-known example is the Nested Logit (NL) model (Williams 1977, Daly & Zachary 1979, McFadden 1978); for an overview of existing GEV model structures, see for example Ben-Akiva & Bierlaire (2003) and Train (2003). Two other types of models, the Multinomial Probit (MNP) model (c.f. Daganzo 1979) and the Mixed Multinomial Logit (MMNL) model (c.f. McFadden & Train 2000), allow for a heightened level of flexibility by specifying the taste coefficients to be randomly distributed across decision-makers. Additionally, these models have the ability to closely replicate the correlation structure of any type of GEV model (McFadden & Train 2000 in fact show that the MMNL model can approximate the behaviour of any random utility model arbitrarily closely). Researchers have recently begun to increasingly exploit the power of the MMNL model in particular.

One specific area in which random utility models have been used repeatedly is the computation of value of travel-time savings (VTTS) measures, with some recent discussions of the topic including Algers et al. (1998), Hensher (2001a,b,c), Lapparent & de Palma (2002), Cherchi & Ortuzar (2003), Jara-Diaz & Guevara (2003), Perez et al. (2003), Cirillo & Axhausen (2004) and Sillano & Ortuzar (2004). The VTTS is an important willingness-to-pay indicator, used for example for cost-benefit analysis in the context of planning new transport systems, or for pricing. In discrete choice models, the computation of VTTS measures is relatively straightforward, especially in the case of models using linear utility functions based on fixed taste coefficients. Indeed, if the deterministic part V of the utilities in the model contains a travel-time attribute TT and a travel-cost attribute TC , the VTTS measure is simply computed as:

$$\frac{\partial V / \partial TT}{\partial V / \partial TC} \tag{1}$$

With the commonly used linear-in-variables utility function, this formula

reduces to β_{TT}/β_{TC} , where β_{TT} and β_{TC} are the time and cost coefficients, giving the marginal utilities of increases by one unit in travel-time and travel-cost respectively. Estimates of these marginal utilities are produced by calibrating the model on the choice data used in the estimation. It should be noted that the calculation in equation (1) is based on the assumption that the derivative of the unobserved part of utility with respect to travel-time and travel-cost is zero; that is, all the effects of these two factors are captured in the deterministic part.

With the increased use of the MMNL model in the area of transportation, researchers have begun to exploit the power of this model to represent a random variation in the marginal utility of travel-time across respondents. However, the extension of the theoretical foundations of the calculation of VTTS to the case where β_{TT} and/or β_{TC} are modelled as random variables is not straightforward. The objective of this paper is to highlight one of several critical issues arising in the computation of VTTS in MMNL models; the possibility of obtaining results that indicate a non-zero share of respondents with negative valuations of travel-time savings. We present a rigorous discussion that questions the validity of such results, including theoretical arguments from the econometric as well as from the microeconomic viewpoints. The theoretical arguments are supported by the results of a brief empirical application. Even though several of the issues highlighted in the paper are separately discussed in the existing literature, the authors are not aware of previous work that has integrated these considerations in the context of the estimation of VTTS.

The remainder of this paper is organised as follows. In the next section, we briefly review the theory on the random-coefficients formulation of the MMNL model, and discuss the issue of the choice of distribution for randomly distributed coefficients. In Section 3, we discuss the interpretation of results showing a non-zero probability of positive travel-time coefficients; the consistency of such estimates with economic theory is discussed in Section 4. Section 5 contains a brief application complementing the theoretical discussions presented in this paper, while Section 6 presents the conclusions of the present paper.

2 Random coefficients model

In the random-coefficients MMNL model, the parameter vector β used in the calculation of the utility is assumed to be randomly distributed rather than fixed, such that the MNL choice probability for alternative i and decision-

maker n , $P_{ni}(\beta, x_{ni})$, is replaced by:

$$P_{ni} = \int_{\beta} P_{ni}(\beta, x_{ni}) f(\beta, \Omega) d\beta, \quad (2)$$

where Ω is a vector of parameters of the distribution of the elements contained in the vector β , giving for example the means and standard deviations across the population. Three main specification issues arise with the use of the MMNL model; the selection of which parameters should be modelled as being randomly distributed across agents, the choice of statistical distribution for these coefficients, and the economic interpretation of randomly distributed coefficients. These three aspects of the specification of heterogeneity are all clearly closely inter-related. In this paper, we concentrate specifically on the latter two points and consider in particular the problems that can arise in the case where the chosen distribution allows for positive as well as negative coefficient values.

One example of a parameter for which such random taste heterogeneity has repeatedly been shown to exist is the marginal utility of travel-time (e.g. [Algers et al. 1998](#), [Cirillo & Axhausen 2004](#)). The choice of distribution for this coefficient plays a crucial role in the modelling process. Indeed, in models that are based on the use of fixed taste coefficients, researchers generally have an *a priori* expectation of obtaining a negative travel-time coefficient, and models producing positive values will normally be rejected on the grounds of model misspecification (or lack of explanatory power in the data). While the sign-issue is thus relatively straightforward in the case of fixed-coefficients models, it becomes more complicated in the case of models allowing for random taste heterogeneity. Indeed, in such models, the use of an unbounded distribution can lead to a non-zero probability of positive as well as negative travel-time parameters. It may be tempting to explain this by the notion that for some decision-makers (or for some activities), travel-time has a positive marginal utility, and there is some evidence in the literature that seems to suggest that this is indeed the case, as discussed in Section 3. However, it is not clear a priori whether model estimates showing a significant probability of a positive travel-time coefficient do in fact indicate the presence of such values in the population, or whether they are simply an artefact of the model specification or the poor quality of the data used in model estimation.

