Estimation of value-of-time using Mixed Logit models

Stephane Hess* Michel Bierlaire† John W. Polak‡

Technical report RO-040528
28 mai 2004

Abstract

Value-of-time is a critical willingness-to-pay indicator in many transportation applications. In this paper, we discuss the computation of this measure in the case of discrete choice models allowing for random taste heterogeneity. We first present the theoretical assumptions associated with models using randomly distributed travel-time coefficients, and highlight several important issues that must not be neglected when such an approach is adopted in practice. We then look in detail at the issue 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 note that by using an unbounded statistical distribution, positive travel-time coefficients are postulated *a priori* by the researcher, rather than being *revealed* by the data. We then illustrate how to compute the value-of-time from randomly distributed travel-time coefficients, using various experiments. Finally, we present a simple application to illustrate some concrete difficulties associated with estimating such models. Our results show that the model providing the

*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

best estimation results (in terms of recovering the true distribution of the value-of-time) is not necessarily the model giving the best fit to the data. Furthermore, our results show that the use of distributions whose behaviour in the tails is inconsistent with the intuitive understanding of the associated coefficients may sometimes lead to better estimates of the moments of the *true* distribution of the value-of-time savings.

1 Introduction

Random utility models have been used extensively in the field of transportation research for over twenty years. Initially, virtually all applications were based on the Multinomial Logit (MNL) model (Luce 1959, Marschak 1960, 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. For a review of these models, which allow for various degrees of flexibility in the representation of the inter-alternative correlation-structure, see for example Koppelman & Sethi (2000), Ben-Akiva & Bierlaire (2003) and Train (2003). Two other types of models, the Multinomial Probit (MNP) model and the Mixed Multinomial Logit (MMNL) model, 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. 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-time measures, with some recent discussions of the topic including Algers et al. (1998), Hensher (2001*a*,*b*,*c*), Lapparent & de Palma (2002), Cherchi & Ortuzar (2003), Jara-Diaz & Guevara (2003), Perez et al. (2003), Sillano & Ortuzar (2003) and Cirillo & Axhausen (2004). The value-of-time 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 value-of-time 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 value of time is simply computed as:

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

With the commonly used linear-in-parameters utility function, this formula reduces to:

$$\frac{\beta_{TT}}{\beta_{TC}} \tag{2}$$

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.

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, with some recent examples including Algers et al. (1998), Sillano & Ortuzar (2003) and Cirillo & Axhausen (2004). However, the extension of the theoretical foundations of the calculation of value-of-time measures to the case where β_{TT} and/or β_{TC} are modelled as random variables is not straightforward. In this paper, we describe the underlying assumptions associated with the use of specific random distributions for these coefficients, and discuss the internal validity of the resulting models. We especially deal with the case where the distributional assumptions lead to a non-zero probability of a positive travel-time coefficient.

The remainder of this paper is organised as follows. In Section 2, we describe the MMNL model, and present the issues associated with its specification. We discuss the interpretation of positive cost and time coefficients in Section 3, and the consistency with theories of rational behaviour in Section 4. Finally, the calculation of value-of-time with the MMNL models is addressed in Section 5, and an application showing the potential bias introduced by inappropriate distributional assumptions is described in Section 6.

2 The Mixed Logit model

The choice probabilities in the MMNL model are calculated as the integral of MNL choice probabilities over the assumed distribution of random terms. Formally, whereas in the MNL model, the utilities that decision-maker n is associating with the J alternatives contained in the choice set are modelled by a random vector

$$U_n = V_n + \varepsilon_n,$$

where $V_n \in \mathbb{R}^J$ captures the deterministic part of the utility, and ε_n is a random term, the corresponding utility in the MMNL model is given by:

$$U_n = V_n + \eta_n + \varepsilon_n.$$

In both cases, the entries of the vector ε_n are assumed to be distributed IID extreme-value over alternatives and decision-makers. But, whereas in the MNL model, this leads to a closed-form expression of the choice probabilities, the presence of the additional error term η_n in the MMNL model usually results in an integral without a closed-form expression. The mean of η_n is set to be zero, and no *a priori* constraints exist on the distribution of η_n; the researcher is free to make an appropriate and convenient choice. The resulting model form is very flexible, and free of the restrictive *Independence from Irrelevant Alternatives* (*IIA*) exhibited by the MNL model. For a given choice of distribution $f()$ for η_n, with parameter vector Ω, the MMNL choice probability is given by:

$$P_{ni} = \int_{\eta_n} \widehat{P}_{ni}\left(i|\eta_n\right) f\left(\eta_n \mid \Omega\right) \mathrm{d}\eta_n \tag{3}$$

where

$$\widehat{P}_{ni}\left(i|\eta_n\right) = \frac{e^{V_{ni}+\eta_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}+\eta_{nj}}}$$

is the probability given by the Multinomial Logit model (conditional on a given value of the vector η_n). Generally, a continuous distribution will be used for $f(\eta_n)$; discrete mixing distributions are used occasionally, leading to the "latent class model", which is popular especially in psychology and marketing (see Train 2003 for a discussion of this model).

Due to the absence of a closed-form notation for the MMNL choice-probabilities, numerical techniques, typically simulation, are required in the estimation and application of this model. The computational cost of these numerical techniques meant that the MMNL model remained largely confined to theoretical discussions for years following its development (the MMNL model was first discussed by Boyd & Mellman 1980 and Cardell & Dunbar 1980). However, with the development of ever more powerful computers and simulation techniques (c.f. Hess et al. 2003), researchers (and to a lesser degree practitioners) have recently begun to increasingly exploit the power of the MMNL model in transport demand modelling and other areas of economics.

Two distinct (though mathematically identical, as illustrated namely by Ben-Akiva & Bierlaire 2003) specifications of the MMNL model exist in the literature; the Random Coefficients Logit (RCL) formulation exploits the error structure of the MMNL model to accommodate a random distribution of tastes across decision-makers, while the Error Components Logit (ECL) formulation allows the model to approximate any GEV correlation structure arbitrarily closely. The two approaches can also be combined to allow for the joint modelling of random taste heterogeneity and flexible substitution patterns. The RCL formulation is used more regularly than its ECL counterpart, with some recent examples in the field of transportation being given by Bhat (2000), Bhat & Castelar (2002), Hess et al. (2003), and Hess & Polak (2004*a*,*b*). The ECL formulation has been used amongst others by Bhat (1998*a*) and Brownstone & Train (1999). In this paper we concentrate on issues related to the use of the RCL formulation, for more details on the ECL formulation, see Walker (2001).

With a linear-in-parameters specification of utility, the observed utility that decision-maker n obtains from choosing alternative i is given by $\beta' \cdot x_{ni}$, where β is a vector of taste coefficients and x_{ni} is a vector of attributes of alternative i, as faced by decision-maker n. In the *RCL* formulation, the parameter vector β used in the calculation of the utility is assumed to be randomly distributed rather than fixed; the error term η_n represents the deviation from the mean observed utility V_i caused by the fact that β is no longer the same for all decision-makers. Three main issues arise with the use of the MMNL model; the selection of coefficients that are to be randomly distributed across agents, the choice of statistical distribution for the different coefficients, and the economic interpretation of randomly distributed coefficients. The first of these three issues is rather straightforward, in that it can be resolved with the help of various statistical tests. Here it should

always be remembered that, while the MMNL technique offers added flexibility, each additional random term adds a dimension of integration; randomly distributed parameters should thus only be used where appropriate. However, while the use of too many random coefficients can lead to problems in estimation, a poor choice of mixing distribution can lead to downright wrong results. This issue of the choice of distribution is strongly interrelated with the issue of interpretation, and significant problems can arise especially in the case where the chosen distribution allows for positive as well as negative coefficient values.

One example of a coefficient 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 (MNL, NL, etc), 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). This issue is discussed in more detail in Section 4. While the sign-issue is relatively straightforward in the case of fixed-coefficients models, it becomes more complicated in the case of models using mixing distributions, where the use of an unbounded statistical distribution can lead to a positive probability for negative as well as positive travel-time coefficients. This thus no longer assumes that the sign of the travel-time coefficient stays negative across travellers, but suggests a positive coefficient for some proportion of the population. 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. This is the main topic of the discussion presented in this paper. To begin with, we now describe some possible distributions that can be used with MMNL models, and discuss their applicability for representing heterogeneity in the marginal utility of travel-time.

