DTU Library

Sample selection and taste correlation in discrete choice transport modelling

Mabit, Stefan Lindhard

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Sample selection and taste correlation in discrete choice transport modelling

Stefan L. Mabit

DTU Transport
Technical University of Denmark

Abstract

The subject of this thesis is discrete choice analysis in transport modelling. Many situations within transportation research may be modelled as a choice from a discrete set of alternatives. The framework of random utility maximisation is well-established to model such choices but there are still many issues that deserve attention. This thesis investigates how sample selection can affect estimation of discrete choice models and how taste correlation should be incorporated into applied mixed logit estimation.

Sampling in transport modelling is often based on an observed trip. This may cause a sample to be choice-based or governed by a self-selection mechanism. In both cases, there is a possibility that sampling affects the estimation of a population model. It was established in the seventies how choice-based sampling affects the estimation of multinomial logit models. The thesis examines the question for a broader class of models. It is shown that the original result may be somewhat generalised. Another question investigated is whether mode choice operates as a self-selection mechanism in the estimation of the value of travel time. The results show that self-selection can at least partly explain counterintuitive results in value of travel time estimation. However, the results also point at the difficulty of finding suitable instruments for the selection mechanism.

Taste heterogeneity is another important aspect of discrete choice modelling. Mixed logit models are designed to capture observed as well as unobserved heterogeneity in tastes. But just as there are many reasons to expect unobserved heterogeneity, there is no reason to expect these tastes for different things to be independent. This is rarely accounted for in transportation research. Here three separate investigations of taste correlation in willingness-to-pay estimation are presented. The first contribution addresses how to incorporate taste correlation in the estimation of the value of travel time for public transport. Given a limited dataset the approach taken is to use theory on the value of travel time as guidance in the specification of the correlation. The second contribution examines how different distributional assumptions are affected by the inclusion of taste correlation. The third contribution investigates the correlation patterns between willingness-to-pay measures for different public transport modes and how to capture them in the simplest possible way. A general feature of the three investigations is that we find scale heterogeneity. Since this induces correlation it is an important aspect of taste correlation to specify the scale correctly. We see that scale heterogeneity may be partly explained by background variables. Looking at the three contributions on taste correlation there seems to be the general conclusion that significant taste correlation is often present and that it sometimes has an effect on willingness-to-pay evaluation. A conclusion for applied work is that it should allow for correlation if this has not been sufficiently captured by the remaining specification of the model.

Resumé

Denne afhandling behandler analysen af diskrete valg i trafikmodellering. Mange situationer inden for trafikforskning kan modelleres som et valg mellem en endelig mængde af muligheder. Teorien om stokastisk nyttemaksimering er veletableret som model for sådanne valg, men der er stadig mange problemstillinger, som fortjener opmærksomhed. Denne afhandling undersøger, hvordan dataudvælgelse kan påvirke estimationen af diskrete valgmodeller, samt hvordan korrelation mellem præferencer er en del af anvendt mixed logit modellering.

Dataudvælgelsen i trafikmodellering er ofte baseret på en observeret rejse. Det kan betyde, at en stikprøve bliver valgbaseret eller påvirket af en selvudvælgelsesmekanisme. I begge tilfælde er der en risiko for, at dataudvælgelsen påvirker estimationen af en model, der beskriver befolkningen. Tilbage i halvfjerdserne blev det etableret, hvordan valgbaseret datudvælgelse påvirker estimationen af multinomial logit modeller. Afhandlingen undersøger valgbaseret dataudvælgelse for en større gruppe af modeller. Det ses, at det oprindelige resultat delvist kan generaliseres. Et andet spørgsmål, som behandles, er, hvorvidt transportmiddelvalg fungerer som en selvudvælgelsesmekanisme i forbindelse med estimationen af værdien af rejsetid. Resultaterne viser, at selvudvælgelse delvist kan forklare kontraintuitive resultater omkring estimationen af værdien af rejsetid. Men de bekræfter samtidig, hvor svært det er at finde brugbare instrumenter til modellen.

Heterogenitet af præferencer er et andet vigtigt aspekt af diskret valgmodellering. Mixed logit modeller er designede til at fange både observeret og uobserveret heterogenitet i præferencer. Men ligesom der er mange grunde til at forvente uobserveret heterogenitet, er der ingen grund til at forvente, at et individs smage for forskellige goder nødvendigvis er uafhængige. Dette forhold tages der sjældent hensyn til i modellering af trafik. Afhandlingen præsenterer tre forskellige undersøgelser af korrelation mellem præferencer i forbindelse med estimation af betalingsviljer. Det første bidrag omhandler, hvordan præferencekorrelation kan indarbejdes i estimationen af værdien af rejsetid for kollektiv trafik. På grund af databegrænsninger benyttes teori om værdien af rejsetid som vejledning til specifikationen af korrelation. Det andet bidrag undersøger forholdet mellem funktionsantagelser og korrelation i estimationen af betalingsviljer. Det tredje bidrag undersøger præferencekorrelationer mellem betalingsviljer for forskellige kollektive transportmidler, samt hvordan disse kan modelleres på den simplest mulige måde. Et gennemgående resultat af de tre undersøgelser er, at der er skalaheterogenitet. Da dette skaber korrelation, bliver det en vigtig del af modelleringen af korrelation at modellere skalaen. Resultaterne viser, at skalaheterogeniteten delvist kan forklares af baggrundsvariable. Samlet set viser de tre bidrag, at der ofte er signifikant smagskorrelation, og at det i nogle tilfælde påvirker evalueringen af betalingsviljerne. En konklusion for anvendte modeller er derfor, at man bør tillade korrelation, hvis den ikke er beskrevet tilstrækkeligt af den øvrige modelspecifikation.

Acknowledgements

I would like to thank my supervisors Senior Researcher Mogens Fosgerau, Danish Transport Research Institute, and Professor Otto A. Nielsen, Centre for Traffic and Transport, both for the role they played in making my PhD study a reality and their participation in the study during the last four years. Mogens is especially thanked for the many hours of discussion about transportation research in general and econometrics in particular. Otto is thanked for keeping me connected with applied modelling and for an always optimistic view on problems.

I would like to thank my examiners, Andrew Daly, Jeppe Rich, and Michel Bierlaire, for the valuable feedback they have given me. This has greatly improved the thesis.

I would also like to thank the two institutions that have supported my study. The Technical University of Denmark is thanked for the funding and the Centre for Traffic and Transport is thanked for all the practical support during my study. I am also thankful to all the nice people at the centre who have made it fun to work here for the past four years.

A special thank to Andersen's Travel Grant to Mathematicians that made it possible for me to spend a year of study at leading universities within the transport modelling community. Also thank you to the researchers who made the travel an inspiration for my work: David Hensher and John Rose, University of Sydney, Marcela Munizaga and Sergio Jara-Diaz, Universidad de Chile, and Michel Bierlaire, Ecole Polytechnique Federale de Lausanne.

Thank you to Trine and my family who support me in whatever I undertake. And, finally, thank you to my friends, both those who have shown real interest in my thesis and those who have pretended out of politeness.

January 4th, 2008

Stefan L. Mabit

Contents

Chapter 1

Introduction

1.1 Discrete choice and transport modelling

This thesis concerns discrete choice analysis in transport modelling. The thesis contributes to a tradition that over the last 30 years have combined the two areas. Many questions within transportation research may be modelled as a choice from a discrete set of alternatives. The discrete choice models were introduced in transport modelling to improve the four-step framework. In a four-step model, transport demand modelling is characterised by four questions: How many trips are taken, what destinations are chosen, what modes are used, and what routes are taken? From a modelling perspective, the answers to each of these questions may be described as a choice by an individual among a finite number of alternatives. Reflecting this, discrete choice modelling has become a main tool in transport modelling. Applications are numerous and especially mode choice, route choice, and willingness-to-pay estimation are enriched by applications of discrete choice models. The interaction between the two areas began with applications of multinomial logit models (MNL) in the seventies, e.g., Domencich & McFadden (1975), and developed into the use of advanced models, such as mixed logit models, e.g., McFadden & Train (2000) and nonparametric methods, e.g., Horowitz (1993).

Choices in transportation have been modelled at three different levels (McFadden 2007). First, they were only modelled indirectly as part of an aggregated outcome. An example is a gravity model used for explaining trip volumes. Second, a choice was modelled as the outcome of rational individuals maximising over preferences based on economic consumer theory. This is mainly represented by the class of random utility maximisation (RUM) models. Finally, choices have been investigated based on findings in psychology, brain science, etc., which contest the classic economic perspective of choices as the outcome of stable preferences. An example is the theory developed by Tversky & Kahneman (1991).

This thesis is concerned with modelling at the second level. It investigates important aspects of discrete choice analysis in behavioural transport modelling. The motivation comes from;

- The need for society to model the daily choices made by travellers.

- The great amount of behavioural aspects RUM models are able to imitate.
- The need to develop and consolidate state-of-the-art discrete choice modelling to advance applied transport modelling.

This thesis focuses on two aspects of discrete choice modelling. The first is sample selection. Sample selection describes the way individuals have entered the sample. Sampling in transport modelling is often based on an observed trip. This may cause the sample to be governed by some sampling mechanism. The question is then whether the way the sample is obtained affects the estimation. A special concern in discrete choice modelling is the issue of choice-based sampling. This happens when sample selection is based on the choices that are the objects of study. This is treated more in Chapter 2. Another mechanism that may affect estimation in transport modelling is self-selection. This happens when something unobserved affects both the choices and the sample selection. This is discussed more in Chapter 3.

The second issue concerns taste heterogeneity and correlation between these tastes. RUM models depict a choice as the outcome of utilities based on economic consumer theory. The utilities are based on attributes that characterise each alternative. The derivative of the utility with respect to an attribute represents the taste for that attribute. In transportation, these tastes are sometimes the main object of interest, e.g., the value of travel time (VTT) which is a ratio between two tastes (the tastes for time and cost). This has resulted in many years of research into the estimation of mean tastes in a population. Advances, e.g., mixed logit models, have made it possible to estimate the distribution of tastes. Most research has done this by focusing on marginal taste distributions. This thesis emphasises the importance of considering the correlation between the different tastes as well.

One reason why heterogeneity has become more important in modelling is that the transport system has become more and more complex. Therefore heterogeneity is an integral part of many policy issues. An example is road pricing where it is essential to include heterogeneity (e.g., Small et al. 2005). The reference is just one of many examples where mixed logit models have shown their value when heterogeneity is important. An opponent to mixed logit would say that this heterogeneity is better modelled by background variables. This is correct but these might not be available in the dataset, and in many cases they only explain part of the heterogeneity. Furthermore, some models may not be able to include them while they can include random heterogeneity.

The thesis is organised as follows: The remainder of this chapter discusses discrete choice and mixed logit models, sample selection, and taste correlation, before concluding with a summary of the contributions of the five papers. The discussion is not intended as a review of the modelling issues but as an overview necessary to place the contributions in relation with state-of-the-art modelling. The remaining chapters present the papers. Chapter 2 is a note on choice-based sampling and nested generalised extreme value (GEV) models, while Chapter 3 is a paper on how sample selection may affect the estimation of VTT. The last three chapters look at taste correlation in mixed logit models. Chapter 4 contains a paper on how theoretically motivated taste correlation structures can improve a mixed logit model. Chapter 5 investigates how different distributional assumptions are affected by taste correlation in both route and mode choice models. Chapter 6 contains a paper that investigates how to specify and test models including taste correlation in the estimation of public transport willingness-to-pay (WTP) indicators.