One potential source of model misspecification can come in the form of an inappropriate choice of mixing distribution for the travel-time coefficient. The distribution most commonly used in MMNL models is the *Normal* (Gaussian) distribution. The fact that the *Normal* distribution is

unbounded means that every real number has a non-zero probability of being produced as a draw; specifying a given coefficient to follow a *Normal* distribution is thus equivalent to making an *a priori* assumption that both positive and negative values for this coefficient may exist in the population. In the case where the true distribution yields strictly negative values, but has a mean close to zero with a long tail into the negative space of numbers, the symmetrical nature of the *Normal* distribution can, in approximation, lead to a significant share of positive values, even though such values are not actually *revealed* by the data (c.f. Section 5). On the other hand, if such a possibility really existed, for whatever reasons (including data impurities), the *Normal* distribution has the potential to reveal the effect. The issue with the *Normal* distribution is thus the problem of deciding whether a non-zero probability of a positive coefficient is *revealed* by the data or is simply an artefact of the symmetrical nature of the distribution.

A number of alternatives to the *Normal* distribution have been used in MMNL models, with variable success. These distributions can be split into two main groups; distributions with fixed bounds, and distributions with bounds that are estimated during the model fitting exercise.

The best known example of a distribution with a fixed bound is the *Lognormal* distribution, which is the most common choice of distribution for coefficients with an explicit sign assumption in MMNL models. While the *Lognormal* distribution has performed well in some applications (e.g. [Bhat 1998, 2000](#), [Train & Somnier 2004](#), [Hess & Polak 2004](#)), its applicability is limited for two prime reasons, its long tail on the unbounded side, and problems with slow convergence in some cases. The problem with long tails especially is a major disadvantage, given that it can for example lead to severe problems with overestimated standard deviations; as an example, [Hess & Polak \(2004\)](#) report that for one coefficient, the *Lognormal* distribution produces a mean of 5 and a standard deviation of 500. Other distributions with a fixed bound include for example the Gamma, Rayleigh and Exponential distributions. For all these distributions, a sign-change in the attribute can be used to allow for negative coefficients only.

Aside from the general problems of long tails, distributions with a fixed bound at zero lack the power to allow for counter-intuitively signed coefficients in the case where such values are *revealed* by the data, for example in the case of data impurities or other model misspecification (e.g. incomplete utility function, as described in Section 3). By ignoring the potential impacts of such problems, researchers lose information contained in the dataset, and limit the explanatory power of the model, hence leading to poorer model fit. Although, in the case of such data impurities or model misspecifications, it

is desirable not to explain a significant probability of a positive travel-time coefficient by the notion that some agents have negative valuations of travel-time savings, it is similarly bad practise to simply constrain the model to only produce negative values for β_{TT} , hence ignoring the impact of data or model imperfections.

In this case, the use of distributions bounded on either side, with bounds directly estimated from the data (i.e. through maximum likelihood estimation), is clearly preferable. With these distributions, there does, thanks to the additional left and right bounding variables, exist no a priori constraint on the domain. While this allows the distribution to be constrained to either purely positive or negative domains, it does, unlike in the case of distributions with fixed bounds, also allow for domains straddling the zero value, thus allowing data or model specification problems to manifest themselves. Moreover, the risk of values with the *wrong* sign being caused by the shape of the distribution, as with the *Normal*, largely disappears (problems may still occur in the case of a significant mass at the endpoints). A simple example of such a distribution is given by the *Triangular*, which is a generalisation of the *Uniform* distribution, allowing for a peak in the density function. The *Triangular* distribution is used rarely with MMNL models, as the linear segments between its bounds and the mode is seen as a restriction. The *Triangular* distribution however not only avoids the long tails of the *Normal* distribution, and the strict bounds of the *Lognormal* distribution, but also allows for asymmetrical shapes.

Recently, very good results have been reported with the use of Johnson’s S_B distribution (Train & Sonnier 2004). The S_B distribution can be obtained as a logit-like transformation of the *Normal* distribution, and with $\xi \sim N(\mu, \sigma)$, a draw from the S_B distribution is given by:

$$c = a + (b - a) \cdot \frac{e^\xi}{e^\xi + 1}, \quad (3)$$

where the shape of the distribution depends on the choice of μ and σ , and where c is bounded between a and b . The S_B distribution has a major advantage over other bounded distributions in that it can be used to approximate a number of very different distributions; for example, it can imitate the shape of the *Normal* and *Lognormal* distributions, with bounds on both sides, and it can also replicate Beta distributions. Furthermore, it can be specified to be symmetrical or asymmetrical, it can have a tail to the left or the right, its density can take the shape of a fairly flat plateau with drop-offs on either side, and it can also be specified to be bi-modal (c.f. Train & Sonnier 2004). While the S_B distribution is very flexible, its use leads to a need

to estimate four parameters. Furthermore, while its performance in terms of bounds is generally very good, its performance in terms of the mean and standard deviation is highly dependent on the shape of the *true* distribution, and in some cases, it can lead to significant bias in these measures (c.f. [Hess & Axhausen 2004](#)).

A further possibility is the use of an empirical distribution, whose shape reflects the actual distribution found in the sample population used in the estimation process. Another possible approach is that of censored distributions; for example, [Train & Sonnier \(2004\)](#) suggest that a *Normal* distribution censored below or above zero could be used for attributes that some respondents are indifferent to, while a strict sign assumption exists for the remainder of the population. By estimating the bound in such a distribution (i.e. not setting it to zero), modellers still allow for data impurities to manifest themselves. Finally, with the aim of allowing for a zero VTTS measure for part of the population, [Cirillo & Axhausen \(2004\)](#) propose the use of a Normal distribution with a mass at zero. For more extensive discussions of the issue of the choice of distribution, the reader is referred to [Hensher & Greene \(2003\)](#) and [Sørensen \(2003\)](#).