2.1 The Normal distribution

The *Normal* (Gaussian) distribution is the most commonly used distribution in MMNL models. It has been used for coefficients without a strict sign assumption (such as ASCs) as well as for coefficients where such an *a priori* assumption exists in principle (e.g. travel-time coefficients). The problems with using the *Normal* distribution arise in this latter group of coefficients. Indeed, the *Normal* distribution, with density function given by:

$$f(\xi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\xi-\mu)^2}{2\sigma^2}}, \tag{4}$$

is defined on $\xi \in (-\infty, +\infty)$, for all values of μ (mean) and σ (standard deviation). The fact that the *Normal* distribution is unbounded means that every real number has a positive 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 exist in the population. In the case of a *normally* distributed cost or travel-time coefficient, it can thus be seen that a positive probability of a non-negative coefficient is postulated by the researcher, and not necessarily *revealed* by the data.

The above discussion suggests that researchers should always attempt to use alternative distribution in the case where the model results based on the Normal distribution indicate a significant probability of a counter-intuitive sign for a coefficient where a strong a priori sign assumption exists. The estimates based on the other distributions can then be used, along with the model fit statistics of the various models, to ascertain whether this result is in fact simply an artefact of the use of the *Normal* distribution, or whether both negative and positive coefficient values actually prevail in the population. In the following subsections, we briefly describe some of these alternative distributions, and discuss their respective advantages and disadvantages. For other reviews of the different available distributions, see for example Hensher & Greene (2001), Train (2003) and Train & Sonnier (2003). Most of the distributions presented in this section generate only draws contained in the positive part of the space of real numbers; in the case where a negative coefficient is expected, the positive coefficient can simply be used with the negative of the associated attribute. Also, some of the distributions presented in this section are initially constrained to produce draws contained between 0 and 1. With ξ being a draw from such a distributions, a transformed draw

contained between a and b can be obtained through $a + (b - a) \cdot \xi$.

2.2 The Uniform Distribution

The *Uniform* distribution is the most basic statistical distribution; it assigns constant probability to all values included in its domain of definition. For ξ distributed *uniformly* on $[a, b]$, we have:

$$f(\xi) = \begin{cases} 0, & \text{for } \xi < a \\ \frac{1}{b-a}, & \text{for } a \leq \xi \leq b \\ 0, & \text{for } \xi > b \end{cases} \tag{5}$$

For the standard *Uniform* distribution, a and b are set to 0 and 1 respectively; draws from all other statistical distributions can be obtained with the help of the standard *Uniform* draws through the integral transform result. The *Uniform* distribution can be used for coefficients with an a priori sign assumption by constraining either the lower or the upper bound to 0, leading to positive, respectively negative draws only.

The *Uniform* distribution has rarely been used in the specification of MMNL models, given that it assigns equal probability to all values in its domain and does thus not allow for a peak in the distribution at the population mode. However, models based on the *Uniform* distribution are generally very easy to estimate, such that the *Uniform* distribution can at least be seen as a first step in the identification of coefficients for which significant random heterogeneity exists in the population.

2.3 The Triangular Distribution

The *Triangular* distribution is a generalisation of the *Uniform* distribution, allowing for a peak in the density function. For ξ distributed *triangularly* on $[a, b]$, with mode c, we have:

$$f(\xi) = \begin{cases} \frac{2(\xi-a)}{(b-a)(c-a)}, & \text{for } a \leq \xi \leq c \\ \frac{2(b-\xi)}{(b-a)(b-c)}, & \text{for } c \leq \xi \leq b \end{cases} \tag{6}$$

The *Triangular* distribution becomes symmetrical when $c = \frac{a+b}{2}$, in which case the mean $\mu = \frac{1}{3}(a + b + c)$ is equal to the mode. As was the case with the *Uniform* distribution, the *Triangular* distribution can be adapted to yield

positive or negative draws only. Like the *Uniform* distribution, the *Triangular* distribution is used rarely with MMNL models, as the linear segments between its bounds and the mode is seen as a restriction. However, the tent-like shape of the *Triangular* distribution can be seen as an approximation of the *Normal* distribution, with finite bounds, and with linearly decreasing probabilities either side of the mode. Furthermore, the *Triangular* distribution avoids the long tails of the *Normal* distribution, and also allows for the mean value to be different from the mode.

2.4 The Lognormal Distribution

The *Lognormal* distribution is the most common choice of distribution for coefficients with an explicit sign assumption. A variable ξ follows a Lognormal distribution if its logarithm is *normally* distributed. The domain of the distribution is the space of strictly positive real numbers, and with $ln\left(\xi\right) \sim N\left(\mu_N, \sigma_N\right)$, we have:

$$f\left(\xi\right) = \frac{1}{\sigma_N \xi \sqrt{2\pi}} \cdot e^{-(ln(\xi)-\mu_N)^2/\left(2\sigma_N^2\right)} \tag{7}$$

The mean and standard deviation of the *Lognormal* distribution can be obtained through:

$$\mu_{LN} = e^{\mu_N + \frac{\sigma_N^2}{2}}$$

and

$$\sigma_{LN} = \mu_{LN} \cdot \sqrt{e^{\sigma_N^2} - 1}$$

With $\mu_N = 0$ and $\sigma_N = 1$, the *Lognormal* distribution reduces to Gibrat's distribution. The *Lognormal* distribution has been tested extensively with MMNL models, and although it performed well in some applications (e.g. Bhat 1998*b*, 2000, Train & Sonnier 2003, Hess & Polak 2004*a*), its applicability is limited for two prime reasons. The main problem with the *Lognormal* distribution is that it is characterised by a long tail; this can lead to major problems with overestimated standard deviations. As an example, Hess & Polak (2004*a*) report that for one coefficient, the *Lognormal* distribution produces a mean of 5 and a standard deviation of 500. Another problem is that of very slow convergence of models using *lognormally* distributed coefficients, although this can be somewhat improved on by using estimates from a *Normal* distribution as starting values (c.f. Hensher & Greene 2001). Despite

its problems, the *Lognormal* distribution can be seen as being preferable to the *Normal* distribution in the case of coefficients with a strong a priori sign assumption, such as cost and time coefficients.

2.5 Johnson's S_B distribution

Recently, very good results have been reported with the use of Johnson's S_B distribution (Train & Sonnier 2003). 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:

$$c = \frac{e^{\xi}}{e^{\xi} + 1}, \tag{8}$$

where the shape of the distribution depends on the choice of μ and σ, and where c is bounded between 0 and 1. The distribution can be further adapted by replacing the 1 in the denominator by a further parameter, which then has an effect especially on the skewedness of the distribution. The fact that the S_B distribution is bounded on both sides gives it a clear advantage over the *Lognormal* distribution, as it avoids problems with thick tails. The S_B distribution has another major advantage 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 2003).

2.6 Other distributions

There are many other random distributions that could be used in MMNL models, with the Exponential distribution being just one example. Another possibility is the use of an Empirical distribution, whose shape reflects the actual distribution found in the population used in the estimation process. For a discussion of this approach, see Hensher & Greene (2001). Another possible approach is that of censored distributions; for example, Train & Sonnier (2003) suggest that a *Normal* distribution truncated below or above 0 could be used for attributes that some respondents are indifferent to, while a strict sign assumption exists for the remainder of the population. With

$F()$ giving the cumulative distribution of a univariate distribution, a draw from this distribution can be calculated as $F^{-1}(\mu)$, where μ is a draw from a standard uniform distribution (assuming that $F()$ is invertible). A draw from a version of this distribution that is truncated between a and b is then given by $F^{-1}(\bar{\mu})$, where $\bar{\mu} = (1-\mu) \cdot F(a) + \mu \cdot F(b)$. For more details on the generation of draws from such distributions, see Train (2003).

3 Interpretation of positive coefficients

Postulating unbounded distributions for travel-time and/or travel cost is equivalent to postulating that individuals belonging to an unidentified segment of the population have positive time and/or cost coefficients. It is very important to emphasise that this is a modelling decision, made a priori, and not a fact revealed by the data. We discuss here the validity of such an assumption.

At first glance, the assumption seems 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, everything else being equal, is inconsistent with the intuitive understanding of rational economic behaviour.* With all correlated factors, such as prestige effects, accounted for, the marginal utility of increases in cost should be negative. Therefore, the use of unbounded distributions for the cost parameter is clearly inappropriate.