1.2 Discrete choice models

Discrete choice models within the RUM framework are based on utility theory from economic consumer theory and random utility theory developed originally by psychologists (McFadden 2000). A discrete choice is any choice from a finite set of alternatives.[1] This choice set is denoted C.[2] A general discrete choice model for a generic individual consists of a set of probabilities attached to the alternatives

$$P(j|x,\beta) \quad \forall j \in C,$$

where x is a vector of explanatory variables and β is a vector of coefficients. This thesis only considers models based on RUM. This can be assured by the assumption that each alternative has attached a stochastic utility function

$$U_j = U_j(x_j, \beta, \varepsilon)$$

where ε follows some distribution, and assuming that the alternative maximising utility is chosen

$$P(i|x,\beta) = P(U_i > U_j, \quad \forall j \neq i). \tag{1.1}$$

This use of strict inequalities follows Train (2003).[3] Discussions of the relationship between discrete choice models and RUM models are given in McFadden (1981) and Ben-Akiva & Lerman (1985), where the general discrete choice models are denoted as probabilistic choice systems. There seem to be different opinions in the literature as to how close the utility functions should be connected with economic theory. At one end there are modellers using conditional indirect utility functions derived from an economic model, e.g., Train & McFadden (1978). At the other end there is the view that the utility function only represents that the attributes may be reduced to a one-dimensional measure, i.e., the function represents an assumption of commensurability. This view is presented in Ben-Akiva & Lerman (1985) where they only present the definition given in Equation 1.1. An in-between solution is that any applied RUM model should be in accordance with economic theory. To be useful the model also needs to allow for Equation 1.1 to be computationally feasible.

As a point of departure it is very important not to place restrictions on the utility function,

[1] While the word finite is sufficient in a mathematical definition it may be of importance in applications whether this is a small number (mode choice) or a large number (route choice).

[2] The choice set may be indexed by individuals if necessary.

[3] As discussed by McFadden & Train (2000) there is no difference between using an inequality and a strict inequality in a well-specified RUM model.

other than those imposed by fundamental theory. The general form as given by Ben-Akiva & Lerman (1985) is

$$(1.2) \qquad U_j = V_j(z_j, s) + \varepsilon(z_j, s),$$

where z_j are attributes of a given alternative, s are background variables, e.g., socioeconomic variables, and ε is a random variable depending on z_j and s. As noted by McFadden & Train (2000), U_j should be decreasing in the cost of alternative j, c_j, for the model to be in accordance with economic theory.

The workhorse of discrete choice models is the MNL model. The most common specification for this model is the linear-in-parameters specification

$$(1.3) \qquad U_j = \beta' x_j + \varepsilon_j,$$

where ε_j is independently and identically Gumbel distributed with unit scale. Here x may include transformations of the variables of interest. What has made this model so useful is the closed-form expression of the resulting choice probabilities. For the model the probability of choosing alternative i is

$$(1.4) \qquad P(i|x,\beta) = \frac{e^{\beta' x_i}}{\sum_j e^{\beta' x_j}}.$$

Two restrictive assumptions of the model are, however, that β is constant within the population or segments and that utilities are independent and homoscedastic, i.e., the covariance matrix is proportional to the identity matrix. Due to the closed-form probabilities, estimation is very fast. This facilitates the search for the best specification given x. This search includes which interactions and transformations of x to use.

1.2.1 The mixed logit model

The mixed logit model[4] allows the coefficients in Equation 1.3 to vary in the population following specified distributions, i.e., it is a mixture of MNL models over either continuous or discrete distributions or both. This allows for individuals to have different tastes for attributes. Another gain is that substitution patterns may be made more realistic through the use of error components (Walker 2001). The choice probabilities for the model are given by:

$$(1.5) \qquad P(i|x) = \int \frac{e^{\beta' x_i}}{\sum_j e^{\beta' x_j}} f(\beta) d\beta,$$

where $f(\beta)$ is the density function assumed for β. The mixed logit model has the same specification issues as the MNL model, but in addition an important issue is the choice of mixture distribution. The choice has been discussed and compared in many applications, e.g., Hensher

[4]Other names are logit kernel (Walker 2001), random coefficient logit (Ben-Akiva & Lerman 1985), or mixed multinomial logit (McFadden & Train 2000).

(2001), Hess et al. (2005), and Fosgerau (2006). Since the subject has been much debated, I will only comment on some aspects with relevance for this thesis. Two distributions that are commonly used are the normal and the lognormal distributions.

The normal distribution is often favoured in applications because of its simplicity; it is easy to generate normal draws. On the negative side is that the distribution has support on the whole of the axis. This is unreasonable for many marginal utilities. An example is a cost coefficient, β_C, which from a microeconomic perspective is required to be negative for all individuals. A mixed logit model with a normal cost coefficient cannot be a RUM model (McFadden 2000).[5] A consequence of unrealistic normal coefficients is seen in Revelt & Train (1998). The paper models the choice of refrigerators. They apply a normal coefficient to interest rate on a loan financing some of the alternatives. Within a certain range their model predicts that market share will rise with higher interest rate, which is unreasonable.

The lognormal is also simple to simulate as it is derived from a normal. This distribution solves the problem with support, since it is only defined on a half axis. As a problem, this distribution has been reported to give unreasonably high mean estimates (Hensher 2006).

Several double-bounded distributions, e.g., the triangular or S_B, have been proposed to remove the problems. The cost is either computational, if the researcher chooses to estimate the end-points of the distribution, or conceptual if the researcher chooses to fix the end-points. In the latter case, if the researcher obtains reasonable estimates, this could be due to the fixed end-points and not because the model is correctly specified. In this sense, a lognormal distribution has the possibility to signal a mis-specification through high mean estimates while this may not be possible with double-bounded distributions.

A natural but complicated extension of the distributions described above consists of allowing for more flexible functional forms. An example of this is the approach developed by Fosgerau & Bierlaire (2007). They present a method of sieves to test the appropriateness of the mixing distribution. In addition to being a test, this can also be seen as a way to make the distributions more flexible. Their approach is compared in Fosgerau & Hess (2006) to an approach that allows for the distributions to be discrete mixtures of continuous distributions.

Following specification of the model comes estimation. One approach to estimation is maximum simulated likelihood (MSL), i.e., maximisation of the objective function

$$SLL = \sum_n ln(\check{P}(i_n|x,\beta)), \tag{1.6}$$

[5] This view depends on the extend to which one thinks that RUM models should be in accordance with economic theory.

where i_n indicates the alternative chosen by individual n. MSL estimates the choice probabilities using simulated probabilities

$$\check{P}(i|x) = \frac{1}{R}\sum_r \frac{e^{\beta_r' x_i}}{\sum_j e^{\beta_r' x_j}}, \tag{1.7}$$

where each β_r is a draw from $f(\beta)$. To apply the simulator, a variety of different drawing approaches may be used. As pseudo-random numbers are inefficient in the simulation, quasi-random number sequences have been used, e.g., Halton sequences (Bhat 2001, 2003). If many dimensions of simulation are needed, the Halton sequences are correlated, therefore other sequences have been investigated. Numbers that appear to have nice properties in several dimensions are the modified latin hypercube sampling (MLHS) draws. These are compared to Halton draws by Hess, Train & Polak (2006).

Estimation by MSL gives consistent and asymptotically normal estimates given certain condition on the number of draws, R, and the sample size, N (Train 2003). However, it has to be decided what estimator to use for the variance-covariance matrix. Given the log-likelihood function in Equation 1.6 it is possible to calculate the covariance of the scores in the population, B, and the expected hessian, A. These sum to zero at the true parameters when the model is correctly specified. In applications, the model is never exactly true so the best we can hope for is that B and $-A$ are similar. If simulation is used as in MSL, this adds a second layer of uncertainty. McFadden & Train (2000) argue to use the estimator $A^{-1}BA^{-1}$ instead of B^{-1} as the latter has shown poor finite-sample performance.[6]

When a mixed logit model has been estimated, it is necessary to examine whether the mixing is significant. The standard approach in a non-linear model estimated by maximum likelihood is to use a likelihood ratio (LR) test. In the case of testing one mixed coefficient, the null hypothesis is that the variance is zero. This value lies on the boundary of the parameter space. Therefore, a standard LR test is invalid. This is noted by Horowitz (1993) in the case of random probit.

As noted by Hjort (2007), Self & Liang (1987) deduce the asymptotic distribution to be a fifty-fifty mix of a point mass at zero and a χ^2_1 distribution. This shows that if a variance is significantly different from zero with the standard test, it will also be so with the correct test.[7]

Inspired by marketing, the use of stated-preference (SP) data has become common in transportation research. These data offer an easy way to collect several observations per individual, i.e., to collect panel data. Mixed logit models are well-suited to these data since the distribution may be assumed to vary over individuals, only. In the case of a mixed logit model estimated on panel data it makes sense to interpret the distributions as taste distributions whereas there is high risk that the distributions cannot separate taste from random noise in the case of cross-section

[6] There does not seem to have been made any comparisons between $A^{-1}BA^{-1}$ and $-A^{-1}$ for mixed logit models.

[7] This result first came to my knowledge after most of the estimation in this thesis. Therefore it has not been incorporated. Looking through the papers, it does not invalidate the conclusions.

data.[8]

A mixed logit model adapted to panel data has the form

$$U_{jnt} = \beta_n' x_{jnt} + \varepsilon_{jnt}, \tag{1.8}$$

where t denotes the t'th choice situation by individual n. Even though the model accounts for correlation over repeated choices by the same individual, it still makes the assumption of independent homoscedastic additive errors. This places a restriction on the unidentified scale. One way to relax the model in case of panel data is to assume that the errors follow individual-specific distributions. This can be attained with

$$U_{jnt} = \beta_n' x_{jnt} + \gamma_n \varepsilon_{jnt}, \tag{1.9}$$

where γ_n may depend on background variables s_n. If γ_n has a meaningful inverse then this specification is still a mixed logit model given by

$$U'_{jnt} = \gamma_n^{-1} \beta_n' x_{jnt} + \varepsilon_{jnt}. \tag{1.10}$$

In this way, the mixed logit model is capable of capturing some additional heteroscedasticity. This resembles the approach used by Caussade et al. (2005) for MNL models. The specification in Equation 1.10 is a standard mixed logit model but the derivation shows that the β in Equation 1.8 should allow the marginal distributions to be correlated and that background variables may have an effect on all coefficients if they affect the scale. This intuition motivates the models in Chapters 4 to 6.

1.2.2 Willingness-to-pay indicators

Besides mode and route choice where the object of interest is the choice probability, a third issue has led to much discrete choice research in transport modelling. This is the estimation of willingness-to-pay (WTP) indicators. These play an important role both in cost-benefit analysis and as input into transport models, e.g., assignment models.

The most common WTP indicator is the value of travel time (VTT). In discrete choice models, the computation of VTT is given by the ratio of the partial derivatives of the utility function with respect to travel time and travel cost (i.e., the marginal rate of substitution between travel time and travel cost). Jara-Diaz (2000) and Mackie et al. (2001) provide overviews of the development in economic theory concerning VTT. In case of a model with errors, which are independent of time and cost, the VTT measure is computed as:

$$VTT = \frac{\partial V / \partial x_T}{\partial V / \partial x_C} \tag{1.11}$$

with V being the systematic part of utility, and x_T and x_C representing the travel time and travel cost attributes respectively. From utility theory we have $\partial V / \partial x_C < 0$ and from the framework of

[8] The problem of seperating noise and tastes still remains with panel data but in most cases it will be diminished.

DeSerpa (1971) it follows that $\partial V/\partial x_T \leq 0$. So within this framework, only non-negative VTT estimates are theoretically consistent.

The distributions chosen for the random coefficient in a mixed logit model carries over to the distributions for the WTP measures. If the cost coefficient is distributed, this generally leads to WTP as a ratio of two distributions. Therefore, the inverse distribution of the cost coefficient is important. The easiest way to accomplish a simple inverse to the cost coefficient is by fixing it, but this is a strong restriction in case of models with homoscedastic errors. The experience from the estimations in this thesis is that heterogeneity in the cost coefficient is very influential.

One way to avoid problems with the WTP as a ratio is by parameterisation in WTP space, see Train & Weeks (2005). This approach is also used in Chapters 3 and 5.