At this point, it is of interest to briefly discuss what should be done in the case where a model produces a significant share of positive travel-time coefficients, when appropriate precautions were taken to guarantee that this is not simply an artefact of the distributional assumptions. It is difficult to make a general recommendation, as the optimal course of action is highly dependent on the modelling issue at hand. It should be clear that it is not generally possible to determine whether the estimates are a result of poor data or an insufficient specification of the utility function. In the absence of improved data or a better understanding of the characteristics of the errors affecting existing data, the scope for addressing data problems is at this point generally very limited. Modellers are thus largely constrained to trying to improve the quality of their utility specification; here, special care should be taken to reduce the impact of correlation in the unobserved part of utility, by including any attributes that are potentially correlated with travel-time. Finally, if all attempts to obtain strictly negative travel-time coefficients fail, modellers should acknowledge the potential impact of unobservables on their estimates, and an appropriate re-labelling of the coefficients is desirable to avoid any confusion.

3 Interpretation of positive coefficients

As alluded to in the previous section, there are several potential reasons why an estimation process can yield a non-zero probability of a positive travel-time coefficient, aside from the effects of the shape of the distribution used. In this section, we described some of these reasons, after first looking at the issue of interpretation of positive travel-time and travel-cost coefficients.

At first glance, positive marginal utilities for cost and time attributes seem inconsistent with the hypothesis of rationality underlying the theory of random utility maximisation. This is particularly the case for a positive cost coefficient, where an increase of the utility would occur when the cost of the associated alternative increases. *Assuming that individuals enjoy paying more for a given good, with all other observed attributes being equal, is inconsistent with the intuitive understanding of rational economic behaviour.* If all correlated factors, such as prestige effects, were properly accounted for, the marginal utility of increases in cost should be negative and the use of unbounded distributions for the cost parameter would be inappropriate. In reality, however, such effects are generally not all explicitly accounted for, and the use of a distribution with flexible bounds may alert us to their importance in a particular empirical context.

The case of travel-time coefficients is slightly different. A negative measure for VTTS for a given individual in effect suggests that this individual would be willing to pay for increases in travel-time. At first sight, this is counter-intuitive. However, several recent papers discuss zero (Richardson 2003) or positive (Redmond & Mokhtarian 2001) elasticity with respect to travel-time. There are interesting statements like: *“I’d rather have an hour-plus commute than a five-minute commute. In the morning, it gives me a chance to work through what I’m going to do for the day. And it’s my decompression time.”* (Sipress 1999, cited by Redmond & Mokhtarian 2001). Also, the conventional interpretation of travel as a derived demand, implying a disutility for time spent travelling, may be questioned. Mokhtarian & Salomon (2001) discuss the phenomenon of undirected travel, that is cases in which travel is not a byproduct of the activity but itself constitutes the activity, and argue that this may explain the evidences of excess travel (longer than absolutely necessary travel-times) observed even in the context of mandatory journeys.

Salomon & Mokhtarian (1998) identify two possible reasons for excess travel. The first reason is the presence of unobserved objective factors. This is the case when the negative marginal utility of travel-time increases is compensated by the gains in utility resulting from simultaneously conducted

activities. The problem here is that our existing conceptual frameworks tend to lead us to think of travel and activity participation as distinct, whereas this is clearly not always the case. This topic is set to become increasingly important in the analysis of travel patterns due to the development of mobile data communication tools that massively expand the capacity for conjoining activities and travel in novel ways. The development of models that are able to analyse such conjoint activity patterns is thus an important avenue for future research.

A similar reasoning to that of conjoint activities applies in the case of desirable travel-experience factors (c.f. [Young & Morris 1981](#)). As an example, commuters walking to work may prefer a slightly longer path through a scenic park to a shorter walk through congested and polluted streets. Similarly, people may prefer to use their car for going shopping for comfort reasons, even though the presence of bus priority lanes would make for a quicker bus journey. On a related issue, the positive impact on utility of this *comfort* factor might outweigh the negative impact of the higher cost (e.g. parking fees) when compared to public transport. The impact of such unobserved attributes is related to the second reason for excess travel cited by [Salomon & Mokhtarian \(1998\)](#); namely the presence of unobserved subjective factors. As an example, the pleasure of driving an automobile, combined with the social positive perception of having and using a car, relayed by the marketing of automobiles, may explain the presence of excess travel. The impacts of such travel-experience factors can be illustrated relatively easily with the help of suitably generated synthetic data. As such, it can be shown that failing to account properly for the impact of travel-experience factors can reverse the sign of coefficients in MNL models, or significantly affect the split between positive and negative coefficients in MMNL models. Furthermore, the model estimates can falsely indicate the presence of significant random taste heterogeneity in the case where only fixed coefficients were used in the data generation process ([Hess et al. 2004](#)).

Clearly, it is often not possible to unambiguously quantify the impact of conjoint activities or travel-experience factors, and there is thus a significant risk of a biased estimate of the travel-time coefficient. Even in the case where a model produces a negative travel-time coefficient, it can be assumed that this coefficient is still biased either upwards or downwards by the failure to include some correlated attributes in the model. However, the issues described above should be considered especially in the explanation of positive travel-time coefficients (or a positive probability of such coefficient values), and researchers should strive to include as many descriptive attributes as possible, to reduce the impact of the correlation between travel-

time and unmeasured variables on the estimation of travel-time coefficients. In fact, it can be seen that by explicitly accounting for all travel-experience attributes, only the actual cost in time as a resource would remain; this would be constant across alternatives (e.g. modes or activities) for a given person at a specific moment in time, yet would most probably vary across individuals and across the time-of-day. As such, obtaining different VTTS for different alternatives in a mode-choice analysis is in fact a sign that some travel-experience attributes have not been included in the utility specification; exploring and exploiting such different VTTS measures is however often one of the main objectives of such studies. Finally, it should be noted that the issue of quantifying the impact of conjoint activities or travel-experience factors is even more difficult in the case of forecasting models.