The case of travel-time coefficients is slightly different. Several recent papers discuss zero (Richardson 2003) or positive (Redmond & Mokhtarian 2001) elasticity with respect to travel-time. There are interesting quotations 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 observed 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. One example of such a conjoined activities scenario is that of the choice between rail and air for short-haul business trips. Even though the total journey time for the *air* alternative may be inferior to that of the *rail* alternative, some travellers will, ceteris paribus, prefer the *rail* alternative. Aside from an inherent dislike of flying, a possible explanation, especially in the case of business travellers, is the fact that business travellers obtain some utility from the "continuity" of travel-time offered by the rail service, which, unlike the segmented journey by air (access-time, waiting-time, flight-time, egress-time) allows them to use the travel-time productively. 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. Indeed, in the examples discussed above, the "time" in question is not just travel-time, but also activity time; conjoining activities and travel in the same time-unit squeezes more utility out of this given unit of time. 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 this way. The development of models that are able to analyse such conjoint activity patterns is thus an important avenue for future research. Another typical example is the choice of residential location, where the commute time coefficient may be significantly positive, because there are several other factors (like the type of neighbourhood, the proximity of relatives, the proximity of schools and shops, to cite but a few) that override the cost of excess travel.

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 nice path 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 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 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 (see, for instance, Lindelof 2000).

Clearly, it is often not possible to unambiguously quantify the impact of conjoint activities or travel-experience factors (given for example the usual problems with quantifying abstract measures such as "pleasantness" or comfort), 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. The issue of quantifying the impact of conjoint activities or travel-experience factors is even more difficult in the case of forecasting models.

To illustrate the impact of correlated attributes, a brief empirical analysis was conducted. A simulated dataset of 1000 observations was constructed, giving respondents the choice between three ways of walking to work. Each alternative potentially comprises a street-level and a park-level segment, where the portion of a given path going through a park varies randomly across alternatives, with eleven different possible levels $(0, 0.1, ..., 1.0)$. For each observation, a street-level walking-time was generated, using integer values drawn *uniformly* between 5 and 15 minutes. It was then assumed that walking through a park is a detour, with the time needed to cover a unit segment being distributed *uniformly* between 1.8 and 2.2 times the equivalent street-level walking time required for that unit segment. This random variation can be seen to represent the variation in the extent of detours across alternatives and individuals. With the help of the street-level walking time, the proportion of park-time, and the park-time detour factor, a total travel-time can then be calculated for each alternative. This process was repeated for the 1000 observations used. This approach leads to high levels of correlation between the travel-time for a given alternative and the associated proportion of park-time, with, for the present dataset, an overall level of correlation (across observations and alternatives) of 0.56.

This data was then used to generate a chosen alternative for each observa-

tion. For this, two fixed taste coefficients were introduced, β_{WT}, which gives the marginal utility of one additional minute of walking-time, and β_{PP}, which gives the marginal utility of increasing the proportion of park-time from 0 to 1. The value of β_{PP} is equivalent to the ASC of a *pure-park* alternative, with the base alternative being a *pure-street-level* alternative. Corresponding values of β_{PP} for non-extreme values of the proportion attribute can be found by multiplication of β_{PP} by the appropriate value. This in effect results in 11 ASCs, with the one for a pure-street-level alternative being equal to zero. For the data generation process, the values of β_{WT} and β_{PP} were chosen to be -0.35 and 5 respectively; walking-time has a negative utility, while increases in the proportion of park-time have a positive utility. Depending on the size of the street-level walking-time, and the value of the random detour-factor, increases in the proportion of park-time can thus lead to increases or decreases in the utility of an alternative (where this utility is a function of β_{WT} and β_{PP}).

We first estimated a model using *walking-time* as well as *park-time-proportion* as explanatory variables, such that the utility of alternative i is given by $U_i = \beta_{WT} \cdot WT_i + \beta_{PP} \cdot PP_i$. This model gives a log-likelihood of -994.24, with estimates for β_{WT} and β_{PP} of -0.3175 and 4.946 respectively, thus closely reproducing the *true values* used in the generation of the data. Attempts were made to estimate a MMNL model on this dataset, using walking-time and park-time-proportion as explanatory factors. As expected, given that fixed coefficients were used in the data generation, this did not lead to any significant gains in model fit, with a new log-likelihood of -993.92, and insignificant estimates for the standard deviations of β_{WT} and β_{PP}.

Next, the model was re-estimated without the park-time-proportion coefficient, such that $U_i = \beta_{WT} \cdot WT_i$. This thus in effect removes the ASCs for the different levels of park-time-proportion, hence no longer explicitly modelling the positive impact of increases in park-time-participation. According to the discussion presented in this section, it should in this case be expected that the positive marginal utility of the park-time-proportion is added to the negative marginal utility of the walking-time coefficient, given the high level of correlation between the respective attributes. Depending on the levels of these two coefficients, the combined coefficient could be negative or positive. The removal of β_{PP} leads to a very significant drop in model fit, with a new log-likelihood of -1062.43. Furthermore, the estimated value for β_{WT} is 0.1116, and the estimate is highly significant, with a t-test value of 8.01.

This shows that the disutility of walking-time is outweighed by the utility of increases in the proportion of park-time, such that the model falsely indicates a positive marginal utility for walking time. As an extension, this model was adapted so that β_{WT} was allowed to vary *normally* across the population. This leads to an increase in the log-likelihood to -1060.39, with estimates of 0.1395 for the mean of β_{WT} and 0.1597 for its standard deviation. Although this approach suggests that the marginal utility of changes in the park-time-proportion does not universally outweigh the marginal disutility of increases in walk-time (there is a probability of 19% of a negative value for the estimated walk-time coefficient), this model shows that ignoring an attribute that is strongly correlated with walking-time can lead to the false conclusion that the walking-time coefficient varies randomly across the population. In both models (MNL and MMNL), attempts were made to include an ASC for pure-street-level alternatives; this coefficient was however insignificant in both cases, and the travel-time coefficient estimates were virtually identical to those of the models not using this ASC.

The above example has shown the potential effects of using a model that does not explicitly account for the effects of attributes that are strongly correlated with travel-time. It should be noted that this example is very basic, and that the effect may not always be as dramatic as in the present analysis. Nevertheless, the sheer scale of the effect does suggest that ignoring the potential impact of correlated attributes can lead to very misleading results. Finally, in the above example, fixed coefficient were used as the *true* values in the data-generation process; extension to the *random-coefficients* case is straightforward.

4 Consistency with economic theory

We have already noted that the occurrence of a positive cost coefficient, even for a small segment of the population, is seemingly inconsistent with the intuitive understanding of theories of rational economic behaviour. We now discuss our claim that a positive travel-time coefficient is similarly inconsistent with such theories.

We first note that the complex behaviour associated with positive aspects of travelling, discussed in the previous section, cannot be addressed by a positive travel-time coefficient in a linear-in-parameters utility function. Indeed, in this case, the utility would go to plus infinity when the travel-time

increases without bounds, such that the alternative would always be chosen; this is completely counter-intuitive, and it is highly doubtful whether this is consistent with the rationality assumptions underlying random utility theory. Clearly, the same reasoning applies in the case of a random-coefficients model producing a positive probability of a non-negative travel-time coefficient. Therefore, assuming an unbounded distribution for the travel-time coefficient, that is, assuming a positive coefficient for a segment of the population, in conjunction with a linear-in-parameters utility function, does lead to very counterintuitive conclusions. However, it is this linear-in-parameters formulation that is generally used, including in studies reporting a positive probability of non-negative travel-time coefficients (e.g. Cirillo & Axhausen 2004). It could be argued more easily that positive travel-time coefficients are possible within a narrow region, requiring the use of a non-linear specification, however it is still not clear that such values are feasible, even over a small range of travel times.

We now look at the issue of positive travel-time coefficients from the angle of general economic theory. According to this theory, time, just like money, should be regarded as a limited resource, which is used in exchange for goods or services. Once a particular unit of an economic resource has been used, it is no longer available to *pay* for other goods or services. Depending on the utility obtained from a certain good or service, the penalty associated with the incurred decrease in the resource can be seen as being smaller or larger. As an example, money spent on a fancy dinner will be seen as a more worthwhile investment than money used to pay a fine for illegal parking. Similarly, spending time watching a movie might be seen as a more worthwhile pass-time than spending time doing the washing-up. Such differences also exist in the case of travel-time; for example, time spent walking through a park on the way to work may be seen as a lesser evil than sitting in a crowded bus.