1.3 Sample selection

This section discusses two kinds of sample-selection mechanisms. The first is choice-based sampling, which is special to discrete choice modelling. The second mechanism is self-selection, which is more general than choice-based sampling. If the mechanism is correlated with unobservables affecting behaviour, then the sampling mechanism can disturb the estimation of a discrete choice model.

1.3.1 Choice-based sampling

Choice-based sampling is both a useful tool and a potential problem when estimating a discrete choice model. Choice-based samples are useful when a specific alternative with low market share is of special importance for a model, e.g., biking in the US or environmental labels on products. The problem with choice-based samples is discussed in the paper by Manski & Lerman (1977). In the paper, they show that choice-based samples in general lead to inconsistent estimates if estimated by exogenous sample maximum likelihood (ESML). Furthermore, they propose a consistent estimator that they name weighted ESML (WESML). The WESML estimator is based on the objective function

$$WLL = \sum_n \frac{Q_i}{H_i} ln(P(i|x,\beta)),$$

where H_i and Q_i are the choice frequencies in the sample and population. For the procedure to work it is necessary to assume that the population shares Q_i are known.[9] Finally in the paper, they present a proof by McFadden that ESML estimation of a MNL model will lead to consistent estimates except for the constants. The intuition is as follows. Suppose a population makes choices following a MNL model as given by Equation 1.4. Let s indicate the event to be

[9] Manski & Lerman (1977) note that it can alternatively be estimated from an auxiliary sample of size going to infinity.

in the sample. Then the proof is based on the result that

$$(1.12) \qquad P(i|x,\beta,s) = \frac{e^{\beta' x_i + \gamma_i}}{\sum_j e^{\beta' x_j + \gamma_j}},$$

where $\gamma_j = ln(H_i/Q_i)$. For an estimation that allows for a full set of alternative specific constants, the above expression shows that ESML estimation on the choice-based sample will only bias the constants. These can afterwards be corrected by γ_j.

Bierlaire et al. (2006) extend this result to a class of models they denote block-additive GEV models. These models have probabilities defined by GEV-generating functions

$$G(y_1,\ldots,y_J) = \sum_m G^m(y_{C_m}),$$

where C_m is a partitioning of the alternatives, y_{C_m} denotes the subvector of alternatives in C_m, and G^m is a μ-homogenous GEV-generating function. For this model, they show that the MNL result can be generalised as long as the relative sampling rates within each block corresponds to the population rates.[10] A nested logit (NL) model has generating function

$$(1.13) \qquad G = \sum_m G^m(y_{C_m}) = \sum_m \Big(\sum_{j \in C_m} y_j^{\mu_m} \Big)^{\frac{\mu}{\mu_m}}.$$

Since each G^m complies with the conditions of the block-additive GEV model, the proof may be generalised to cover NL models and more generally nested GEV models. This is the approach used in Chapter 2.

1.3.2 Self-selection into samples

Choice-based sampling illustrates that it may be a problem in discrete choice modelling if the choice affects the sampling. Another problem is if individuals self-select into the sample based on a mechanism that is correlated with the dependent variable without being controlled for. One version of this problem has been investigated in the field of labour supply. Heckman (1979) presents a model to control for self-selection. The model has been relaxed in many aspect since then (Vella 1998). However, Heckman's model is useful to describe the mechanism behind self-selection. Therefore, it is worthwhile to sketch out the original model.

In labour supply, the interest is to estimate a wage equation, i.e., the mean wage as a function of some explanatory variables. The problem is that the expected wage affects participation in the labour market and that inference on wage equations is based on individuals observed in the labour market. Thus, individuals select to be part of the labour market, and hence the sample population, based on their expected wage. The model described by Heckman consists of a wage equation and a selection model describing whether an individual participates in the labour market. The selection model is

$$S = 1\{U_1 > 0\}, \qquad U_1 = \gamma' x_1 + u_1,$$

[10] Their result is more generally also taking care of sampling of alternatives. Here I only restate the part relevant for choice-based sampling.

where $u_1 \sim N(0,1)$ and x_1 are explanatory variables including a constant. For individuals in the sample, i.e., $S=1$, the relationship between wage and other explanatory variables x_2 is captured by a linear model

$$ln(wage) = \beta' x_2 + \varepsilon,$$

where ε is mean zero and uncorrelated with x. Central to this model are the explanatory variables that enter the selection equation and not the wage equation. These variables act as instruments for the selection. Self-selection corresponds to correlation between the errors in the two equations.

A variation of the model is applied by Vella (1992) to the case where both the selection equation and the main equation are discrete. This type of model is highly dependent on the specification of the error distributions. The dependence makes it difficult to use the nonparametric approaches developed for the original model.

Chapter 3 presents a model developed to capture this kind of sample selection in the modelling of VTT. The selection model serves to separate user type and mode type effects. These are discussed by Wardman (2004). Another investigation of self-selection and VTT is presented in Fosgerau et al. (2007). Other applications involving self-selection include energy consumption and choice of appliance (Dubin & McFadden 1984, Bolduc et al. 2001) and trip length and VTT (Daly & Carrasco 2006). Bolduc & McFadden (2001) also discuss it in the context of access/egress walking and public transport.

1.4 Correlation in discrete choices

The main limitation in the MNL model is the assumption that the covariance matrix for the utilities is a homoscedastic diagonal matrix, i.e.,

$$cov(U,U) = cov(\beta' x + \varepsilon, \beta' x + \varepsilon) = kI_J,$$

where I_J is a diagonal matrix of dimension J and k is a constant. The generalisations belonging to the GEV family allow flexibility in $cov(U,U) = cov(\varepsilon, \varepsilon)$. Although $cov(\varepsilon, \varepsilon)$ can only take positive values, GEV models can approximate any model where $cov(U,U)$ is independent of x (Dagsvik 1994). The covariance matrix describes correlation between alternatives - this is known as inter-alternative correlation.

In a mixed logit model we can only make the reduction

$$cov(U,U) = x cov(\beta, \beta') x' + kI_J.$$

This shows how the distributions in β induce inter-alternative correlation. The off-diagonal elements represent inter-alternative correlation. The correlation corresponding to $cov(\beta, \beta')$ will be denoted as taste correlation. Though not evident from the notation, independent marginal distributions in β create inter-alternative correlation as long as they are shared between alternatives. The relationship shows that it is difficult to separate taste correlation from correlation between alternatives as discussed by Hess, Polak & Bierlaire (2006). In binary choices there is no relevant inter-alternative correlation. This makes them suitable for the investigation of taste correlation.

Correlated coefficients have been estimated in some applications, e.g., Revelt & Train (1998), Huber & Train (2001) or Train & Sonnier (2005), and indirectly in models applying WTP space (Train & Weeks 2005). But correlation is rarely allowed for in transport modelling. Models allowing for taste correlation are investigated further in Chapters 4 to 6.

Both the literature and this thesis restrict the investigation to distributions derived from normal distributions through a monotone transformation. Just as a multivariate normal may be described by its mean and covariance matrix the same is true for the derived distributions, i.e., $\beta = h(\mu + Lu)$, where h is a vector of monotone transformations of the marginal distributions, L is a lower triangular matrix such that $\Sigma = LL'$ and $u \sim N(0,I)$. The parameterisation of Σ through L is known as a Choleski factorisation. In case of a model with distributions not derived from a normal distribution, one would have to use a projection matrix similar to the Choleski factorisation. This is mentioned by Greene et al. (2006), but they do not apply it.

The approach above models correlation as part of a specific distribution. This depends heavily on the distributional assumptions as shown in Chapter 5. Therefore it would be nice to have a distribution-free assessment. A heuristic to assess correlation prior to the choice of distribution is presented in Sørensen (2003).

1.5 Contribution of the papers

Chapters 2 to 6 consist of five papers. Paper 1 and 2 are concerned with two different aspects of sample selection relevant for discrete choice models. Paper 3 through 5 represent my work on taste correlation in discrete choice models. Next, the contributions of the papers are described in more detail.

Paper 1: A note on choice-based sampling and nested GEV models

The first paper is a note on choice-based sampling. It is motivated by McFadden's result on choice-based sampling for MNL models. The note establishes a sufficient condition on sampling for a nested GEV consistency result, similar to that for MNL models. When we began working on the paper, a wrong claim circulated that McFadden's result could be generalised to NL models. The claim was shown to be wrong by Bierlaire et al. (2006) (BBM), and this was confirmed by our simulations. This led us to investigate why the claim was wrong. Based on that, we found the correction presented in the note for NL models and more general nested GEV models.[11] Essentially, the note points out that under certain restrictions the MNL result may be

[11] A nested GEV model has disjoint nests, i.e., a MNL structure, at the top level and any GEV structure at the lower levels.

extended to a broader class of models.

During the same period BBM added a section to their paper on block-additive GEV models. This was essentially the same result, though they mention that their proof does not cover NL models. In my opinion the results are the same. Due to the similarity to the results by BBM, we have decided not to submit the note for journal publication. Our result still serves to highlight that ESML estimation may be used in applied work when a NL model is appropriate under certain sampling restrictions.

Paper 2: Controlling for sample selection in the estimation of the value of travel time

The second paper investigates whether self-selection into a transport mode affects the estimation of VTT for that mode. The setup is that individuals in a SP sample for VTT estimation enter the sample based on their mode choice in a reference trip. Hence, by their mode choice individuals have self-selected into the sample for that mode. To describe this self-selection mechanism, we propose a new model. We adapt the Heckman model used in labour supply econometrics to VTT estimation. To do this, a linear regression equation is replaced by a mixed logit model used to estimate VTT. So in addition to the mixed logit model, the model also includes a probit model that accounts for the possible self-selection through instrumental variables. In the literature, the models are connected through correlation of the error terms. A new feature of our model is that we model correlation between an error term and a random coefficient.

The results show that self-selection can explain at least part of the difference in VTT between modes. The investigation highlights that it is difficult to find appropriate instruments. Furthermore, it appears that the results on self-selection depend on the specification of the mixed logit model.

The weak point of the model is the strong distributional assumptions we make. Especially the probit assumption in the selection equation is strong. The analogous assumption has been relaxed in the econometric labour supply literature. In our case the presence of a mixed logit model instead of a linear regression complicates the model considerably. Therefore it is difficult to use the nonparametric methods developed in the labour supply literature.

Paper 3: Estimation of correlated value of travel time in public transport

The third paper examines how theoretically motivated correlation structures affect the estimation of WTP measures in public transport. In the paper we apply mixed logit models to SP data on public transport route choice to investigate if the WTP indicators are correlated. The estimation of models with correlated coefficients shows that the standard independence assumption is unreasonable for the data investigated. Part of the correlation can be attributed to scale heterogeneity. Some of the scale heterogeneity may be explained by background variables. Based on LR and non-nested hypotheses tests, the standard mixed model is rejected. The mean WTP is similar across the mixed logit models with and without correlation.

The main conclusion of the paper is that mixed logit models should allow for taste correlation instead of mechanically using an independence assumption. The results are used in Nielsen & Mabit (2007). The paper applies the estimated distributions to assignment in the Copenhagen region.

Paper 4: Correlation and willingness-to-pay indicators in transport demand modelling

Many papers have investigated how different assumptions affect the estimation of mixed logit models. But these only look at the choice of a mixing distribution as a choice of marginal distributions. This paper investigates how different distributional assumptions are affected by taste correlation in a mixed logit model.

We assume distributions with different correlation structures on two different datasets. The starting point for the investigation is two approaches, which represent different distributional assumptions, known as modelling in preference space and WTP space. We investigate how both specification types compare to models allowing for more general correlation structures. In the case of models with all lognormal coefficients, these two approaches represent restricted correlation patterns and they can be embedded in a common model. The effect of correlation is also investigated in models where the scale and WTP is parameterised by background variables.

The results show that correlation is significant and that it may have an effect on WTP evaluation. In addition, the paper shows that the question is not whether to use preference space or WTP space but more generally how to include correlation as part of the specification of the mixing distribution. Again the results show that part of the correlation can be attributed to scale heterogeneity that may be explained by background variables.