4 Consistency with microeconomic time allocation theory

The conventional approach to estimating the VTTS from discrete choice models involves, as we have seen, calculating the marginal rate of substitution between travel-time and travel-cost, at constant utility. Although this is an intuitively plausible approach, it is important to appreciate that the justification for this approach to the valuation of travel-time savings rests not on plausibility but rather on a substantial body of microeconomic theory that addresses the issue of how individuals allocate time amongst alternative activities, including travel. It is useful to briefly review some key features of these microeconomic foundations, since they provide useful insights into the issues being considered in this paper.

The topic of time allocation and valuation has been the subject of intense study from a variety of different perspectives for several decades (see, among others, [Becker 1965](#), [Oort 1969](#), [De Serpa 1971](#), [Evans 1972](#), [Truong & Hensher 1985](#), [Bates 1987](#) and [Jara-Diaz & Guevara 2003](#)). The papers by [Jara-Diaz \(2000\)](#) and [Mackie et al. \(2001\)](#) provide excellent overviews of the development of this literature. The currently accepted position is that individuals are assumed to potentially derive utility both from the consumption of goods and from the time they spend in different activities (though of course this may vary across individuals). This is represented by a *direct* utility function that includes both goods consumed and activity time as arguments. Individuals are assumed to organise their consumption of goods and their allocation of time between activities (e.g. work, travel and leisure) such that this direct utility is maximised, subject to constraints on

the total amount of time and wealth available, and technical constraints on the minimum amount of time that it is necessary to allocate to a particular activity and/or to the consumption of a good.

The framework in this form was first crystalised in the work of [Oort \(1969\)](#) and, especially, [De Serpa \(1971\)](#), which serves as a useful point of reference for the discussion. A simple version of this framework would consider the allocation of time between say work, leisure and travel. Within this framework, DeSerpa defined three concepts of the value of time. The first is the *resource* value of time, which arises because the total amount of time available for allocation to all activities is fixed by the total time constraint. The resource value of time is equal to the ratio of the marginal utility of time and the marginal utility of income and is given by the ratio of the Lagrange multiplier associated with the total time constraint (μ) and the Lagrange multiplier associated with the income constraint (λ). The second is the value of time *allocated* to a particular activity, which arises because time (including travel-time) itself is seen as a potential source of (positive or negative) utility (not simply as a factor contributing to the production of other goods). This is equal to the rate of substitution between activity time and income in the direct utility function. The third concept is that of the value of *saving time in a particular activity*, which arises because of the technical constraints on the minimum amount of time that must be allocated to particular activities (for example in our case, the minimum time for a trip). This is equal to k/λ , where k is the Lagrange multiplier associated with the corresponding minimum travel-time constraint. It follows from the first order optimality conditions of this model that (see [Jara-Diaz 2000](#))

$$\frac{\mu}{\lambda} = \frac{\partial U/\partial L}{\partial U/\partial G} = w + \frac{\partial U/\partial W}{\partial U/\partial G} \quad (4)$$

and

$$\frac{k}{\lambda} = w + \frac{\partial U/\partial W}{\partial U/\partial G} - \frac{\partial U/\partial t}{\partial U/\partial G} \quad (5)$$

and hence that

$$\frac{k}{\lambda} = \frac{\partial U/\partial L}{\partial U/\partial G} - \frac{\partial U/\partial t}{\partial U/\partial G} \quad (6)$$

where W is the time allocated to work, L is the time allocated to leisure, G is the consumption of goods, w is the wage rate, t is the time allocated to travel, and μ , λ and k are Lagrange multipliers as defined above.

A number of authors (see [Jara-Diaz 2000](#)) have shown that the marginal rate of substitution between the time and cost parameters in the (conditional indirect) utility of a discrete choice model is precisely equal to the ratio k/λ . Hence it follows from equation (6) that the VTTS which we are considering in this paper is, from a microeconomic perspective, composed of two distinct components; the value associated with the ability to use time released by reductions in travel-time in other activities (such as work or leisure) and the value associated with the change in utility derived directly from the travel experience itself. Moreover, we should note that the Karush-Kuhn-Tucker optimality conditions guarantee that $k \geq 0$, with the equality condition (i.e., zero VTTS) applying if and only if the individual allocates more than the minimum required amount of time to the trip. For these circumstances to come about, the individual would have to derive a *positive* utility from time spent travelling at a rate exactly equal to μ/λ . That is to say, the traveller would be indifferent as between time spent in leisure and time spent travelling. Note further that in this model, there is no circumstance under which $k < 0$ could be observed. Assuming that $\lambda \geq 0$, this implies no circumstances in which a negative VTTS could be observed.

We are aware of only one recent attempt ([Jara-Diaz & Guevara 2003](#)) to disentangle these two components of the VTTS, where the empirical results reported suggested that for the sample of Chilean commuters studied, the VTTS was dominated by the strongly negative utility associated with the travel-time experience itself.

The preceding discussion has demonstrated that *if* one accepts the conventional microeconomic time allocation framework as providing an adequate basis for evaluating travel-time savings, then positive and zero values of travel-time savings are theoretically possible, but not negative ones. Of course, it could be objected that the empirical results reported in the literature regarding negative VTTS provide prima facie evidence that the existing time allocation theory is incorrect or inadequate. However, while there are certainly many respects in which the existing theory could and should be improved (see for example the discussion in [Mackie et al. 2001](#) and the recent work of [Jara-Diaz 2003](#)), we believe that on balance, it is rather more likely that some of the recent findings of negative VTTS in the literature are econometric artifacts associated with the complexities of the specification of taste heterogeneity in discrete choice models.