While the purchase of a given good may lead to a gain in utility for a decision-maker, it can be seen that this gain in utility is the result of the increase in utility resulting from obtaining the good outweighing the decrease in utility from spending monetary resources in the purchase of this good. In theory, it is similarly possible to split the estimated *travel-time* coefficient into an actual negative travel-time coefficient, and a set of (positive or negative) travel-experience coefficients. Indeed, no matter how large the gain in utility resulting from a purchase of a certain good, the decision-maker would prefer to obtain this utility without spending any money. Similarly, no matter how large the enjoyment of a certain activity, a rational economic

agent would prefer to obtain this enjoyment without reducing the availability of time for other activities. As such, any decrease in limited resources, such as money and time, must be seen as a negative factor. Agents will want to minimise the cost that an activity has in terms of money and time (while maximising the benefit in terms of utility), and hence the associated coefficients should be negative, although the resulting *purchase* may lead to a comparatively larger gain in utility than the decrease in utility resulting from spending money and time. Depending on the quality of the data available for modelling, random utility models can be used to explain the interaction between the negative marginal utility of decreases in economic resources (time and money) and the positive or negative marginal utility resulting from the related purchase. Again, depending on the quality of the available data, this can allow researchers to explain the increase in utility associated with excess travel-time without having to link this increase in utility to the increase in travel-time, but rather, by linking it to associated travel-experience attributes (or conjoint activities).

The above discussion can be extended in theory even to activities that an individual would seemingly prefer to last for as long as possible, such as a holiday. The observation that such activities seem to have a large positive time coefficient can be explained by the fact that the gain in personal utility from these activities is larger than that of any other activity. Nevertheless, the above reasoning still applies; if individuals were able to obtain the maximum available benefit from such an activity (thus use up all *available* activity), they would prefer to do so without affecting their resources in terms of time and money, such as to be able to turn their attention to other, although slightly less attractive, activities. This shows that the availability of time is seen as a tool for increasing utility through activities, such that any decreases in time availability should be seen as a penalty, given that they decrease the scope for conducting further activities. A possible counter-argument to this theory is the case where individuals over-indulge in one activity, such as to voluntarily reduce their time-budget with the aim of having less time available to conduct alternative, undesirable activities. However, it seems that if there was some alternative way for these individuals to avoid the undesirable activities without having to compromise their time-budget, then this would certainly be preferable to them. This again suggests that travel-time coefficients should be negative.

Again, it should be noted that such model misspecification is often not avoidable given the quality of the data, and the estimated travel-time coef-

ficients will indeed often capture the effects of a whole range of unmeasured variables. However, the potential impact of such unmeasured variable on coefficient estimates should be taken into account in the interpretation of the results, especially so in the case where the estimates indicate the presence of positive travel-time coefficients in the population. Indeed, if the best fitting model produces a travel-time coefficient estimate whose value is not consistent with the intuitive understanding of this coefficient, this does not necessarily suggest that the model is wrong, but it is in this case highly desirable not to name this coefficient "travel-time" parameter, for the reasons given above. Furthermore, such a coefficient must not be used for the calculation of accurate measures of the value of travel-time. Indeed, such a calculation requires estimates of the marginal utilities of travel-time and cost, net of any other effects, and, with the above reasoning, a positive estimate for the coefficient associated with travel-time should not be used as a travel-time coefficient. Rather, it should be seen as a combined travel-time and travel experience coefficient, giving the compound marginal utility of travel-time and any factors related to travel-time that are not explicitly included in the model.

In summary, we claim that cost and travel-time coefficients may not be positive within the framework of neoclassical economic theory in general, and random utility theory in particular. Consequently, these coefficients cannot be represented with the help of an unbounded distribution (like the *Normal*) in a RCL model. The complex behavioural issues related to positive utility of travel as an activity must be explicitly modelled, with a clear distinction in the model between the satisfaction obtained through travelling, and the actual impact of travel-time, which will probably require a nonlinear specification. A RCL model with linear-in-parameters utility functions is in no way designed to capture such complex behavioural issues.

5 Calculation of value of travel-time savings

Another major issue with randomly distributed cost and travel-time coefficients arises in the calculation of the value of time measures. Despite detailed discussions of the ill-effects of ignoring the full distribution of coefficients in the calculation of trade-offs (e.g. Hensher & Greene 2001), it is still common practise to calculate the value of time as the ratio between the mean of the travel-time coefficient and the mean of the cost coefficient. It should

be recognised that such a ratio of means is different from a mean of ratios, and that the use of the former measure can lead to serious bias in the estimated value of time. Furthermore, the ratio of means approach does not enable the calculation of the distribution of the value of time measures across decision-makers.

In the case of a coefficient that is assumed to follow a *Normal* distribution, the use of a ratio of means not only ignores the spread of values around this mean, but also ignores the potential existence of positive as well as negative values for the coefficient. This thus leads to further bias. Indeed, assuming that the cost coefficient is negative throughout the population (either fixed or following a signed distribution), the use of a Normal distribution for the travel-time coefficient will lead to negative values of travel-time for some individuals. The interpretation of such values is not clear a priori, and it is also doubtful whether negative values of travel-time are consistent with the theory of rational economic behaviour, as discussed in Section 4.

To illustrate the difference between the two approaches, and to show the effect of positive travel-time coefficients in the calculation of trade-offs, three sets of experiments were conducted, using different distributional assumptions.

In the first set of experiments, we assume that $\beta_{TC} \sim N(-6, 1.2)$ and $\beta_{TT} \sim N(-4, 0.8)$, where the units are dollars and minutes respectively. With 100,000 independent draws from either distribution, the maximum observed values were -0.5276 for β_{TC} and -0.7268 for β_{TT}, and the probability of a positive coefficient is around $2.8 \cdot 10^{-7}$ for both coefficients. When using the simple calculation based on the means of the two distributions, we obtain a value of time of \$0.67 per minute (\$40 per hour). If, on the other side, we calculate the mean over the ratios of each pair of coefficients, we obtain a value of time of \$0.6972 per minute (\$41.83 per hour). While these values are very similar, the use of the simple ratio of means approach leads to a loss of all information concerning the distribution of the value of time across the population. A calculation of the variance for the ratio of coefficients in the population yields a value of 0.04866, such that a 95% quantile interval of the value of time per minute is given by $[0.2648, 1.1296]$, or equivalently, lower and upper quantiles for the distribution of the value of time of \$15.89 and \$67.77 per hour respectively. Clearly, ignoring this spread in values leads to an important loss of information. Also, with the above specification, and in the absence of correlation between β_{TC} and β_{TT}, the maximum observed measure of value of time is an unrealistically high \$585.51 per hour, which is

a direct result of the long tails of the *Normal* distribution used for the two coefficients.

The calculation of the value of time becomes slightly more complicated in the case where the two attributes of travel-time and travel cost are correlated. As an example, suppose that the covariance between β_{TC} and β_{TT} is equal to 0.2, which, with the above specified variances, leads to a correlation between β_{TC} and β_{TT} of 0.21. The two coefficients are now distributed according to a multivariate *Normal* distribution, and the draws can be produced quite easily using a Choleski transformation (c.f. Train 2003). To remove any additional source of error, the same random draws that were used in the calculation of the draws in the first part of the example were reused here. The value of time produced by the ratio of means approach remains unchanged at \$40 per hour, while the value resulting from the mean of ratios approach is slightly lower than in the first part of the example, at \$41.45 per hour, with an associated 95% quantile interval of $[18.71, 64.19]$. The incorporation of correlation also reduces the impact of the tails of the *Normal* distribution on the calculation, with the largest outlier now being a value of time of \$496.21 per hour. Still, this value is very high, and again, a failure to acknowledge the variation in the distribution of the value of time leads to a significant loss of information.

In the second set of experiments, we choose the parameters of the distribution of β_{TT} so as to lead to a significant probability of a positive travel-time coefficient. For this, we assume that $\beta_{TT} \sim N(-4, 3.13)$, meaning that $P(\beta_{TT} > 0) \simeq 10\%$. We leave the distribution of β_{TC} unchanged at $N(-6, 1.2)$. The estimated value of time using the ratio of means approach remains unchanged at \$40 per hour, given that the coefficient mean values have been kept the same. We first assume that the two coefficients are uncorrelated, leading to a mean value of time of \$41.78 per minute (using the mean of ratios approach), with an associated 95% quantile interval of $[-27.38, 110.95]$. The range of this interval highlights the effect of allowing for a positive travel-time coefficient, and furthers the doubt about the validity of using the *Normal* distribution, given that the results indicate a lower 95% quantile at a value of time of -\$27.38 per hour, and a probability of a negative value of time of around 10%. We conclude the second set of experiments by looking at the case where β_{TC} and β_{TT} are correlated, using the same level of correlation as in the first experiment. Due to the higher variance used for β_{TT} (when compared to the first experiment), it now requires a covariance of 0.79 to obtain a correlation of 0.21. This approach again brings the estimated value of time closer to that produced by the ratio of

means approach, with the mean of ratios approach giving a value of \$40.27 per hour, while the 95% interval is now equal to $[-24.56, 105.09]$. This example illustrates the risk of not explicitly looking at the distribution of the value of time across the population. Indeed, in this example, the mean of ratio and ratio of means approach produce virtually the same mean value of time, which might suggest that the choice of approach is not important. The differences only become visible by looking at the full distribution of values, where the mean of ratios approach reveals a huge spread of values of time, with minimum and maximum values of -\$224.57 and \$613.79 respectively.