Paper 5: Studies of correlated willingness-to-pay indicators for public transport

Paper 3 and 4 show that it is reasonable to expect taste correlation in mixed logit models. Based on this evidence, the final paper estimates correlation between WTP indicators in the case of public transport. The data consist of unlabelled route choices. The size of the dataset allows us to estimate a separate model for each mode. The estimation uses mixed logit models where the mixing distributions are lognormal distributions. In the final models, the distributional assumptions are tested.

For each mode, we develop a model that cannot be rejected by a model allowing for a full correlation matrix. This is done to get as simple a description as possible. The preferred model is then developed using background variables to see if they can explain part of the correlation. For all modes both observed and unobserved heterogeneity in the scale is present. Again this shows that part of the correlation may be attributed to a individual-specific scale.

The contribution of the paper is two-fold. First, it estimates correlation between the different WTP indicators relevant for public transport modelling. Second, it investigates how partial correlation structures and deterministic heterogeneity can explain correlation. This is novel as compared to papers in the literature that either assume independence or estimate a full correlation matrix.

As all three papers address taste correlation it is appropriate to discuss how they differ and what their combined contribution is. Concerning Paper 3 and 5 they share the same purpose as they seek to capture the taste correlation in WTP for public transport. Paper 3 precedes Paper 5 by 2 years and represents a procedure given a limited dataset. Paper 5 based on new data shows how it is possible to proceed given a large dataset for each public transport mode. Paper 4 concerns the question how the taste correlation interact with the other distributional assumptions made in a mixed logit model. Together the papers show that in all cases allowing for correlation leads to better model fit. An effect on WTP is seen in Paper 4 and for 2 out of 3 modes in Paper 5. In all papers, scale heterogeneity is present. It is seen to be partly explained by reference travel time and other background variables.

1.6 Perspectives

Looking at the contributions of the papers and the areas of research to which they are related, several directions for future investigations arise.

First, there is the question of how to direct state-of-the-art discrete choice modelling, e.g., mixed logit modelling, toward applied transport modelling. Whereas transportation research has seen an enormous amount of work on mixed logit models, the main part of application are still based on MNL models.[12] A simple explanation is that state-of-practice will always be behind state-of-the-art research. A second explanation is that the field of mixed logit modelling has not yet matured. In research, mixed logit modelling is shown to improve the modelling of most discrete choices both statistically and as an approximation to true behaviour. But there has not been much thought about when it is worthwhile to transfer these complex models to large-scale applications. In the same way, both identification and testing issues still seem largely overlooked in the research on mixed logit modelling.

Second, there are some technical aspects of mixed logit modelling that deserve more attention. One aspect is how many draws to use in the simulation (Walker 2001). In applications this is often based on a judgement about stability of the simulated log-likelihood and the parameter estimates. It would be interesting if something less abstract could be used. Another technical issue is how to test the distributions in a mixed logit model. Due to the long run times, especially in estimation, it is important to develop suitable tests. Some examples are the tests developed by McFadden & Train (2000) and Fosgerau & Bierlaire (2007). One source of new tests could be the tests that have been developed for logit models. It may be fruitful if they can be transferred to the case of mixed logit. One example of this is the White test. This is a useful tool to detect mis-specification for binary logit models (Lechner 1991). Maybe it could also be used to avoid mis-specification in the case of mixed logit models.

[12]Many applications done by consultants are not public so the view of state-of-practice presented here is only based on partial information.

As a final comment, I think there is more work to be done about the influence of prior choices on SP experiments. The model presented in Chapter 3 is one example as it controls for mode choice in the case of VTT estimation. Train & Wilson (2007) also focus on how prior choices affect estimation based on SP data. Together the two papers highlight that SP designs used today should be examined in the same sense as Bradley & Daly (2000) investigate adaptive designs.

References

Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, MA, USA. [3, 4]

Bhat, C. R. (2001), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model', *Transportation Research Part B: Methodological* **35**(7), 677–693. [6]

Bhat, C. R. (2003), 'Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences', *Transportation Research Part B: Methodological* **37**(9), 837–855. [6]

Bierlaire, M., Bolduc, D. & McFadden, D. (2006), 'The estimation of generalized extreme value models from choice-based samples', Forthcoming in Transportation Research B. [9, 11]

Bolduc, D., Khalaf, L. & Moyneur, E. (2001), 'Joint discrete/ continuous models with possibly weak identification [draft]', *5th tri-annual Invitational Choice Symposium on Hybrid Choice Models, UC Berkeley, June 2001* . [10]

Bolduc, D. & McFadden, D. (2001), Methodological developments in travel behaviour modelling, *in* D. Hensher, ed., 'Travel Behaviour Research, The Leading Edge', Pergamon Press, Australia, pp. 227–237. [10]

Bradley, M. & Daly, A. J. (2000), New analysis issues in stated preference research, *in* J. de D. Ortúzar, ed., 'Stated Preference Modelling Techniques', PTRC, London, chapter 2.6, pp. 121–136. [15]

Caussade, S., Ortúzar, J. de D., Rizzi, L. I. & Hensher, D. A. (2005), 'Assessing the influence of design dimensions on stated choice experiment estimates', *Transportation Research Part B: Methodological* **39**(7), 621–640. [7]

Dagsvik, J. (1994), 'Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes', *Econometrica* **62**(5), 1179–1205. [10]

Daly, A. & Carrasco, J. (2006), *The influence of trip length on marginal time and money values*, paper presented at the 11^{th} International Conference of Travel Behaviour Research, Kyoto, Japan. [10]

DeSerpa, A. (1971), 'A theory of the economics of time', *The Economic Journal* **81**, 828–846. [8]

Domencich, T. A. & McFadden, D. (1975), *Urban Travel Demand : A behavioural analysis*, North Holland, Amsterdam. [1]

Dubin, J. A. & McFadden, D. (1984), 'An econometric analysis of residential electric appliance holdings and consumption', *Econometrica* **52**, pp. 345–362. [10]

Fosgerau, M. (2006), 'Investigating the distribution of the value of travel time savings', *Transportation Research Part B: Methodological* **40**(8), 688–707. [5]

Fosgerau, M. & Bierlaire, M. (2007), 'A practical test for the choice of mixing distribution in discrete choice models', *Transportation Research Part B* **41 (7)**, 784–794. [5, 14]

Fosgerau, M. & Hess, S. (2006), *Competing methods for representing random taste heterogeneity in discrete choice models*, European Transport Conference, Strasbourg, 2006. [5]

Fosgerau, M., Hjorth, K. & Lyk-Jensen, S. V. (2007), *Between mode differences in the value of travel time: Self-selection or strategic behaviour?*, European Transport Conference, Leiden, 2007. [10]

Greene, W. H., Hensher, D. A. & Rose, J. M. (2006), 'Accounting for heterogeneity in the variance of unobserved effects in mixed logit models', *Transportation Research Part B: Methodological* **40**(1), 75–92. [11]

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47(1)**, 153–162. [9]

Hensher, D. A. (2001), 'The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications', *Transportation* **28**, 101–118. [4]

Hensher, D. A. (2006), 'The signs of the times: Imposing a globally signed condition on willingness to pay distributions', *Transportation* **33**, 205–222. [5]

Hess, S., Bierlaire, M. & Polak, J. W. (2005), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236. [5]

Hess, S., Polak, J. W. & Bierlaire, M. (2006), *Confounding between taste heterogeneity and error structure in discrete choice models*, European Transport Conference, Strasbourg. [11]

Hess, S., Train, K. & Polak, J. W. (2006), 'On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice', *Transportation Research Part B: Methodological* **40**(2), 147–163. [6]

Hjort, K. (2007), 'Om forskelle i værdien af rejsetid på tværs af transportmidler: Selv-selektion eller strategisk opførsel'. Master Thesis, Department of Economics, University of Aarhus, (in Danish). [6]

Horowitz, J. (1993), 'Semiparametric estimation of a work trip mode choice model', *Journal of Econometrics* **58**, 49–70. [1, 6]

Huber, J. & Train, K. E. (2001), 'On the similarity of classical and bayesian estimates of individual mean partworths.', *Marketing Letters* **12 (3)**, 259–269. [11]

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. [7]

Lechner, M. (1991), 'Testing logit models in practice', *Empirical Economics* **16**, 177–198. [14]

Mackie, P. J., Jara-Diaz, S. R. & Fowkes, A. S. (2001), 'The value of travel time savings in evaluation', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 91–106. [7]

Manski, C. & Lerman, S. (1977), 'The estimation of choice probabilities from choice based samples', *Econometrica* **45(8)**, 1977–1988. [8]

McFadden, D. (1981), Econometric models of probabilistic choice, *in* C. F. Manski & D. McFadden, eds, 'Structural analysis of discrete data with econometric applications', MIT Press, Cambridge, MA, pp. 198–272. [3]

McFadden, D. (2000), *Disaggregate Behavioral Travel Demand's RUM Side*, IATBR Conference. [3, 5]

McFadden, D. (2007), 'The behavioral science of transportation', *Transport Policy* **14**, 269–274. [1]

McFadden, D. & Train, K. E. (2000), 'Mixed mnl models for discrete response', *Journal of Applied Econometrics* **15 (5)**, 447–70. [1, 3, 4, 6, 14]

Nielsen, O. A. & Mabit, S. L. (2007), *The effect of correlated Value of Travel Time Savings in public transport assignment*, working paper. [13]

Revelt, D. & Train, K. (1998), 'Mixed logit with repeated choice: Household's choices of appliance efficiency level', *The Review of Economics and Statistics* **80**(4). [5, 11]

Self, S. G. & Liang, K.-Y. (1987), 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions', *Journal of the American Statistical Association* **82**(398), 605–610. [6]

Small, K., Winston, C. & Yan, J. (2005), 'Uncovering the distribution of motorists' preferences for travel time and reliability', *Econometrica* **73**(4), 1367–1382. [2]

Sørensen, M. V. (2003), Discrete Choice Models. Estimation of passenger traffic., PhD thesis, Danish Technical University, DK-2800 Kgs. Lyngby, Denmark. [11]

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY, USA. [3, 6]

Train, K. & McFadden, D. (1978), 'The goods/leisure tradeoff and disaggregate work trip mode choice models', *Transportation Research* **12**, 349–353. [3]

Train, K. & Sonnier, G. (2005), Mixed logit with bounded distributions of correlated partworths, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 7, pp. 117–134. [11]

Train, K. & Weeks, M. (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 1, pp. 1–16. [8, 11]

Train, K. & Wilson, W. (2007), 'Estimation on stated-preference experiments constructed from revealed-preference choices', *Transportation Research B, forthcoming* . [15]

Tversky, A. & Kahneman, D. (1991), 'Loss aversion in riskless choice: A reference-dependent model', *The Quarterly Journal of Economics* **106 (4)**, 1039–61. [1]

Vella, F. (1992), 'Simple tests for sample selection bias in censored and discrete choice models', *Journal of Applied Econometrics* **7 (4)**, 413–421. [10]

Vella, F. (1998), 'Estimating models with sample selection bias: A survey', *The Journal of Human Resources* **33(1)**, 88–126. [9]

Walker, J. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables, PhD thesis, MIT. [4, 14]

Wardman, M. (2004), 'Public transport values of time', *Transport policy* **11**, 363–377. [10]

Chapter 2

A note on choice based sampling and nested GEV models

Stefan L. Mabit
Centre for Traffic and Transport, Technical University of Denmark

Mogens Fosgerau
The Danish Transport Research Institute, Technical University of Denmark

Abstract

It is well known that all coefficients except the constants in a multinomial logit (MNL) model are estimated consistently on a choice-based sample with exogenous sample maximum likelihood (ESML) estimation. It has recently been demonstrated that this is not a general feature of generalised extreme value (GEV) models. This note establishes a sufficient condition for a nested GEV consistency result, similar to that for MNL models. As a special case the result is valid for nested logit (NL) models. Simulation is used to illustrate that ESML estimation of NL models may lead to inconsistent estimates on choice-based samples in cases where the sufficient condition is not fulfilled.