5 Application

It is sometimes tempting to justify the use of an unbounded distribution (e.g. *Normal*) for travel-time coefficients, and an implied positive probability of non-negative coefficient values, by the better model fit obtained with this distribution. While this is correct from a strictly mathematical point of view, it should not serve as a proof for the existence of positive travel-time coefficients and negative VTTS. Indeed, the models should rather be regarded as being misspecified; although the model allowing for a positive marginal utility of travel-time is mathematically superior, the interpretation of the coefficient as the marginal utility of travel-time is not necessarily correct.

There are two potential reasons why a better model fit can be obtained when using an unbounded distribution, such as the *Normal*. One is that its shape in the negative space of numbers might be better able to approximate the shape of the *true* distribution than is the case for any of the alternative distributions that have been tried in the estimation. The other potential reason is the existence of a positive factor that is strongly correlated with travel-time. In this application, we illustrate the impact of the distributional assumptions in the approximation to an alternative *true* distribution; the detailed analysis of the effects of unmeasured factors and conjoint activities is the topic of ongoing research.

The data used in the present analysis are based on a dataset assembled by the Canadian Rail Operator VIA Rail in 1989 to predict demand levels for a high-speed rail line in the Toronto-Montreal corridor. For a detailed description of the dataset, see [KPMG Peat Marwick & Koppelman \(1990\)](#). The sample used in the present analysis contains 4,306 observations, looking at the choice between air, car and rail. Rather than using the actual choices observed in the data, it was decided to use the attribute vectors contained in the dataset, in conjunction with a preset vector of taste parameters, to produce a set of simulated choices. This allows us to test the performance of various distributional assumptions on a dataset where the “true” values of the taste coefficients are known.

For the generation of the travel-time coefficients, a Normal distribution truncated at zero was used, thus allowing for a group of people who are indifferent to changes in travel-time (zero VTTS). However, rather than using simulation over this distribution in the calculation of the choice probabilities for the different alternatives and observations, a separate draw from this distribution was produced for each observation, leading to 4,306 individual-specific travel-time coefficients. This approach is arguably more consistent

with the interpretation of the MMNL as a model with varying taste coefficients across individuals, as it allows us to calculate distributional parameters based on the actual distribution of taste coefficients across respondents, taking the sampling into account. The draws were produced by generating 4,306 independent draws from $N(-0.0375, 0.0375)$, and by setting any positive values to zero, leading to a mean travel-time coefficient of -0.0407 , with an associated standard deviation of 0.0328, and a mass at zero of 16.5%. Fixed values were used for the alternative specific constants (ASC) for air and train, as well as for the cost and frequency coefficients (c.f. first part of table 1). For each observation, we now had a vector of taste coefficients along with a vector of explanatory attributes, and this information was used to calculate for each individual the MMNL choice probabilities for the different alternatives contained in that individual’s choice-set. A process based on random draws was then used to determine the chosen alternative with the help of the calculated choice-probabilities.

Using the estimation software BIOGEME (Bierlaire 2003), four different models were estimated on this simulated choice data; one MNL model and three MMNL models. The MNL model was estimated to illustrate the effect of not allowing for a variation in the marginal utility of travel-time across coefficients. The three MMNL models estimated on the data made different distributional assumptions with regards to β_{TT} ; with one model using a *Normal* distribution, one model using a *Lognormal* distribution, and one model using an S_B distribution. The S_B distribution was specified with additional bounding parameters a and b , as given in equation (3), and although both a and b were negative (thus implying negative values only), only b was significantly different from zero, such that a was constrained to zero, with no visible impact on model fit statistics.

For software reasons, it was at present not possible to estimate a model using the *true* distribution; a *Normal* with a mass at zero. It is in this case important to establish whether the actual distribution of the taste coefficients across the 4,306 respondents is close to the hypothetical *true* distribution, or whether it is biased by the random draws used to generate the individual-specific taste coefficients. A brief analysis showed that the impact of sampling bias plays only a minor role in this case; the actual sample distribution is virtually indistinguishable from the theoretical *true* distribution. Nevertheless, to eliminate the effects of sampling altogether, any comparative measures calculated in this analysis were based on the 4,306 individual-specific coefficients actually used, rather than on the theoretical distribution.

The results of the estimation are shown in table 1. The results show that

each of the three MMNL models leads to a very significant improvement in log-likelihood over the MNL model, by 162.64, 139.87 and 160.83 units respectively. This shows the importance of acknowledging the presence of significant levels of heterogeneity in the marginal utility of travel-time. For the three MMNL models, table 1 gives the estimated parameters of the distribution of β_{TT} , along with the implied mean and standard deviation of the coefficient. For the models using the *Lognormal* distribution, a sign change was used in the presentation of the results for the travel-time coefficient, to reflect the negative impact of the associated attribute.

The first observation that can be made from table 1 with regards to the MMNL models is that the three different distributions lead to quite similar improvements in model fit, when compared to the much poorer performance of the MNL model. The biggest improvement in model fit is obtained by the model using the *Normal* distribution, ahead of the model using the S_B distribution. Finally, the lowest log-likelihood of the three MMNL specifications is obtained by the model using a *Lognormal* distribution for the travel-time coefficient.

The next step looks at the implied willingness to pay for frequency increases, given by the negative value of the ratio between the frequency coefficient and the cost coefficient, with the true value of this ratio being equal to \$2.29. The first observation that can be made is that the MNL model considerably underestimates this ratio, at a value of \$1.07; this is a result of the overestimated cost coefficient in this model. The three MMNL models (in the order used in table 1) give values for this ratio of \$3.22, \$2.5 and \$3.26 respectively. This shows that the models based on the use of the *Normal* distribution and the S_B distribution provide point estimates which overestimate the true ratio.