To further illustrate the effects that the distributional assumptions for time and cost coefficients have on the distribution of the value of time, plots were produced that show the density of the travel and cost coefficients as well as the value of time for the two examples described above. In both cases, the specifications incorporating correlation between β_{TT} and β_{TC} were used. The resulting plots are shown in figure 1, with the plots in the first row referring to the example with strictly positive values of time, and the plots in the second row referring to the example allowing for negative values of time. The plots in the first row show a much longer tail to the right in the distribution of the value of time, which is a result of it being bounded at zero due to the use of strictly negative time and cost coefficients. The plots for the second example show that the tail of the value of time to the *right* of the mean is in this case far less pronounced, although the maximum value of time is still 2.17 times further away from the mean value than is the case for the minimum value (the corresponding value in the first example is 13.6). Finally, the plots in the second row clearly show the effect of allowing for positive time coefficients, with a probability of around 10% of a negative value of time.

The two brief examples described above have shown that the use of the ratio of means approach leads to a small bias in the calculation of the value of time. This bias has been shown to be larger in the case of asymmetrical distributions, such as *Lognormal* (c.f. Hess & Polak 2004*a*). While this bias in the mean values of time might be acceptable in some cases, the fact that the use of the ratio of means approach leads to a complete loss of information about the effects of the distribution of coefficients on the distribution of the value of time in the population constitutes a major disadvantage for this method.

As an extension, a further experiment was conducted in which the distribution of the cost coefficient was adapted so as to allow for a small prob-

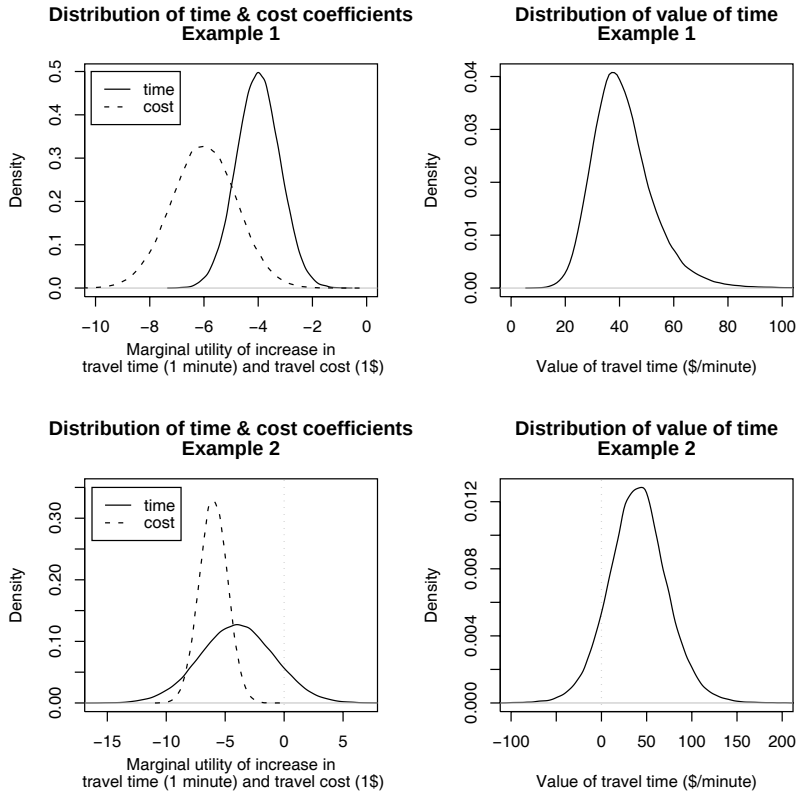

Figure 1: Distribution of time and cost coefficients and implied value of time measures

ability of a positive cost coefficient. Although positive cost coefficients are clearly counter-intuitive, a positive probability of a non-negative cost coefficient could result from a poorly specified model. One example is the use of a *Normal* distribution in the case of a negative cost coefficient with a large standard deviation. In our example, we maintain $\beta_{TT} \sim N(-4, 3.13)$ and set $\beta_{TC} \sim N(-6, 3.65)$, leading to a probability of a positive cost coefficient of 5%. Again, the correlation between β_{TC} and β_{TT} was kept at 0.21. The aim of this example is to show the effect of cost coefficients that are close to zero in the calculation of value of time measures. While the value of time resulting from the ratio of means approach remains unchanged at \$40 per hour, the value resulting from the mean of ratios approach is now \$51.94 per hour, showing a significant bias with the other approach. With the above

specifications, the 95% quantile interval is extremely wide, with limits of -$4682.856 and $4786.74, where the extreme values are a direct result of a division by a cost coefficient that is close to zero. This example shows that important problems can arise with models that use a *Normal* distribution for the cost coefficient, or for that matter, any distribution that allows values that are very close to zero. With the specification used for this example, there is a 13.3% probability of a negative value of time. To illustrate the distribution of the value of time with this specification, the kernel of the distribution was estimated for the range $[-500, 500]$, which contains 97.6% of the simulated values. A density plot for this region is shown in figure 2. The plot illustrates the narrow peak around the mean value of time measure, with long tails to either side, where the tails are equally long, but where more mass (65.88%) is placed to the left of the mean, with the range between 0 and $51.94 containing some 52.58% of the mass.

In the example described above, a correlation level of 0.21 was used between β_{TT} and β_{TC}. With higher correlation levels, the tails of the distribution of the value of time would have been lower, while, with lower correlation levels, even longer tails would be expected. Indeed, from a theoretical viewpoint, the ratio of two independent *normally* distributed random variables with zero mean follows a Cauchy distribution. The Cauchy distribution is uni-modal and symmetric, with much heavier tails than the *Normal* distribution. A key characteristic of this distribution is the absence of moments. Therefore, theoretical confidence intervals based on the mean and the variance of the distribution cannot be computed (see Evans et al. 2000). Finally, it is conceivable that in an application using real-world data, the correlation between β_{TT} and β_{TC} would be negative, which could in turn lead to more extreme measures of the value-of-time (e.g. high value-of-time for high-earners).

It should be noted that the distributional assumptions used in the three sets of examples described in this section have generally led to very high (absolute) extreme values of time. This is a direct result of using unbounded distributions, such that, with a sample of 100,000 random draws, there is a significant probability of obtaining some very extreme outlying values. While the impact of such values can be reduced by removing the upper or lower few percentiles of the distribution (c.f. Hensher & Greene 2001), this can arguably be seen as *tinkering* with the results produced by the estimation. A preferable approach seems to be to use distributions bounded on both sides, such as for example the S_B distribution, thus avoiding problems with

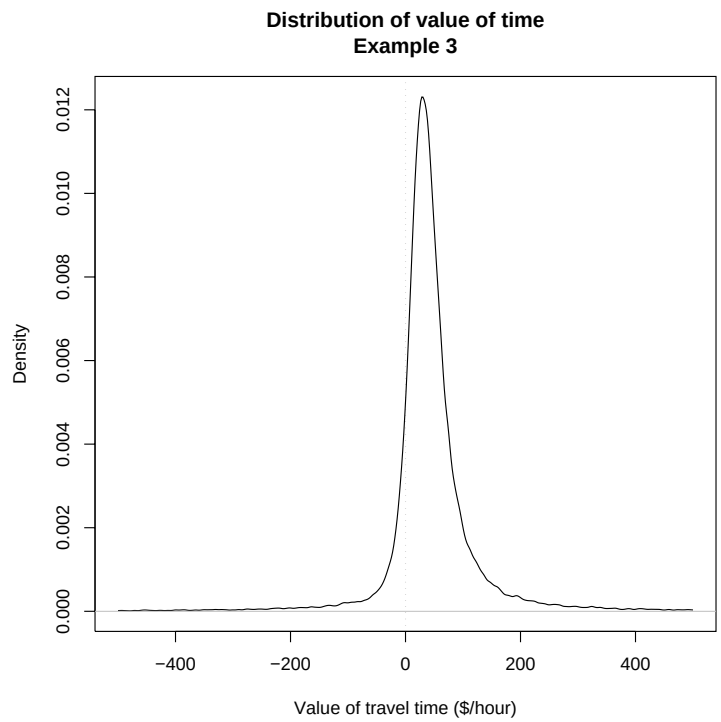

Figure 2: Distribution of time and cost coefficients and implied value of time measures

outliers altogether.