2.1 Introduction

In general exogenous sample maximum likelihood (ESML) estimation is inconsistent under stratified sampling for discrete choice models (see Manski & Lerman 1977). This note shows that, for some sampling schemes, ESML estimation of nested generalised extreme value (GEV) models is consistent for the coefficients except for the alternative specific constants. It also shows how to correct the inconsistencies in the alternative specific constants. Hence this is a generalisation of McFadden's result for multinomial logit (MNL) models described in Manski & Lerman (1977).

A recent article by Bierlaire et al. (2006), hereafter BBM, shows that choice-based sampling may lead to inconsistencies in coefficients. Furthermore, they identify a GEV-type model where estimation can be done by ESML.

We propose a general correction to nested GEV, e.g., nested logit (NL) models, under a specific assumption on the sampling scheme. This note is intended to supplement the BBM paper. It does this through a small technical generalisation. Though only slightly more general, it has an important implication for practice since the result covers NL models.

We will look at choice-based sampling when the distribution of choices in each nest of a nested GEV model is independent of sampling. A nested GEV is a two level GEV model where the upper level is governed by a MNL model and choice in each nest follow a general GEV model. A precise definition is given in Section 2.4.

As an example of an application think of a NL model of mode choice with the choice between car and public transport on top. If we are especially interested in one specific public transport mode we can oversample all public transport keeping the within nest frequencies the same as in the population.

The note is organised as follows. Section 2.2 introduces notation on GEV models. In Section 2.3 theory on sampling is presented and in Section 2.4 the main result on choice-based sampling and nested GEV models is established. In Section 2.5 we give two examples of estimation on choice-based samples that underline the theoretical result. In the last section we conclude on the results.

2.2 Discrete choice and GEV models

Suppose a discrete choice between J alternatives $j \in C$ can be described through probabilities $P(i|x)$ where we condition on explanatory variables x, e.g., attributes and socioeconomic variables. A common approach in discrete choice modelling is to assume that choices follow a GEV model. A GEV model, where V_j is the systematic part of utility, has probabilities given by

$$(2.1) \qquad P(i|x) = \frac{e^{V_i}G_i}{\mu G} = \frac{e^{V_i}G_i}{\sum_j e^{V_j}G_j},$$

where G (μ-GEV generating function) is a function of all e^{V_j} and G_i is the derivative of G with respect to the ith variable, (Ben-Akiva & Lerman 1985). To simplify notation the overall scale μ is assumed to be unity in this note unless otherwise stated.

Two special cases of GEV models are most prominent in applications: the MNL model and the NL model. The MNL model has the generating function $G = \sum_j e^{V_j}$, hence

$$P(i|x) = \frac{e^{V_i}}{\sum_j e^{V_j}}.$$

In a NL model the choice set C is partitioned into subsets or nests A_m, i.e., each alternative i belongs to exactly one nest A_m.[1] These models are based on GEV generating functions;

$$(2.2) \qquad G = \sum_m \Big(\sum_{j \in A_m} e^{V_j \mu_m} \Big)^{\frac{1}{\mu_m}}.$$

2.3 Sampling and estimation

Suppose a sample has been collected to estimate a discrete choice model. The most common assumption is that the sample is random given the population of interest.[2] Suppose the population has N_p individuals and that N_s of them are collected in a sample. Let s be a dummy variable indicating whether an individual is included in the sample. For a random sample

$$P(s) = P(s|i) = \frac{N_s}{N_p}$$

where s is short for $s = 1$. In a random sample the fraction, H_i, that chooses alternative i is the same as the fraction, $Q_i = \sum_x P(i|x)P(x)$, that chooses i in the population. In a choice-based sample, H_i is set by design. In a given application H_i will be derived from the sample after data cleaning it may then differ from the value set by design. Besides Q_i must be assumed to be known or derived from an auxiliary sample.

By definition of a choice-based sample we have

$$P(s|i,x) = P(s|i) \text{ or equivalently } P(x|i,s) = P(x|i).$$

For a given choice-based sampling scheme we define r_i by $r_i = P(s|i) = \frac{H_i N_s}{Q_i N_p}$. The probability for alternative i in the sample is

$$\begin{aligned} P(i|x,s) &= \frac{P(i,s|x)}{P(s|x)} = \frac{P(i|x)P(s|i,x)}{\sum_j P(j,s|x)} \\ &= \frac{P(i|x)P(s|i,x)}{\sum_j P(j|x)P(s|j,x)} = \frac{P(i|x)P(s|i)}{\sum_j P(j|x)P(s|j)} \\ &= \frac{P(i|x)r_i}{\sum_j P(j|x)r_j}. \end{aligned}$$

Now we can calculate the probabilities given a choice-based sample in a GEV population. Let $\alpha_j = log(\frac{H_j}{Q_j})$. Since N_s/N_p cancels out:

$$(2.3) \qquad P(i|x,s) = \frac{r_i e^{V_i} G_i / \sum_k e^{V_k} G_k}{\sum_j (r_j e^{V_j} G_j / \sum_k e^{V_k} G_k)}$$

$$(2.4) \qquad = \frac{e^{V_i + \log r_i} G_i}{\sum_j e^{V_j + \log r_j} G_j} = \frac{e^{V_i + \alpha_i} G_i}{\sum_j e^{V_j + \alpha_j} G_j}.$$

[1] These models are also known as tree logit models.

[2] Here random sampling signifies that the population in sampled uniformly.

For the special case of a MNL model, where $G_i = 1 \forall i$, we get

$$P(i|x,s) = \frac{e^{V_i+\alpha_i}}{\sum_j e^{V_j+\alpha_j}}, \tag{2.5}$$

where $\alpha_j = log(\frac{H_j}{Q_j})$. This shows that the probabilities in the sample are described by a MNL model that only differs from the population MNL model with respect to the alternative specific constants. ESML estimation consists of maximising the likelihood function as if the sample population was random. As the name indicates the procedure can be applied to any exogenous sample not just random. Since this gives consistent estimates for a MNL model under random sampling, the sample probabilities in Equation 2.5 establishes directly the correction in Manski & Lerman (1977) for MNL models under choice-based sampling.[3] Since the alternative specific constants are shifted the model has to allow for $J-1$ alternative specific constants in the estimation.

2.4 Choice-based sampling and nested GEV models

First we need to introduce some notation and define nest-level choice-based sampling. Suppose that we obtain a choice-based sample such that in any given nest, the distributions of choices are the same in the sample and the population, i.e.,

$$P(i|m,s) = P(i|m). \tag{2.6}$$

This is termed nest-level choice-based sampling. Denoting $H_i^m = P(i|s,m)$ and $Q_i^m = P(i|m)$, let $\alpha_i^m = ln(\frac{H_i^m}{Q_i^m})$ and $\alpha_m = ln(\frac{H_m}{Q_m})$ where H_m(resp. Q_m) is the fraction choosing nest m in the sample (resp. population). Note that

$$\alpha_i = ln(\frac{H_i}{Q_i}) = ln(\frac{H_m H_i^m}{Q_m Q_i^m}) = \alpha_i^m + \alpha_m.$$

Assumption 2.6 is equivalent to $\alpha_i^m = 0$.

Next we define nested GEV models. Suppose that A_m is a partitioning of a choice set C in mutually exclusive nests and that for each nest we have a μ_m-GEV generating function, $G^m(e^{V_j}; j \in A_m))$. G^m is a function of e^{V_j} for $j \in A_m$ only. Since the family of GEV functions is stable under addition and raising to powers we have that

$$G = \sum_m (G^m(e^{V_j}; j \in A_m))^{\frac{1}{\mu_m}}$$

is a GEV function. This is what we call a nested GEV model.[4] It is 1-homogenous hence the derivative with respect to the j*th* argument, G_j, is 0-homogenous, and G_j only depend on V_j for $j \in A_m$.

[3] This same idea is used by BBM. It is repeated here because we use the procedure in Section 2.4.

[4] Another way to characterise a nested GEV is to say that we have disjoint nests at the top level, i.e., a MNL structure, and any multilevel GEV structure at the lower levels.

Now we are ready to show the result about nested GEV models under nest-level choice-based sampling. We use a star to emphasise the true parameter values in the population. From Equation 2.4 we get that the sample model is

$$P(i|x,s) \quad = \quad \frac{e^{V_i^\star+\alpha_i}G_i}{\sum_{j\in C} e^{V_j^\star+\alpha_j}G_j} = \frac{N_i}{D}. \tag{2.7}$$

Now

$$\begin{aligned} N_i &= e^{V_i^\star+\alpha_i}G_i(e^{V_j^\star}; j\in A_m) \\ &= e^{V_i^\star+\alpha_m}G_i(e^{\alpha_m}e^{V_j^\star}; j\in A_m) \end{aligned}$$

where $G_i(e^{V_j}; j\in A_m) = G_i(e^{\alpha_m}e^{V_j}; j\in A_m)$ follows because of $0-$homogeneity of G_i, and $\alpha_i = \alpha_m$ because of nest-level choice-based sampling. The same could be done to D hence Equation 2.7 becomes

$$P(i|x,s) = \frac{e^{V_i^\star+\alpha_m}G_i(e^{\alpha_m}e^{V_j^\star})}{\sum_{j\in C} e^{V_j^\star+\alpha_{m(j)}}G_j(e^{\alpha_{m(j)}}e^{V_j^\star})} = \frac{e^{V_i'}G_i(e^{V_j'})}{\sum_{j\in C} e^{V_j'}G_j(e^{V_j'})}. \tag{2.8}$$

The result in Equation 2.8 shows that the sample model only differs from the population model in the constants. Therefore, exactly as in the case for MNL, ESML estimation will give consistent parameter estimates, except for the constants that need to be shifted by α_m. Furthermore, if we assume $V_i^\star = \gamma_i^\star + \beta^\star x_i$ then $V_i' = \gamma_i' + \beta^\star x_i$ with

$$\gamma_i' = \gamma_i^\star + \alpha_m. \tag{2.9}$$

Since $\alpha_m = ln(\frac{H_m}{Q_m})$, the correction is easily calculated. So a consistent estimator of the constants is obtained by subtraction of a correction factor for each alternative specific constant. Hence the procedure is to estimate γ_i', to correct them using α_m, and then to normalise the constants by setting a suitable constant equal to zero. Again the model has to allow for $J-1$ alternative specific constants in the estimation for the correction to work.

The above result for nested GEV models under nest-level choice-based sampling is seen to have the block additive GEV model described by BBM as a special case. For applications a more important result is that NL models under nest-level choice-based sampling can be estimated by ESML estimation.

2.5 Examples

To illustrate the corrections for estimation based on choice-based samples we consider two examples using a simulated population. We have constructed a population following a NL model. In the first example one nest is oversampled. Here Assumption 2.6 is fulfilled. In the second example we oversample one alternative in a two alternative nest. In this example, Assumption (2.6) is not fulfilled. All estimation is performed with Biogeme (Bierlaire 2005).

Table 2.1: ESTIMATION RESULTS FOR EXAMPLE I

coefficient	true	ESML		t test	WESML		t test
LL	—	-4010.0		—	-7548.0		—
asc_{car}	0.0	0.0		—	0.0		—
asc_{bus}	1.0	-1.73	(4.7e-2)	-58.1	0.97	(5.1e-2)	-0.59
asc_{rail}	1.0	-1.74	(4.3e-2)	-63.7	0.96	(4.6e-2)	-0.87
β_{cost}	-1e-2	-9.34e-3	(5.7e-4)	1.16	-9.62e-3	(8.6e-4)	0.44
$\beta_{time-car}$	-2e-2	-2.03e-2	(1.1e-3)	-0.27	-2.01e-2	(1.4e-3)	-0.07
$\beta_{time-bus}$	-3e-2	-2.86e-2	(1.4e-3)	1.00	-2.87e-2	(1.5e-3)	0.87
$\beta_{time-rail}$	-3e-2	-2.88e-2	(1.3e-3)	0.92	-2.90e-2	(1.4e-3)	0.71
μ	4.0	4.15	(3.2e-1)	0.47	4.11	(3.2e-1)	0.34

We created a dataset with 112110 observations choosing between 3 alternatives: car, bus and rail. Each alternative has the two attributes cost and travel time. The generated attributes were created based on revealed travel costs and times from a recent Danish survey, DATIV. It was assumed that this population followed a NL model with parameters given as true parameters in Table 2.1 where bus and rail are in the same nest. The choices were then simulated. Following the simulation the choice frequencies (Q_i) became 0.2937, 0.1920, and 0.5143.