Even more important differences exist across models in the estimates for the mean and standard deviation of the implied value of travel-time reductions. For the *true* coefficients, the VTTS measure was calculated for each of the 4,306 individual travel-time coefficients, and the mean and standard deviations were calculated on the basis of these values (differing only marginally from the simple ratio using the mean and standard deviation of β_{TT}). The results presented in table 1 show that the MNL model considerably underestimates the mean VTTS, which is a result of the overestimated cost coefficient along with the underestimated travel-time coefficient. The results further show that the point estimates of the three MMNL models overestimate the true mean and standard deviation. The bias is biggest for the model using a *lognormally* distributed coefficient, especially when looking at the implied standard deviation. This is a direct result of the long

	True coefficients	MNL	MMNL $\beta_{TT} \sim N(\mu, \sigma)$	MMNL $\beta_{TT} \sim LN(\mu, \sigma)$	MMNL $\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$
<i>ASC air</i>	-1.5	1.9650 (7.81)	-2.4852 (-3.78)	-2.0409 (-3.96)	-2.3393 (-4.34)
<i>ASC rail</i>	0.4	-0.1061 (-1.60)	0.4812 (3.36)	0.6930 (5.39)	0.5856 (4.47)
<i>Travel-cost</i>	-0.035	-0.0478 (-17.51)	-0.0218 (-4.04)	-0.0312 (-7.83)	-0.0247 (-5.42)
<i>Frequency</i>	0.08	0.0513 (14.31)	0.0703 (7.34)	0.0780 (9.19)	0.0804 (7.52)
<i>b</i>	-	-	-	-	-0.0771 (-4.48)
<i>Travel-time μ</i>	-0.0375	-0.0111 (-19.47)	-0.0395 (-11.56)	-3.4271 (-38.44)	0.3352 (0.48)
<i>Travel-time σ</i>	0.0375	-	0.0342 (9.34)	1.1456 (19.23)	3.0559 (3.67)
<i>Travel-time mean</i>	-0.0408	-0.0111	-0.0395	-0.0626	-0.0415
<i>Travel-time std.dev.</i>	0.0328	-	0.0342	0.1031	0.0284
<i>VTTTS mean \$(CAN)/hour</i>	69.73	13.93	108.72	120.42	100.78
<i>VTTTS std.dev. \$(CAN)/hour</i>	56.21	-	94.13	200.10	69.05
<i>Parameters</i>	-	5	6	6	7
<i>Log-likelihood</i>	-	-3171.07	-3008.43	-3031.2	-3010.24
<i>ρ^2</i>	-	0.2860	0.3227	0.3175	0.3223

Table 1: Estimates for MNL and MMNL models (t-statistics in brackets)

tail of the *Lognormal* distribution. The S_B distribution leads to the lowest overall bias, especially in the standard deviation. At this point, it should be noted (c.f. table 1) that for the model using the S_B distribution, the parameter μ of the S_B -distributed travel-time coefficient β_{TT} is not statistically significant.

The results presented in table 1 show that while the model using the *Normal* distribution leads to the best model fit, it leads to an apparently poorer performance than the model using the S_B distribution, especially in terms of recovering the true standard deviation of the VTTS. This suggests that model fit on its own may not always be an appropriate indicator of model performance, when the aim is to recover the characteristics of economic indicators such as VTTS.

In the context of the discussion presented in this paper, it is of interest not just to look at model fit and at the parameters of the distribution of the VTTS, but to also consider the bounds of the distribution. While the *Lognormal* and S_B distribution are both bounded by zero, the *Normal* distribution does, with the estimated parameters given in table 1, lead to a probability of 12.41% of a positive (non-zero) travel-time coefficient (and hence negative VTTS) despite the fact that no such strictly positive coefficient values were used in the generation of the data. This result confirms the notion described in Section 2 that the use of the *Normal* distribution can lead to false conclusions, indicating a probability of a positive travel-time coefficient when such values do not exist in the population.

To further illustrate the differences in the tail behaviour of the different distributions, 95% percentile bounds for the VTTS distribution were calculated empirically for the four models, each time making use of a sample of 1,000,000 random draws from the appropriate distribution. Corresponding bounds for the *true* distribution were calculated from the 4,306 draws actually used in the data generation. The respective limits are reproduced in table 2. The results of this analysis show the effect of allowing for positive values of β_{TT} , with a lower 95% percentile limit on the VTTS of -\$75.78 per hour when using the *Normal* distribution. Furthermore, the *Normal* distribution quite considerably overestimates the upper 95% percentile. While the *Lognormal* distribution performs well for the lower percentile, it massively overestimates the upper percentile. On the other hand, a near-perfect approximation to the *true* percentiles is obtained when using the S_B distribution.

In combination with the results from table 1, this shows that the S_B distribution leads to the best performance in recovering the *true* mean, standard deviation and upper and lower 95% percentiles. The *Normal* dis-

	Lower 95% percentile limit	Upper 95% percentile limit
<i>True</i> distribution	0	191.97
$\beta_{TT} \sim N(\mu, \sigma)$	-75.78	293.29
$\beta_{TT} \sim LN(\mu, \sigma)$	6.62	589.89
$\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$	0.65	186.95

Table 2: 95% percentile intervals for distribution of value of travel-time savings (\$/hour)

tribution performs similarly well in terms of the mean, but overestimates the standard deviation, and leads to biased lower and upper percentiles. Furthermore, it falsely indicates a significant probability of negative valuations of travel-time savings. Finally, the *Lognormal* distribution massively overestimates the standard deviation and by implication also the upper 95% percentile.