Finally, it is important to keep in mind that there is a significant source of uncertainty in the coefficients due to sampling. As is well known, the estimators of coefficients in a MNL are *normally* distributed random variables (due to sampling errors), and the distribution of the estimator for the value of time is already an issue in non-mixed models (c.f. Armstrong et al. 2001). In a RCL context, the estimators of the parameters of the *Normal* distribution are themselves *normally* distributed, and a theoretical analysis of the value-of-time distribution is a very complicated issue. Therefore, we strongly suggest a simulation-based analysis of the distribution of the value-of-time, as illustrated by the examples above. A more advanced, and more complicated, approach to calculating the distribution of the value-of-time across respondents is to use posterior analysis to calculate the most likely

values of β_{TT} and β_{TC} for each individual, given the observed choices and the fitted distributions for β_{TT} and β_{TC} across respondents. These values can then be used to calculate a value-of-time measure for each respondent, and the calculation of the distribution of these values can be performed using kernel analysis tools. However, Bayesian (posterior) analysis can be computationally quite expensive (c.f. Train 2003), and specific code may have to be produced for a given model. The calculation becomes potentially even more complicated in the case of correlated measures of β_{TT} and β_{TC}.

6 Application

It is sometimes tempting to justify the use of the *Normal* distribution for travel-time coefficients, and the 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 values of time. 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. Unfortunately, it is a common misconception that the assumptions underlying an approach are validated because the approach seems to fit the data. A related example is that of structural parameters in Nested Logit models; although values above 1 for such coefficients are not generally consistent with utility maximisation, they can lead to a better model fit and thus give a better mathematical representation of the data.

There are two potential reasons why a better model fit can be obtained when using a *Normal* distribution. One is that the shape of the *Normal* distribution in the negative space of numbers might be better able to approximate the real shape of the distribution of coefficient values than is the case for any of the bounded distributions that have been tried in the estimation. The other, and potentially more likely reason, is the existence of a positive factor that is strongly correlated with travel-time. If the gains in utility obtained from this associated activity exceed the disutility of travel-time for some agents, then the sum will be positive for these agents. If the model is in turn misspecified, such that the other factor is not explained individually in the model, then the *Normal* distribution will give a better fit,

as it explicitly allows for such positive values.

The existence and impact of unobserved factors that are correlated with travel-time are highly dependent on the application at hand. As such, their impact is not easy to illustrate in an example. For this reason, the scope of the application presented in this section is limited to showing the effect of poor distributional assumptions. 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 is 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), for previous applications using this data, see Bhat (1997) and Wen & Koppelman (2001). The sample used in the present analysis contains 4,306 observations, and looks at the choice between air, car and rail, where except for car, the different modes are not necessarily all available to every respondent.

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. In addition to two alternative specific constants (ASC) for air and train, three coefficients were used in the generation of the choice vector; the marginal utility of cost ($), the marginal utility of the frequency of service (only for air and train), and the marginal utility of total travel-time (minutes). Fixed values were used for all coefficients except travel-time, and these values are shown in table 1. The travel-time coefficient was assumed to be distributed randomly across the population following a *Triangular* distribution. However, rather 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 RCL model as a model with varying taste coefficients across individuals. The other advantage of this approach is that it allows us to calculate distributional parameters based on the actual distribution of taste coefficients across respondents, rather than the theoretical distribution, which takes into account only the distributional assumptions and ignores the impact of the specific random draws used in the generation of the actual travel-time coefficients. The parameters used for the

Coefficient	Value
ASC air	-1.5
ASC rail	0.4
Travel cost	-0.035
Frequency	0.08
Travel-time (μ)	-0.03751
Travel-time (σ)	0.02293

Table 1: Coefficient values used in generation of data

Triangular distribution were set to be $a = -0.1$ (lower bound), $b = -0.001$ (upper bound) and $m = -0.01$ (mode), leading to a theoretical mean of -0.037, with an associated standard deviation of 0.02235. The actual mean and standard deviation in the sample of $4,306$ draws used in this application are very similar to the theoretical values, at -0.03751 and 0.02293 respectively. This shows that the extra sampling variation introduced by selecting 4,306 values rather than simulating over the actual distribution has not biased the distributional parameters in any significant way.

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 choice probabilities for the different alternatives contained in that individual's choice-set. A simple Monte-Carlo exercise was then conducted to determine the chosen alternative from these choice probabilities.

Five different models where estimated on this simulated choice data; one MNL model and four 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 four 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 two models using an S_B distribution. Two different versions of the S_B distribution were used, one with:

$$c = \frac{e^{\xi}}{e^{\xi} + 1}, \tag{9}$$

and one with:

$$c = b \cdot \left(\frac{e^{\xi}}{e^{\xi} + 1} \right), \tag{10}$$

where, in both cases, $\xi \sim N(\mu, \sigma)$. The distribution shown in equation 10 uses the transformation described towards the end of section 2.1, with $a = 0$. No further gains could be made by additionally estimating the offset parameter a, its value being indistinguishably close to zero.

The results of the estimation are shown in table 2. The results show that each of the four MMNL models leads to a very significant improvement in log-likelihood over the MNL model, by -80.8, -90.57, -84.01 and -88.53 units respectively. This shows the importance of acknowledging the presence of significant levels of heterogeneity in the marginal utility of travel-time. For the four MMNL models, table 2 gives the estimated parameters of the distribution of β_{TT}, along with the implied mean and standard deviation of the coefficient. For the *Normal* distribution, there is a one-to-one correspondence between the estimate μ and the mean of β_{TT}, and the estimate σ and the standard deviation of β_{TT}. For the *Lognormal* distribution, the mean and standard deviation were calculated from μ and σ using the formulae given in section 2.4, while for the S_B distribution, the distributional statistics were produced with the help of simulation using $1,000,000$ draws based on the estimated parameters μ and σ. For the model using a scaled S_B distribution (multiplication by b), this scaling was also taken into account in the presentation of the results. Finally, for the models using the *Lognormal* and S_B distribution, a sign change was used in the presentation of the results, to reflect the negative impact of the associated attribute.

The first observation that can be made from table 2 with regards to the MMNL models is that the three different distributions lead to quite similar model fit, when compared to the much poorer performance of the MNL model. The best fit is obtained by the model using a *Lognormal* distribution, ahead of the two models using an S_B distribution, where the model using a scaling coefficient has a slight advantage. Finally, the lowest log-likelihood of the four MMNL specifications is obtained by the model using a *Normal* 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 (arising from table 1) 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 \$1.37; this is a result of the overestimated cost coefficient in this model. The four MMNL models (in the order used in table 2) give values for this ratio of \$2.55, \$2.70, \$2.74 and \$2.85 respectively. This shows that all four

	MNL		MMNL $\beta_{TT} \sim N(\mu, \sigma)$		MMNL $\beta_{TT} \sim LN(\mu, \sigma)$		MMNL $\beta_{TT} \sim S_B(\mu, \sigma)$		MMNL $\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$	
ASC air	1.3135	(4.42)	-1.9379	(-3.23)	-2.3828	(-3.96)	-2.3804	(-3.84)	-2.3209	(-4.09)
ASC rail	-0.3488	(-4.44)	0.22	(1.66)	0.3181	(2.40)	0.3123	(2.38)	0.2657	(2.07)
Travel cost	-0.0528	(-16.29)	-0.0339	(-6.68)	-0.0374	(-8.60)	-0.0368	(-8.38)	-0.0348	(-7.67)
Frequency	0.0723	(16.57)	0.0866	(10.95)	0.101	(11.18)	0.1009	(10.85)	0.0993	(11.00)
b	-		-		-		-		0.1022	(4.26)
Travel-time μ	-0.0152	(-20.56)	-0.0355	(-11.68)	-3.3534	(-37.95)	-3.3069	(-34.09)	-0.5166	(-1.61)
Travel-time σ	-		0.0201	(8.93)	0.8263	(16.76)	0.8685	(14.67)	1.3662	(6.35)
Travel-time mean	-0.0152		-0.0355		-0.0493		-0.0484		-0.0414	
Travel-time std.dev.	-		0.0201		0.0491		0.0437		0.0255	
Parameters	5		6		6		6		7	
Log-likelihood	-2582.84		-2502.04		-2492.27		-2498.83		-2494.31	
ρ^2	0.4185		0.4367		0.4389		0.4374		0.4384	

Table 2: Estimates for MNL and MMNL models (t-tests in brackets)

models slightly overestimate the true ratio, where the bias is at its lowest in the MMNL model using a *normally* distributed coefficient.