2.5.1 Example I

From the population we pick a choice-based sample with 10000 observations using the following frequencies (H_i): 0.8587, 0.0384 and 0.1029. In this way the relative within nest probabilities are the same in the sample and the population such that Assumption 2.6 is satisfied. Estimating a NL model on this sample we obtain the results in Table 2.1 with robust standard errors in parenthesis. We report the results of a corresponding Weighted ESML (WESML) estimation as well (Manski & Lerman 1977).[5] The t tests are against the true parameters. We see that the coefficients from ESML and WESML are similar except for the alternative specific constants. We also see that the standard deviations for WESML are slightly larger as expected. The result shows that the constants are rejected, while the other coefficients are not rejected in a t test against the true values. Using the ESML estimates of the alternative specific constants we calculated the alternative specific constants using the correction derived in Section 2.4. The corrected constants are reported as normalised in Table 2.2, i.e., we derive the corrected constants and then we normalise the car constant to zero to make them comparable to the true constants. The corrections come from Equation 2.9. The corrected alternative specific constants are now acceptably close to the true ones.

[5] The WESML is seen to give a very different log-likelihood. This is a result of the weighting. The t tests are still asymptotically valid.

Table 2.2: CORRECTION OF CONSTANTS

coefficient	true	estimated	corrected
car	0.0	0.0	0.0
bus	1.0	-1.73	0.95
rail	1.0	-1.74	0.94

Table 2.3: ESTIMATION RESULTS FOR EXAMPLE II

coefficient	true	ESML		t test	WESML		t test	ESML-fix	
LL	—	-2825.8		—	-7503.4		—	-2838.0	
asc_{car}	0.0	0.0		—	0.0		—	0.0	
asc_{bus}	1.0	3.64	(7.5e-2)	35.2	1.01	(8.2e-2)	0.12	3.43	(4.7e-2)
asc_{rail}	1.0	2.40	(8.3e-1)	1.69	0.99	(8.9e-2)	-0.11	2.42	(4.9e-2)
β_{cost}	-1e-2	-1.29e-2	(8.5e-4)	-3.41	-1.10e-2	(1.2e-3)	-0.83	-1e-2	
$\beta_{time-car}$	-2e-2	-2.31e-2	(2.1e-3)	-1.48	-2.17e-2	(2.4e-3)	-0.71	-2e-2	
$\beta_{time-bus}$	-3e-2	-3.66e-2	(1.8e-3)	-3.67	-3.26e-2	(2.6e-3)	-1.00	-3e-2	
$\beta_{time-rail}$	-3e-2	-3.83e-2	(2.1e-3)	-3.95	-3.27e-2	(2.9e-3)	-0.93	-3e-2	
μ	4.0	3.10	(2.0e-1)	-4.50	3.72	(3.8e-1)	-0.74	4.0	

2.5.2 Example II

In this example we select a different choice-based sample from the same population. Recall that the population choice frequencies are 0.2937, 0.1920, and 0.5143. We pick a choice-based sample with 10000 observations, this time using the frequencies (H_i): 0.05, 0.9 and 0.05. The oversampling of alternative 2 violates Assumption 2.6. Estimating a NL model on this we obtain the results in Table 2.3 with robust standard errors in parenthesis. The t tests are against the true parameters.

We see that the parameters are recovered with WESML estimation. For the ESML estimation we see that the coefficient on cost, time in bus and rail together with μ all would be rejected in t tests against the true parameter. As a second test we estimated a model where we fixed the coefficients other than the alternative specific constants to the true values. In a likelihood ratio (LR) test of this model against the ESML model we get an LR statistic of 24.4 which has to be compared with the 0.05 level for a χ^2_5 distribution. Since this is 11.07 we can reject the ESML-fixed model at the 0.05 level. So we have a rejection of the model estimated by ESML both with t tests and with the LR test. This supports that ESML estimation for NL models is inconsistent under choice-based sampling in general.

2.6 Conclusion

We have shown how to correct nested GEV models in a simple way when they are estimated using ESML under nest-level choice-based sampling. This generalises the result by BBM in an important way since the nested GEV model family includes NL models. This could be very useful in applications.

A natural question for further research is whether the consistency result can be generalised to a broader class of models or to a broader class of sampling schemes or whether the present result exhausts all possible cases.

Acknowledgment
Part of this research was carried out during a research stay of the first author at EPFL, Switzerland. The stay was funded by Andersen's Travel Grant to Mathematicians. The authors would like to thank Professor Michel Bierlaire for valuable comments.

References

Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, MA, USA. [20]

Bierlaire, M. (2005), 'An introduction to biogeme (version 1.4)', biogeme.epfl.ch. [23]

Bierlaire, M., Bolduc, D. & McFadden, D. (2006), 'The estimation of generalized extreme value models from choice-based samples', Forthcoming in Transportation Research B. [20]

Manski, C. & Lerman, S. (1977), 'The estimation of choice probabilities from choice based samples', *Econometrica* **45(8)**, 1977–1988. [19, 22, 24]

Chapter 3

Controlling for sample selection in the estimation of the value of travel time

Stefan L. Mabit
Centre for Traffic and Transport, Technical University of Denmark

Mogens Fosgerau
The Danish Transport Research Institute, Technical University of Denmark

Abstract

It is often found that the value of travel time (VTT) is higher for car drivers than for public transport passengers. Here we investigate whether self-selection into transport mode on the basis of unobserved individual VTT can explain this finding. We use a mixed logit model to estimate VTT together with a probit model to account for the possible self-selection using instrumental variables. We find that self-selection seems to explain at least part of the difference in VTT between modes. The investigation highlights that it is difficult to find appropriate instruments and that the results on self-selection are highly dependent on the specification of the mixed logit model.

3.1 Introduction

This paper investigates how value of travel time (VTT) estimates are affected by self-selection into transport modes. The term self-selection means that individuals choose a specific mode partly based on their VTT. Self-selection has implication for the interpretation of VTT estimates, e.g., a relatively high VTT for a sample of car drivers may reflect either discomfort of travelling by car or a self-selection effect whereby individuals with high VTT are more likely to choose to travel by car. We find that accounting for self-selection can reduce the difference in VTT obtained for car and public transport (PT). Our ability to account for self-selection depends of course on the availability of good instruments. Such instruments are generally hard to find and our case is no exception. With better instruments, we conjecture that results can be correspondingly stronger.

The VTT can be estimated within the framework derived from DeSerpa (1971) together with a general theory on discrete choices, see e.g., Train (2003). This framework has shown its use both with revealed-preference (RP) data and stated-preference (SP) data, for recent applications and references, see e.g., Hensher (2001), Axhausen et al. (2004), and Sillano & Ortúzar (2004). The VTT is found as the ratio of the estimated time and cost coefficients in the discrete choice model.

Much research has been devoted to the specification of discrete choice models and in the case of mixed logit models to the choice of distribution for the coefficients. Very little research has focused on the role of the sample except for the case of choice-based sampling. A general question concerning the effect of the sample would be the following: If we want to estimate the average VTT for a given population, which part of the population do we sample and how do we correct the results if we sample a subpopulation that is not representative.

Suppose our interest is to estimate the average VTT in car for some population, based on SP data. Then we would choose a subpopulation and have them make SP choices, e.g., route choices by car. Two possible ways of choosing the subpopulation would be either a random sample from the general population or a sample of people using car. The first sample fulfills the statistical condition of random sampling that is often assumed in most applied work but raises the effort needed to assure realism in the choice task. This happens because we do not know what experience a person never using a car refers to when choosing between alternative car trips. The second sample has the opposite characteristics because every individual has a reference trip but the sample is not necessarily random in the population. The second sample is necessary for SP samples based on pivoting, see Train & Wilson (2007) for references.

The problem that could arise from using the second sample is best illustrated by an example. Suppose everybody in a population has the same VTT in both car and bus. But the population is divided into two equal-sized groups: one with high VTT and one with low. Suppose that for this population car is fastest and the only thing that matters in a mode choice is travel time and cost. Furthermore, suppose costs are such that half of the population travels by car. In this population, everybody with high VTT uses car and everybody with low VTT uses bus. Hence a SP experiment to infer VTT based on a sample of car users would not yield the average VTT in car for the population. The above is a simplified illustration of user-type effects. As discussed by Wardman (2004) these are often confounded with mode effects in applications.

One indication that this could be a real problem is that VTT in car is often estimated to be higher than VTT in PT for individuals who have identical observable characteristics, see references in Axhausen et al. (2004) or results in Section 3.3.4. This is counter-intuitive following the theory on VTT as the difference between modes mainly should reflect differences in comfort. According to the theory, higher VTT in car implies that people should feel less comfortable in car than in PT.

Self-selection has been investigated in a different context in the transportation field. In the case of choice-based sampling the problem is to estimate a discrete choice model describing some choice of interest based on a sample influenced by the same choice. Hence individuals self-select into a choice-based sample. It has been shown that random sample estimation procedures lead to inconsistent estimates on choice-based samples (Manski & Lerman 1977). So if self-selection is present in the VTT context it may also lead to inconsistent estimates.

Self-selection resembles the mechanism behind certain endogeneity issues. Train & Wilson (2007) discusses the endogeneity that arises when SP choices are designed based on RP choices, e.g., some pivoted SP designs. The endogeneity they discuss arises, e.g., when SP mode choices are based on a RP mode choice. The self-selection we describe arises when SP choices on route choice are based on a RP mode choice.

In the field of labour supply the problem of self-selection has been studied for many years, see Heckman (1979) for an early reference. In a labour supply model the interest is to estimate a wage equation, i.e., the mean ln(wage) as a function of some explanatory variables. The problem is that the expected wage affects participation in the labour market and that inference on wage equations is based on individuals observed in the labour market. Thus individuals select to be part of the labour market, and hence the sample population, based on their expected wage. The model described by Heckman consists of a selection model describing whether an individual participates in the labour market and a wage equation. In the original model the selection model is a probit model and the wage equation is a linear regression model. So the standard model is enlarged by a selection model to capture the process of self-selection. Self-selection corresponds to correlation between these two equations. Central to this model are the explanatory variables that enter the selection equation and not the wage equation. These variables act as instruments for the selection. The approach by Heckman has also been applied by Vella (1992) to the case where both the selection and the main equation are discrete. In both cases, the models are connected through correlation of the error terms. This approach has also been used in contingent valuation where sample selection was seen to have a significant effect on willingness-to-pay estimation (Eklof & Karlsson 1999).

Here we adapt the approach from labour supply to the VTT context. We use a mixed logit model as VTT model together with a probit model to model the selection into modes. The VTT model allows for random VTT reflecting unobserved heterogeneity. Since interest is on VTT the concern is whether the sampling is connected with the coefficients in the model and not the additive error. Therefore our model, though resembling the model discussed by Vella, is different since the correlation is captured as an interaction between the selection equation and the random coefficients of the mixed logit model.

The remainder of this paper is organised as follows. In Section 3.2.1 the selection model is presented and in Section 3.2.2 the VTT model is presented with several specifications. The simultaneous model is described in Section 3.2.3 and calculation of the mean VTT is discussed in Section 3.2.4. Section 3.3.1 contains a discussion of the data and Section 3.3.2 discusses the choice of explanatory variables in both models. Section 3.3.3 discusses the estimation of the models. The resulting VTT estimates are presented in Section 3.3.4 with a discussion of results in Section 3.3.5. The final Section 3.4 contains some concluding remarks.