In summary, this brief application has shown that the use of the *Normal* distribution puts researchers at risk of reaching false conclusions with regards to the potential existence of positive measures of the marginal utility of travel-time and resulting negative VTTS measures. On the other hand, the *Normal* distribution does, at least in the present application, lead to an acceptable approximation of the mean and (to a lesser degree) the standard deviation, which suggests that, if researchers are not interested in the implied behaviour in the tails of the distribution, the use of the *Normal* may be acceptable. The *Lognormal* distribution avoids problems with negative VTTS, but has the disadvantage of a very heavy tail. The S_B distribution on the other hand seems to avoid all of these problems, although its use admittedly also led to an overestimation of the mean VTTS in the present analysis. Overall, these results suggest that while, in some applications, the *Normal* distribution may be used to produce an estimate of the mean and standard deviation of the VTTS across the population, it should not be used to produce estimates of the bounds of this distribution, especially so in the case where the mean value of β_{TT} is close to zero.

It should be noted that important further insights into the distribution of taste coefficients can be obtained by conducting a posterior analysis to determine the individual-specific taste coefficients conditional on the observed choices. Indeed, even in the case where the original estimation results indicate a significant probability of positive travel-time coefficients, it is conceivable that, in such an analysis, a positive coefficient would only be

associated with a very low number of respondents (c.f. [Sillano & Ortuzar 2004](#)). This further underlines the risk of misinterpretation with MMNL models, and suggests that a model indicating a non-zero probability of positive travel-time coefficients should not be used for VTTS calculation or forecasting without first conducting an appropriate posterior analysis.

6 Conclusions

The objective of this paper was to raise important issues associated with the estimation of VTTS using MMNL models. They are related to the difficulty of maintaining consistency between the theoretical assumptions on which the models are based, the actual behaviour of decision-makers, and the data collection and model specification constraints.

We acknowledge the need for more and more sophisticated models, due to empirical evidence that not everyone behaves like a “homo-economicus”. The MMNL model significantly contributes to this objective, by accounting for the effects of random taste heterogeneity. However, our results suggest that researchers should avoid the use of unbounded distributions (like the *Normal*) as a means of capturing heterogeneity in estimated time and cost coefficients, as this approach can lead to conclusions that are not supported by the data used. The *Lognormal* distribution, although more consistent with the underlying economic theory, is too strict in imposing non-positive coefficients, and has a heavy tail. The former problem can lead to a loss of information about the impact of data impurities or other specification problems (c.f. Section 2), while the latter problem can lead to seriously biased parameter estimates (c.f. Section 5). Therefore, we suggest the use of bounded distributions such as *Triangular* or S_b , **where the bounds are estimated from the data used.**

In summary, we note that under the microeconomic theory of time allocation, positive as well as zero VTTS measures are possible, but negative measures are not. In the presence of estimates showing positive travel-time coefficients (and hence negative VTTS), care should be taken to refine the model specification, notably by reducing the impact on the estimation of the travel-time coefficient of any travel-experience attributes that are strongly correlated with travel-time, as well as activities that are pursued in the same time-interval as the travelling itself. If these phenomena cannot be modelled due to the lack of explanatory power in the data, and a model with a positive time (or cost) coefficient (either as a deterministic coefficient, or as a random coefficient with a significant probability of being positive) is obtained, then

it is critical to acknowledge the limitations of the model, and to interpret it appropriately. Specifically, the name of the estimated parameter should be changed in order to emphasise that it captures more than one specific effect, and its use to compute VTTS, and/or to perform cost-benefit analysis, should be avoided. With regards to zero VTTS, we believe that the most appropriate solution is to explicitly identify the portion of the population where individuals are insensitive to travel-time savings, in the spirit of the hypothetical distribution proposed by [Cirillo & Axhausen \(2004\)](#). A latent class approach would be useful here to reconcile the economic theory with the behavioural evidence. Finally, in the present article, we have focussed solely on the population-based estimates of the distribution; as mentioned at the end of [Section 5](#), it should be noted again that individual-based parameters (e.g. by conditioning on choice) may be preferable (c.f. [Train 2003](#), [Sillano & Ortuzar 2004](#)), and further exploration of the potential of this approach is an important avenue for future research.

Acknowledgments

The authors would like to thank Kay Axhausen, John Bates, Andrew Daly, Mogens Fosgerau, Gerard de Jong, Patricia Mokhtarian, Juan de Dios Ortuzar, Goran Vuk, and two anonymous referees for comments on an earlier version of this paper.

References

- Algers, S., Bergstroem, P., Dahlberg, M. & Dillen, J. L. (1998), *Mixed Logit Estimation of the Value of Travel Time*, Working Paper, Department of Economics, Uppsala University, Uppsala, Sweden.
- Bates, J. (1987), ‘Measuring travel time values with a discrete choice model: a note’, *The Economic Journal* **97**, 493–498.
- Becker, S. G. (1965), ‘A theory of the allocation of time’, *Economic Journal* **75**(299), 493–517.
- Ben-Akiva, M. & Bierlaire, M. (2003), Discrete choice models with applications to departure time and route choice, *in* R. Hall, ed., ‘Handbook of Transportation Science, Second Edition’, Kluwer, pp. 7–37.