While the differences across MMNL models are relatively minor in the case of the fixed coefficients used for travel cost and frequency, more important differences exist across models in the estimates for the mean and standard deviation of β_{TT}. Given the differences across models in the scale of the coefficients (which can bias comparisons), this comparison was performed with help of the ratio between the travel-time coefficient and the cost coefficient, giving the implied willingness to pay for travel-time reductions. The mean and standard deviation of this ratio were calculated for the true values as well as for the four MMNL models, while the simple mean was used for the MNL model. The results of this calculation are presented in table 3, using multiplication by 60 to give hourly values. The results show that the MNL model considerably underestimates the mean value of time, which is a result of the overestimated cost coefficient along with the underestimated travel-time coefficient. The results further show that the model using a *Normal* distribution performs remarkably well, underestimating the true mean and standard deviation by a mere 2.3% and 9.5% respectively. The use of the bounded distributions leads to an overestimation of the mean and standard deviation, which is especially severe in the case of the standard deviation for the model using a *lognormally* distributed coefficient. This overestimation is a result of the long tails of the distributions. While the S_B distribution also leads to an overestimation of the mean and standard deviation, this bias is corrected downwards in the model using the additional scaling parameter b, with an overestimation of the mean and standard deviation of 11% and 11.9% respectively. At this point it should be noted (c.f. table 2) that in the model using the additional scaling parameter b, the parameter μ of the S_B-distributed travel-time coefficient β_{TT} is significant only at the 89% level.

The results presented in tables 2 and 3 thus show that while the model using the *Lognormal* distribution leads to the best model fit (out of the four MMNL models), it leads to the poorest performance in terms of recovering the true mean and standard deviations of the value of travel-time distribution. On the other side, the *Normal* distribution leads to the poorest model fit, yet performs best in terms of recovering the mean and standard deviation of the value of travel-time distribution. The performances of the models using the S_B distribution lie in between the these two extreme cases, in terms of log-likelihood as well as in terms of the recovery of the true parameter values. This suggest that model fit on its own may not always be an appropriate

	$\mu\,(VOT)$	$\sigma\,(VOT)$
True distribution	64.30	39.31
MNL model	17.27	-
$\beta_{TT} \sim N(\mu, \sigma)$	62.83	35.58
$\beta_{TT} \sim LN(\mu, \sigma)$	79.09	78.77
$\beta_{TT} \sim S_B(\mu, \sigma)$	78.91	71.25
$\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$	71.38	43.97

Table 3: Distribution of value of travel-time savings ($/hour)

indicator of model performance.

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 value-of-time, 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 2, lead to a probability of 3.87% of a positive travel-time coefficient despite the fact that strictly negative coefficient values were used in the generation of the data. Although the probability of a positive coefficient is very low in this case, this result nevertheless confirms the notion described in section 2.1 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. This probability can be expected to be higher in the case where the true distribution used in the generation of the draws has a larger standard deviation, while keeping the mean close to zero.

To illustrate the differences in the bounds of the different distributions, 95% quantile bounds for the value-of-time savings 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 4. The results of this analysis show the effect of allowing for positive values of β_{TT}, with a lower 95% quantile limit on the value-of-time of -$6.97 per hour when using the *Normal* distribution. On the other hand, the results show that the *Lognormal* distribution and the unscaled S_B distribution massively overestimate the upper 95% quantile limit. Overall, the best performance is obtained with the *scaled* S_B distribution, which overestimates the upper 95% quantile by a

	Lower 95% quantile limit	Upper 95% quantile limit
True distribution	9.44	149.25
$\beta_{TT} \sim N(\mu, \sigma)$	-6.97	132.49
$\beta_{TT} \sim LN(\mu, \sigma)$	11.05	283.49
$\beta_{TT} \sim S_B(\mu, \sigma)$	10.89	272.72
$\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$	6.90	157.91

Table 4: 95% quantile intervals for distribution of value of travel-time savings ($/hour)

mere 5.8%, while the *Normal* distribution underestimates it by 11.2%.

Overall, the results indicate that the *Lognormal* and the S_B distributions offer good performance in terms of recovering the true lower limit on the value-of-time savings, and the scaled S_B distribution also performs very well in recovering the upper 95% quantile. Both the *Lognormal* and the S_B distribution overestimate the mean and standard deviation of the true value-of-time, where this bias is however at an acceptable level in the case of the *scaled* S_B distribution. While the *Normal* distribution performs well in terms of recovering the true mean and standard deviation, as well as the the upper 95% quantile, it falsely indicates a possibility of negative values of time.

An even stronger indicator than the 95% quantiles can be given by looking at the minimum and maximum values of time measures used in the data generation, and the corresponding minima and maxima implied by the estimated model parameters using the different distributions. For this, simulation processes using 1,000,000 draws were used with the distributional assumptions from the four models, and the results are shown in table 5. The results show that the use of the *Normal* distribution leads to a very unrealistic lower bound of a negative value-of-time of -$89.96 per hour, while it overestimates the maximum bound by around 29%. Both the *Lognormal* and the unscaled S_B distribution lead to very significant overestimation of the upper bound, while correctly identifying the lower bound. Finally, the *scaled* S_B distribution slightly underestimates the lower bound, but offers the best overall performance with regards to the upper bound.

As a final illustration of the effect of the different distributional assumptions on the implied distribution of the value-of-travel-time savings, the re-

	Minimum VOT	Maximum VOT
True distribution	2.05	168.56
$\beta_{TT} \sim N(\mu, \sigma)$	-89.96	217.73
$\beta_{TT} \sim LN(\mu, \sigma)$	2.09	1726.97
$\beta_{TT} \sim S_B(\mu, \sigma)$	2.18	953.81
$\frac{\beta_{TT}}{b} \sim S_B(\mu, \sigma)$	0.38	175.40

Table 5: Lower and upper bounds for distribution of value of travel-time savings ($/hour)

sulting distributions are plotted in figure 3. In addition to the four estimated distributions, the plot contains the density of the actual 4,306 values used in the data generation process. The plot clearly shows the problems with the lower bound with the use of the *Normal* distribution, and with the upper bound with the use of the *Lognormal* distribution and the *unscaled* S_B distribution. While the *Lognormal* and S_B distributions perform very well in terms of recovering the true mode of the distribution, this is shifted to the right when using the *Normal* distribution, which is an effect of the symmetrical nature of this distribution and the fact that it correctly retrieves the mean to be in the area of $65 per hour. It is also of interest to note that the behaviour of the *Lognormal* and the *unscaled* S_B distributions is very similar. Finally, unlike the use of the *Lognormal* and the *unscaled* S_B distributions, the use of the *scaled* S_B distribution does not lead to problems with excessive weight in the tail of the distribution.

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 value-of-time measures, even though this probability was at a relatively low level in the present example. Also, while the *Normal* distribution does, at least in the present application, lead to a very good approximation of the mean and standard distribution of the true distribution of the value-of-time measure, the equivalence between mean and mode in the *Normal* can lead to problems. The *Lognormal* and S_B distributions avoid problems with negative values-of-time, but have the disadvantage of a heavy tail in the distribution; in the case of the S_B distribution, this problem can however be alleviated by identifying an additional scaling parameter. Overall, these results suggest that in some applications, the *Normal* distribution can be used to produce an estimate of the mean and standard de-