3.2 Model formulation

This section presents the general model. The model estimates VTT from a panel mixed logit model and simultaneously estimates a probit model that controls for the effect of sampling. The probit model will be referred to as the selection model and the mixed logit model as the VTT model.

3.2.1 Selection model

The selection model is a binary probit model describing choice between car and PT. Let n denote a given individual. Then the binary outcome of the selection, Y_n, depends on a latent variable U_{1n}

$$Y_n = 1\{U_{1n} > 0\},$$

where $1\{\}$ denotes the indicator function and U_{1n} is a latent variable determining the choice. Let $Y_n = 1$ denote that the individual chooses car. Then U_{1n} denotes the utility of car minus the utility of PT. We assume that conditional on explanatory variables and coefficients the latent variable is the sum of a deterministic part and an independent random error term:

$$U_{1n} = \gamma' x_{1n} + u_{1n},$$

where $u_{1n} \sim N(0,1)$ and x_{1n} denotes the explanatory variables including a constant. From these assumptions we find that

$$P(Y_n = 1|\gamma, x_{1n}) = \Phi(\gamma' x_{1n}).$$

We require that x_{1n} contains at least one variable that does not enter the VTT model and at the same time, is uncorrelated with the random part of VTT. It is through such variables, generally known as instruments, the model derives its ability to control for the effect of self-selection. The choice of instruments depends on the data available therefore this is discussed in Section 3.3.2.

3.2.2 VTT models

The VTT model is a mixed logit model for panel data. The model applies to a panel of binary choices, in our case we have unlabelled route choices. We introduce the notation where each alternative has just two attributes: cost and time. The model allows for repeated choices from each individual:

$$\begin{aligned} Z_{nt} &= 1\{U_{2nt} > 0\}, \\ Z_n &= \{Z_{n1}, Z_{n2}, ..., Z_{nT_n}\}. \end{aligned}$$

Each choice depends on the latent variable U_{2nt} that corresponds to the difference in utility between the two alternatives. The latent variable is decomposed into a random term and a systematic term conditional on random taste coefficients

$$U_{2nt} = V_{nt} + \varepsilon_{nt} = \alpha_n^C \Delta C_{nt} + \alpha_n^T \Delta T_{nt} + \varepsilon_{nt},$$

where ε_{nt} is independent mean zero logistic and we have omitted the alternative specific constants because the SP experiment is unlabelled. The ΔC_{nt} and ΔT_{nt} refer to the difference in the cost attributes and time attributes between the two alternatives in the SP choice. Under the above assumptions we have

$$P(Z_{nt} = z_{nt} | V_{nt}) = \frac{e^{z_{nt} V_{nt}}}{1 + e^{V_{nt}}}$$

and

$$P(Z_n = z_n | V_n) = \prod_{t=1}^{T_n} \frac{e^{z_{nt} V_{nt}}}{1 + e^{V_{nt}}}. \tag{3.1}$$

Here V_n denotes the vector $\{V_{nt}\}$. An important question in a mixed logit model is which distribution to choose for the coefficients. In this paper the parameters α_n^C, α_n^T are chosen so that VTT is lognormally distributed, i.e., the fraction $\frac{\alpha_n^T}{\alpha_n^C}$ follows a lognormal distribution. This distribution ensures positive VTT, and that the VTT as well as the inverse VTT has a well-defined mean. Another reason for the choice is that on the same dataset it has been shown to perform well in a nonparametric investigation of VTT, see Fosgerau (2006). The model allows for parameterisation of VTT with explanatory variables. This induces heterogeneity in VTT.

It is seen from Equation 3.1 how the probabilities depend on V_n. We will use different specifications of V_n that all lead to lognormal VTT. For all specifications the following notation is used:

$$VTT = \frac{\alpha_n^T}{\alpha_n^C} = e^{\beta + \sigma u_{2n}},$$

where $u_{2n} \sim N(0,1)$ signifies that VTT follows a lognormal distribution such that $\ln VTT \sim N(\beta, \sigma^2)$ distribution.

We investigate 3 different specifications. The first is a specification in preference space with independent coefficients:

$$V_{nt} \quad = \quad e^{-\beta_C' x_{2n} - \sigma_2 u_{2n}} \Delta C_{nt} + e^{\beta_T + \sigma_3 u_{3n}} \Delta T_{nt}, \tag{3.2}$$

where the coefficients are independently lognormally distributed and explanatory variables, x_{2n} enter the cost coefficient. Note that x_{2n} may overlap with x_{1n}. We use the notation $x_n = (x_{1n}, \tilde{x}_{2n})$ where $\tilde{x}_{2n}$ are variables in x_{2n} not in x_{1n}. Implicitly in the model is the standard, but strong, assumption that $u_n = (u_{1n}, u_{2n}, u_{3n})$ is independent of x_n. The choice to enter x_n in the cost coefficient is discussed in Section 3.3.3. The signs on the coefficients are chosen to ease comparison of coefficients across specifications. This specification is denoted Model I.

The second specification is an inverse VTT-space specification:

$$(3.3) \qquad V_{nt} \quad = \quad e^{-\beta' x_{2n} - \sigma_2 u_{2n}} e^{\beta_T + \sigma_3 u_{3n}} \Delta C_{nt} + e^{\beta_T + \sigma_3 u_{3n}} \Delta T_{nt}.$$

The name comes from the fact that we parameterise directly with the inverse of VTT (see Train & Weeks 2005, for a discussion of VTT space). Reasons for doing this are discussed in Section 3.3.3.[1] For another account of VTT space, see Fosgerau (2007), where the estimation in VTT space performs well when compared with alternative specifications. In the VTT-space specification correlation between selection and VTT can be modelled directly whereas it will only be indirectly through correlation with the cost or time coefficient in the model in Equation 3.2.

Model I and II are quite similar. The only difference is in the specification of the correlation structure between the coefficients. A different model is obtained by a log transformation of the data together with a change from an additive ε to a multiplicative ε relative to VTT. This leads to a mixed logit model in log VTT space. This model is investigated in more detail in Fosgerau (2007). The specification becomes:

$$(3.4) \qquad V_{nt} \quad = \quad -ln(-\Delta C_{nt}/\Delta T_{nt}) + \beta' x_{2n} + \sigma_2 u_{2n},$$

where $u_{2n} \sim N(0,1)$. In this model $VTT = e^{\beta' x_{2n} + \sigma_2 u_{2n}}$. We will refer to the model as Model III.

We will later refer to a standard estimation of VTT. This refers to estimation of the models above ignoring selection. The VTT calculated from the standard model estimation will be denoted VTT_s.

3.2.3 Simultaneous model

The model for selection from Section 3.2.1 and the models for VTT estimation above allow for interaction through correlation of the different normally distributed random terms. This gives a model where it is possible to test whether the selection equation influences the VTT estimation. Assuming that u_{1n} and u_{2n} follow a joint normal distribution we can use a Choleski factorisation to write

$$(3.5) \qquad u_{1n} = v_{1n},\ \sigma_2 u_{2n} = s_1 v_{1n} + s_2 v_{2n} \text{ and } \sigma_3 u_{3n} = s_3 v_{3n}$$

where v's are iid normal with mean zero and variance one. The model captures correlation between u_{1n} and u_{2n}, which is natural for Model II and III. For Model I it means that correlation is captured in the cost coefficient. This issue is discussed in Section 3.3.3.

Now we can derive the simultaneous likelihood for the selection and the SP choices. We observe a vector of choices Z_n when $Y_n = 1$. We condition on $x_n, \Delta T_n, \Delta C_n$, but leave this out of the

[1] To our knowledge, this is the first application of an inverse VTT specification.

notation.

$$
\begin{aligned}
P(Z_n = z_n, Y_n = 1) &= E\left(P\left(Z_n = z_n, Y_n = 1|v_n\right)\right) \\
&= E\left(P\left(Z_n = z_n|Y_n = 1, v_n\right)P\left(Y_n = 1|v_n\right)\right) \\
&= E\left(\prod_{t=1}^{T_n} \frac{e^{z_{nt}V_{nt}}}{(1+e^{V_{nt}})} 1\left\{v_{1n} > -\gamma' x_n\right\}\right) \\
&= \int\int\int_{-\gamma x_{1n}}^{\infty} \prod_{t=1}^{T_n} \frac{e^{z_{nt}V_{nt}}}{(1+e^{V_{nt}})} \phi\left(v_{1n}, v_{2n}, v_{3n}\right) dv_n,
\end{aligned}
$$

where we parameterise by $v = (v_1, v_2, v_3)$ and V_{nt} is given by Equations 3.2-3.4 with the factorisation in Equation 3.5. Together with $P(Y_n = 0) = \Phi(-\gamma' x_{1n})$ the expression above can be used to form the partial likelihood necessary for estimation. It is seen from the parameterisation in Equation 3.5 that s_1 captures the correlation between the selection and the VTT estimation. It is also seen that $s_1 > 0$ corresponds to positive correlation between VTT and Y.

We have focused on the case $Y_n = 1$(car) in the above calculations. The equivalent model for a panel of binary choices for individuals with $Y_n = 0$(PT) can be deduced in a similar way.

Even though the two models are estimated simultaneously, the causal interpretation is that selection occur prior to the SP choices.

3.2.4 VTT estimation

The VTT is lognormally distributed in all the models. To evaluate the models with selection and compare them to the standard estimation without selection we evaluate the mean VTT. This can be done either by averaging VTT over the sample or by choosing a representative individual. The first is appropriate in many applications but for the purpose at hand where we compare different models the second approach is more useful since model differences are not confused with sample characteristics. Therefore we evaluate the mean VTT using an individual with mean socioeconomic variables, i.e., $\bar{x}_k = \frac{1}{N}\sum_n x_{kn}$ and $\bar{x} = (\bar{x}_1, \bar{\bar{x}}_2)$.

We will describe the mean VTT calculation based on Model II. The calculations are similar for the other models. So assume that the systematic utility is given by Equation 3.3. First suppose that we estimate a model using a standard model without selection. Then we have

$$\frac{\alpha_T^s}{\alpha_C^s} = e^{\beta_s' x_2 + s_s u_2}.$$

From this the mean VTT evaluated at $\bar{x}$ is

$$E\left(\frac{\alpha_T^s}{\alpha_C^s}|\bar{x}\right) = e^{\beta_s'\bar{x}_2 + \frac{1}{2}(s_s^2)}. \tag{3.6}$$

From the model with selection in Equation 3.3 we have

$$\frac{\alpha_T}{\alpha_C} = e^{\beta' x_2 + s_1 v_1 + s_2 v_2},$$

where v_1, v_2 are independent normal. Therefore we get

$$E(VTT|\bar{x}) = E(\frac{\alpha_T}{\alpha_C}|\bar{x}) = e^{\beta'\bar{x}_2+\frac{1}{2}(s_1^2+s_2^2)}. \tag{3.7}$$

In the same way we can estimate the mean conditional on the individual $\bar{x}$ being a car user, i.e.,

$$E(\frac{\alpha_T}{\alpha_C}|\bar{x}, Y=1) = \int\int_{-\gamma'\bar{x}_1} e^{\beta'\bar{x}_2+s_1v_1+s_2v_2} f(v_1)f(v_2)dv_1dv_2 \tag{3.8}$$

$$= \frac{\Phi(\gamma'\bar{x}_1+s_1)}{\Phi(\gamma'\bar{x}_1)} e^{\beta'\bar{x}_2+\frac{1}{2}(s_1^2+s_2^2)}. \tag{3.9}$$

The expression is found by integration over the truncated lognormal distribution. Conditional on being a PT user the expression becomes:

$$E(\frac{\alpha_T}{\alpha_C}|\bar{x}, Y=0) = \frac{\Phi(-\gamma'\bar{x}_1-s_1)}{\Phi(-\gamma'\bar{x}_1)} e^{\beta'\bar{x}_2+\frac{1}{2}(s_1^2+s_2^2)}. \tag{3.10}$$

These expressions depend on γ, β and s. We could use the point estimates for these, but this ignores the fact that they are estimates (Hensher & Greene 2003). It is more appropriate to use the estimated asymptotic distribution of (γ, β, s). To do this for the model in Equation 3.7 we draw M times (β_m, s_m) and calculate

$$E(VTT_m|\bar{x}) = e^{\beta_m'\bar{x}_2+\frac{1}{2}(s_{1m}^2+s_{2m}^2)}.$$

Then we use the average

$$E(VTT|\bar{x}) = \frac{1}{M}\sum_m E(VTT_m|\bar{x})$$

as the mean VTT estimate from the model with selection. As an estimate of the variation in the mean VTT we calculate

$$std = (\frac{1}{M}\sum_m (E(VTT_m|\bar{x}) - (VTT|\bar{x}))^2)^{\frac{1}{2}}$$

and report this as the standard deviation of the estimated mean. Equation 3.6, 3.8, and 3.10 are simulated in the same way.