- Bhat, C. R. (1998), ‘Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling’, *Transportation Research* **32A**(7), 495–507.
- Bhat, C. R. (2000), ‘Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling’, *Transportation Science* **34**(2), 228–238.
- Bierlaire, M. (2003), *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.
- Cherchi, E. & Ortuzar, J. de D. (2003), *Alternative specific variables in non-linear utilities: influence of correlation, homoscedasticity and taste variations*, Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.
- Cirillo, C. & Axhausen, K. W. (2004), *Evidence on the distribution of values of travel time savings from a six-week diary*, Arbeitsbericht Verkehrs- und Raumplanung 212, IVT, ETH Zurich.
- Daganzo, C. F. (1979), *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York.
- Daly, A. & Zachary, S. (1979), Improved multiple choice models, in D. Hensher & Q. Dalvi, eds, ‘Identifying and Measuring the Determinants of Mode Choice’, Teakfields, London.
- De Serpa, A. C. (1971), ‘A theory of the economics of time’, *The Economic Journal* **81**(324), 828–846.
- Evans, A. (1972), ‘On the theory of the valuation and allocation of time’, *Scottish Journal of Political Economy*, February, pp. 1–17.
- Hensher, D. (2001a), ‘Measurement of the valuation of travel time savings’, *Journal of Transport Economics and Policy* **35**(1), 71–98.
- Hensher, D. (2001b), ‘The sensitivity of the valuation of travel time savings to the specification of unobserved effects’, *Transportation Research* **37E**, 129–142.
- Hensher, D. (2001c), ‘The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications’, *Transportation* **28**, 101–118.

- Hensher, D. & Greene, W. H. (2003), ‘The mixed logit model: the state of practice’, *Transportation* **30**(2), 133–176.
- Hess, S. & Axhausen, K. W. (2004), *Checking our assumptions in value-of-travel-time modelling: Recovering taste distributions*, CTS Working Paper, Centre for Transport Studies, Imperial College London.
- Hess, S., Bierlaire, M. & Polak, J. W. (2004), *A note on the effects of unobserved travel-experience factors on the estimation of value of travel-time savings*, CTS Working paper, Centre for Transport Studies, Imperial College London.
- Hess, S. & Polak, J. W. (2004), *Mixed Logit estimation of parking type choice*, paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC.
- Jara-Diaz, S. R. (2000), Allocation and valuation of travel time savings, in D. Hensher & K. Button, eds, ‘Handbooks in Transport, Vol. 1: Transport Modelling’, Pergamon Press, Oxford, pp. 303–319.
- Jara-Diaz, S. R. (2003), ‘On the goods-activities technical relations in time allocation theory’, *Transportation* **30**, 245–260.
- Jara-Diaz, S. R. & Guevara, C. A. (2003), ‘Behind the subjective value of travel time savings - the perception of work, leisure, and travel from a joint mode choice activity model’, *Journal of transport economics and policy* **37**, 29–46.
- KPMG Peat Marwick & Koppelman, F. S. (1990), ‘Proposals for analysis of the market demand for high speed rail in the Quebec/Ontario corridor’, Submitted to Ontario/Quebec Rapid Train Task Force.
- Lapparent, M. D. & de Palma, A. (2002), *Non-linearities in the valuations of time estimates*, Proceedings of the PTRC Annual Meeting, 2002, PTRC, London.
- Mackie, P. J., Jara-Diaz, S. R. & Fowkes, A. (2001), ‘The value of travel time savings in evaluation’, *Transportation Research* **37E**, 91–106.
- McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, in P. Zarembka, ed., ‘Frontiers in Econometrics’, Academic Press, New York.

- McFadden, D. (1978), Modelling the choice of residential location, *in* A. Karlquist, ed., ‘Spatial Interaction Theory and Planning Models’, North Holland, Amsterdam, chapter 25, pp. 75–96.
- McFadden, D. & Train, K. (2000), ‘Mixed MNL Models for discrete response’, *Journal of Applied Econometrics* **15**, 447–470.
- Mokhtarian, P. L. & Salomon, I. (2001), ‘How derived is the demand for travel? Some conceptual and measurement considerations’, *Transportation Research* **35A**, 695–719.
- Oort, O. (1969), ‘The evaluation of travelling time’, *Journal of Transport Economics and Policy* **3**, 279–286.
- Perez, P. E., Martinez, F. J. & Ortuzar, J. de D. (2003), ‘Microeconomic formulation and estimation of a residential location choice model: Implications for the value of time’, *Journal of Regional Science* **43**(4), 771–789.
- Redmond, L. S. & Mokhtarian, P. L. (2001), ‘The positive utility of the commute: modeling ideal commute time and relative desired commute amount’, *Transportation* **28**, 179–205.
- Richardson, A. J. (2003), ‘Some evidence of travelers with zero value-of-time’, *Transportation Research Record* **1854**, 107–113.
- Salomon, I. & Mokhtarian, P. L. (1998), ‘What happens when mobility-inclined market segments face accessibility-enhancing policies?’, *Transportation Research* **3D**(3), 129–140.
- Sillano, M. & Ortuzar, J. de D. (2004), *Willingness-to-pay estimation from mixed logit models: some new evidence*, Environment and Planning, forthcoming.
- Sipress, A. (1999), ‘Not all commuters driven crazy’, *Washington Post*, p. A1 and A12.
- Sørensen, M. (2003), *Discrete Choice Models: Estimation of Passenger Traffic*, PhD thesis, Centre for Traffic and Transport, Technical University of Denmark, Lyngby, Denmark.
- Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.

- Train, K. & Sonnier, G. (2004), Mixed logit with bounded distributions of correlated partworths, *in* A. Alberini & R. Scarpa, eds, ‘Applications of Simulation Methods in Environmental and Resource Economics’, Kluwer Academic Publishers, Boston, MA.
- Truong, P. & Hensher, D. (1985), ‘Measurement of travel time savings and opportunity cost from a discrete choice model’, *The Economic Journal* **95**, 438–451.
- Williams, H. C. W. L. (1977), ‘On the formulation of travel demand models and economic evaluation measures of user benefit’, *Environment and Planning* **9A**(3), 285–344.
- Young, W. & Morris, J. (1981), ‘Evaluation by individuals of their travel time to work’, *Transportation Research Record* **794**, 51–59.