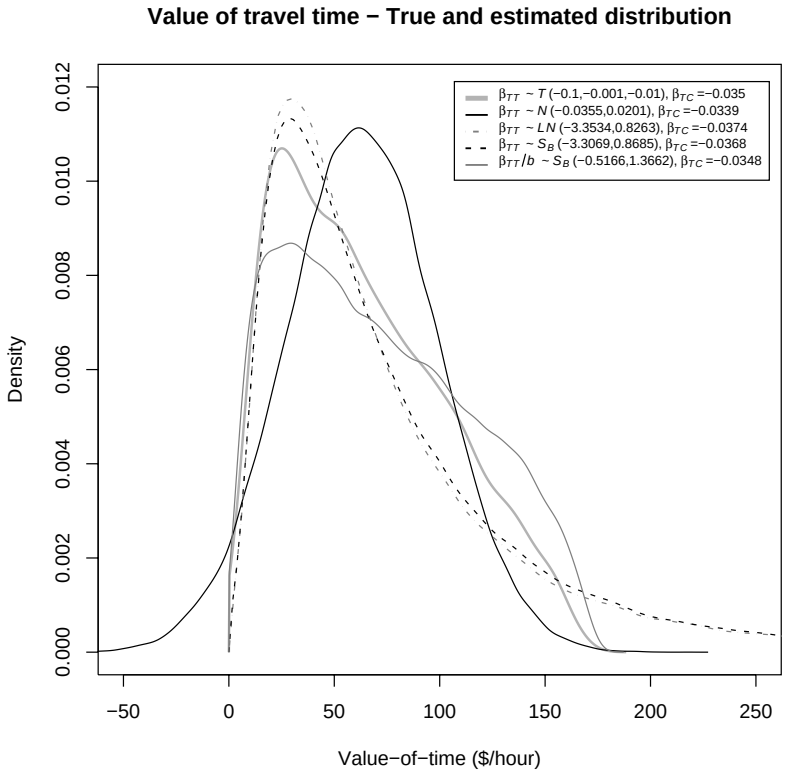

Figure 3: Distribution of value-of-travel-time under different modelling assumptions

viation of the value-of-time savings across the population, but that 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. Finally, this application has confirmed the notion described at the beginning of this section that model fit on its own may not always be an appropriate indicator of model performance; indeed, the model with the highest log-likelihood (using a *Lognormal* distribution) leads to very poor performance in terms of the upper bounds as well as the mean and standard deviation of the distribution of the value-of-time, while the best performance in terms of the mean and standard deviation is offered by the model with the poorest fit (*Normal* distribution), and the best performance in terms of bounds is offered by the

model with the second-best model fit (*scaled* S_B distribution).

7 Conclusions

In this paper, we have discussed issues arising with the identification of randomly distributed travel-time coefficients in discrete choice models. Our discussion has shown that, with the commonly used *Normal* distribution, researchers in effect make an a priori assumption that there exist some travellers with positive travel-time coefficients, leading to negative measures of value-of-time. Results indicating the presence of such positive values for this coefficient need thus not necessarily reveal the existence of such values in the population, but may equally well be seen as a direct effect of the use of the *Normal* distribution.

Aside from this purely technical reason, we have shown that there are various other possible causes that can lead to the erroneous conclusion that negative values of time exist in a population. These sources fall into two main groups; unobserved travel-experience attributes that are strongly correlated with travel-time, and conjoint activities that are pursued in the same time-interval as the travelling itself. In the case where the marginal utility of these activities is not modelled explicitly, this marginal utility will be incorporated into the marginal utility of travel-time, and depending on its sign, can bias this marginal utility of travel-time upwards or downwards. In the case where the total marginal utility of these travel-experience attributes and conjoint activities is positive and exceeds the marginal disutility of travel-time per se, the sum of these components will be positive, falsely indicating a positive marginal utility of travel-time.

In summary, there thus seem to be two potential causes that can lead to false conclusions with regards to the existence of negative values-of-time; model-misspecification in terms of poor distributional assumptions, and model-misspecification in terms of the presence of unmeasured attributes or conjoint activities. In both cases, it is possible that a model allowing for positive travel-time coefficient leads to better model fit; this is however simply a reflection of the fact that the model is better able to mathematically reproduce the actual choice behaviour, and the estimated travel-time coefficient should be seen as a biased estimate of the marginal utility of travel-time per se. Researchers whose results indicate a non-zero probability of a positive travel-time coefficient should thus always consider these potential model-

misspecifications before claiming that these results indicate the presence of individuals with negative values-of-time.

We have also discussed the calculation of value-of-time measures in the case where time and/or cost coefficients are allowed to vary randomly across the population, and have highlighted the importance of incorporating the full distribution of the value-of-time measure, rather than just calculating the mean across the population. We have described the complications that can arise in the case where the chosen random distributions allow for positive as well as negative coefficients, and in the case where the distribution used for the cost coefficient allows for values that are close to zero.

Finally, we have described how, from an economics point of view, time should be seen as a limited resource, and no matter how large the gain in utility resulting from an activity is, a rational economic agent would always prefer to obtain this gain without affecting the availability of time for other activities. This suggests that results indicating a positive marginal utility of travel-time, net of the effects of conjoint activities or unmeasured attributes, are not consistent with the rationality assumptions underlying economic theory.

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.

Armstrong, P., Garrido, R. & Ortuzar, J. D. (2001), 'Confidence interval to bound the value of time', *Transportation Research E* **37**(1), 143–161.

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. (1997), 'Covariance Heterogeneity in Nested Logit models: economic structure and application to inter-city travel', *Transportation Research* **31**(1), 11–21.

Bhat, C. R. (1998*a*), 'Accommodating flexible substitution patterns in multidimensional choice modeling: formulation and application to travel mode and departure time choice', *Transportation Research* **32B**(7), 425–440.

Bhat, C. R. (1998*b*), '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.

Bhat, C. R. & Castelar, S. (2002), 'A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area', *Transportation Research* **36B**, 593–616.

Boyd, J. & Mellman, J. (1980), 'The effect of fuel economy standards on the u.s. automotive market: a hedonic demand analysis', *Transportation Research* **14A**, 367–378.

Brownstone, D. & Train, K. (1999), 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics* **89**, 109–129.

Cardell, S. & Dunbar, F. (1980), 'Measuring the societal impacts of automobile downsizing', *Transportation Research* **14A**, 423–434.

Cherchi, E. & Ortuzar, 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.

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.

Evans, M., Hastings, N. & Peacock, B. (2000), *Statistical distributions*, third edition edn, John Wiley & Sons.

Hensher, D. (2001*a*), 'Measurement of the valuation of travel time savings', *Journal of Transport Economics and Policy* **35**(1), 71–98.

Hensher, D. (2001*b*), 'The sensitivity of the valuation of travel time savings to the specification of unobserved effects', *Transportation Research* **37E**, 129–142.

Hensher, D. (2001*c*), 'The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications', *Transportation* **28**, 101–118.

Hensher, D. & Greene, W. H. (2001), *The Mixed Logit Model: The State of Practice and Warnings for the Unwary*, Working Paper, Institute of Transport Studies, Faculty of Economics and Business, The University of Sydney, Sidney, Australia.

Hess, S. & Polak, J. W. (2004*a*), *Mixed Logit estimation of parking type choice*, paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. & Polak, J. W. (2004*b*), *Mixed Logit modelling of airport choice in multi-airport regions*, paper submitted to Transportation Research A.

Hess, S., Train, K. & Polak, J. W. (2003), *On the use of randomly shifted and shuffled uniform vectors in the estimation of a Mixed Logit model for vehicle choice*, submitted to Transportation Research B.

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.

Koppelman, F. S. & Sethi, V. (2000), Closed form discrete choice models, *in* D. Hensher & K. J. Button, eds, 'Handbook of Transport Modeling', Pergamon Press, Oxford.

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.

Lindelof, B. (2000), 'Many commuters savor private luxury of drive time', *The Sacramento Bee* p. A1 and A12.

Luce, R. (1959), *Individual choice behavior: a theoretical analysis*, J.Wiley and Sons, New York.

Marschak, J. (1960), Binary choice constraints on random utility indications, *in* K. Arrow, ed., 'Stanford Symposium on Mathematical Methods in the Social Science', Stanford University Press, Stanford, CA, pp. 312–329.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York.

Mokhtarian, P. L. & Salomon, I. (2001), 'How derived is the demand fro travel? some conceptual and measurement considerations', *Transportation Research A* **35**, 695–719.

Perez, P. E., Martinez, F. J. & Ortuzar, J. 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 desirec 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 D* **3**(3), 129–140.

Sillano, M. & Ortuzar, J. (2003), *WTP Estimation with Mixed Logit Models: Some New Evidence*, Working Paper, Department of Transport Engineering, Pontificia Universidad Catlica de Chile, Santiago, Chile.

Sipress, A. (1999), 'Not all commuters driven crazy', *Washington Post* p. A1 and A12.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.

Train, K. & Sonnier, G. (2003), *Mixed Logit with Bounded Distributions of Partworths*, Working Paper, Department of Economics, University of California, Berkeley, CA.

Walker, J. L. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, PhD thesis, MIT.

Wen, C.-H. & Koppelman, F. S. (2001), 'The Generalized Nested Logit Model', *Transportation Research* **35B**(7), 627–641.

Young, W. & Morris, J. (1981), 'Evaluation by individuals of their travel time to work', *Transportation Research Record* **794**, 51–59.