3.3 Data and Estimation

3.3.1 Data

The data are from the 2004 Danish VTT study known as DATIV, see Fosgerau et al. (2006). We use observations of commuters using car or PT, giving us 1425 individuals. Each individual was asked 9 SP choices in an unlabelled experiment referring to a current commute trip, i.e., car users only make car choices. Every choice was a binary choice where the alternatives were only described by travel time and cost. One of the SP choices was a check question, i.e., a choice where one alternative is slower and more expensive than the other. A total of 177 persons chose

Table 3.1: STATISTICS ON THE ATTRIBUTES

variable	mean	std.dev.	min	max	mean(pt)	mean(car)
ΔT (min)	-7.24	6.9	-60	-1	-7.77	-6.42
ΔC (DKK)	6.94	11.2	0.5	200	7.11	6.68

Table 3.2: DESCRIPTIVE STATISTICS, $0-1$ INDICATORS, SHARE$=1$ IN PERCENT

variable	All	PT	Car	description
x_{area}	37.9	47.8	22.6	residence in Copenhagen
x_{area2}	34.2	24.5	49.5	residence in small town (<10000)
x_{carno}	21.5	33.6	2.7	no cars in household
x_{cars}	16.5	8.3	29.4	more than one car in household
x_{carsin}	8.8	5.3	14.3	one car in single adult household
x_{child}	10.7	10.7	10.7	child in household
x_{grp2}	6.0	2.0	12.2	travel with family
x_{grp3}	5.9	4.3	8.4	travel with non-family
x_{hinc}	32.2	33.7	29.8	household income $>$ 600.000 DKK
x_{noinc}	5.5	5.0	6.3	income unknown
x_{lic}	91.0	85.5	99.4	holding a drivers licence
x_{lug}	11.7	10.0	14.3	travel with large luggage
x_{female}	53.4	49.8	58.9	female
x_{tripf}	41.9	53.2	24.3	travel less often than daily
$x_{weekend}$	4.7	2.8	7.6	travel on weekend
x_{workh}	80.8	81.9	79.3	work home less than once a week
x_{occup}	98.1	97.2	99.6	wage earner or self employed

this dominated alternative. Since we could not be sure that these people understood the SP task they were taken out of the sample. Of the remaining 1248 individuals 3 had unrealistic reported travel times, 5 had unrealistically large travel costs, 24 had unrealistic travel speeds and 1 did not complete all of his choices. This left 1216 individuals of which 739 used PT and 477 car.

We exclude the dominant check question from the estimations since the information given by this choice is uninformative in the framework of DeSerpa (1971) that implies nonnegative VTT. To estimate the model the alternatives have been arranged such that alternative 1 is the fastest. With this rearrangement the differences between the attributes of the SP choices can be seen in Table 3.1, where ΔT denotes the time attribute of the first alternative minus the time attribute of the second, etc.

Table 3.2 summarises the $0-1$ dummies used as explanatory variables.[2] The reason why a few car users have no cars in the household is that they are car passengers. Descriptions for the continuous variables are shown in Table 3.3. Here x_{age} is the age, x_{dis} is the log of the distance

[2]One Euro is 7.5 DKK (301007).

Table 3.3: MORE DESCRIPTIVE STATISTICS

variable	mean	std.dev.	min	max	mean(pt)	mean(car)
x_{age}	42.8	11.2	16	73	41.4	45.0
x_{dis}	3.02	1.13	0	6.40	2.98	3.08
x_{inc}	1.26	0.51	0	2.4	1.25	1.27
x_{time}	3.22	0.80	1.10	5.86	3.25	3.17
x_{cong}	0.10	0.13	0	0.49	—	0.10

in kilometres between origin and destination. If distance was zero, x_{dis} is set to zero. The variable x_{inc} is the log of gross personal income for the people with reported income.[3] The variable x_{time} is the log of reported travel time i minutes. Each alternative in the SP choices concerning car also included the attribute congested time. Since this attribute in all SP choices had a fixed ratio to the total time depending on reported congestion we choose to use only total travel time and include the congestion ratio as an explanatory variable.[4] This approach was also used in Fosgerau (2006). It is worth noting that car users are older and travel longer distances in shorter time but income is the same in the two segments. The fact that income is similar in the two segments is somewhat unusual. In this dataset the explanation could be that people working in central Copenhagen have higher income in general together with better service by PT.

3.3.2 Instruments

Now we will return to the question of instruments introduced in Section 3.2.1. In the selection equation we have a vector of explanatory variables. These are divided into two groups: the variables that also enter the VTT equation and the instruments. The first choice to make is which explanatory variables to include in both equations. These should be variables with a causal effect on VTT. The effect of other variables on VTT is only included indirectly through the mode choice. As explanatory variables in the VTT equation we choose to include income and time, since it is restrictions on these two resources that cause the VTT to exist. Furthermore we choose to include age and sex since the causality between these variables and VTT is clear.[5] The inclusion of these two variables has the same purpose as segmentation.

For the remaining variables we do not have theory to support their inclusion in the VTT equation. Some of them, like car ownership, are clearly correlated with VTT but since it is very probable that higher VTT leads to higher car ownership it would violate the assumptions of the model to enter car ownership in the VTT equation. The remaining variables are therefore used as instruments in the selection equation. Since these variables are not collected for a mode choice model but as background variables in a SP experiment special care has to be taken. They must be independent of the chosen mode. A variable indicating if working on the reference trip is an example of a variable that cannot be used based on these criteria.

[3]Income is not continuous, but discrete with 11 levels where level 1 is income below 100,000 DKK, level 2 is income between 100,000 DKK and 200,000 DKK, etc., until level 11 which is income above 1,000,000 DKK.

[4]The statistics for congestion is only calculated over car users.

[5]The VTT does not affect age and sex.

Table 3.4: MODEL FITS

	Car	Car	Car	PT	PT	PT
	Model I	Model II	Model III	Model I	Model II	Model III
seq selection	-509.6	-509.6	-509.6	-509.6	-509.6	-509.6
$\overline{\rho}^2$	0.373	0.373	0.373	0.373	0.373	0.373
seq VTT eq.	-2110.9	-2028.3	-2040.4	-2821.7	-2712.7	-2689.5
$\overline{\rho}^2$	0.198	0.229	0.225	0.309	0.336	0.342
simultaneous	-2619.7	-2536.6	-2550.0	-3330.8	-3221.7	-3198.9
$\overline{\rho}^2$	0.240	0.264	0.261	0.320	0.342	0.347

The instruments used are area, number of cars, travel group, license, luggage, trip frequency, weekend, work home and occupation. An ideal instrument would be a variable having a large influence on selection of mode but being uncorrelated with VTT. None of the instruments above are obvious candidates. The number of cars is very likely to affect mode selection but it is doubtful if it is uncorrelated with VTT. Luggage and licence seem like the best candidates but they have low variation in the population.

3.3.3 Estimation results

Estimation was performed using a program written in Ox (Doornik 2001). The program used simulated maximum likelihood, see Train (2003), and Halton draws where the first 20 periods were removed from each series. The final results were based on 1500 Halton draws.

For each of the 3 specifications discussed in Section 3.2.2 two models were estimated - one with correlation between the selection equation and the VTT equation and one without correlation. Since each model is estimated for both car and PT users this gives a total of 12 models. The model fits are summarised in Table 3.4. In Table 3.4, Model I refers to the specification in preference space, Model II refers to the specification in inverse VTT space, and Model III refers to the specification in log VTT space. Now we will comment on the estimations for each of the 3 specifications. The selection models are based on the following specification:

$$U_{1n} = \gamma_0 + \gamma_{age}x_{age,n} + \cdots + \gamma_{workh}x_{workh,n} + u_{1n},$$

where $u_{1n} \sim N(0,1)$. The results for the sequential selection model are in Table 3.5. The results for these coefficients in the simultaneous models are similar so they are not reported. The VTT model estimates are given in Table 3.6.

Model I estimation results

Model I without correlation consists of two independent models: a probit model describing selection and a mixed logit model for the SP experiment with specification

$$V_{nt} = e^{-\beta_0-\beta_{age}x_{age}-\cdots-\beta_{cong}x_{cong}+s_1v_1+s_2v_2}\Delta C_{nt} + e^{\beta_T+s_3v_3}\Delta T_{nt}.$$

Table 3.5: ESTIMATION RESULT FOR THE SELECTION EQUATION

	estimate	t test
γ_0	-3.11	-4.27
γ_{age}	0.12	2.77
γ_{area}	-0.40	-3.42
γ_{area2}	0.33	2.89
γ_{carno}	-1.42	-8.15
γ_{cars}	0.87	7.02
γ_{carsin}	0.54	3.54
γ_{child}	-0.27	-1.93
γ_{dis}	-0.15	-3.69
γ_{grp2}	1.53	7.21
γ_{grp3}	0.59	3.10
γ_{hinc}	-0.24	-2.30
γ_{lic}	1.70	4.60
γ_{lug}	0.44	3.05
γ_{female}	-0.32	-3.43
γ_{tripf}	-1.08	-10.96
$\gamma_{weekend}$	0.66	2.77
γ_{occup}	1.88	3.14
γ_{workh}	-0.28	-2.32

Table 3.6: ESTIMATION RESULTS FOR THE VTT EQUATIONS

Model I								
	Car		PT		Car		PT	
	estimate	t test	estimate	t test	estimate	t test	estimate	t test
β_T	-1.42	-15.73	-1.01	-18.45	-1.43	-16.26	-1.01	-18.28
β_0	-0.61	-1.46	-1.80	-6.34	-0.57	-1.40	-1.84	-7.14
β_{age}	-0.23	-3.86	-0.12	-2.61	-0.23	-3.96	-0.09	-2.14
β_{inc}	0.24	1.29	0.69	5.70	0.22	1.32	0.67	5.58
β_{ninc}	1.03	2.90	1.02	3.88	1.12	3.35	1.03	4.12
β_{female}	-0.29	-2.19	0.06	0.60	-0.34	-2.72	0.04	0.46
β_{time}	0.71	8.11	0.51	9.18	0.67	9.03	0.52	10.27
β_{cong}	1.07	2.21			1.07	2.31		
s_1					0.23	2.00	0.17	1.21
s_2	1.00	14.70	0.98	20.40	0.98	14.68	0.96	19.70
s_3	0.91	10.86	0.66	13.83	0.94	13.30	0.68	13.59
Model II								
	Car		PT		Car		PT	
	estimate	t test	estimate	t test	estimate	t test	estimate	t test
β_T	-1.21	-10.41	-0.79	-10.12	-1.19	-10.41	-0.79	-10.14
β_0	-1.45	-5.50	-2.37	-15.16	-1.34	-6.03	-2.35	-14.04
β_{age}	-0.27	-7.50	-0.15	-4.59	-0.27	-7.73	-0.15	-4.47
β_{inc}	0.43	4.44	0.84	10.70	0.41	5.40	0.85	10.07
β_{ninc}	0.94	4.60	1.24	7.12	0.94	5.41	1.32	8.01
β_{female}	-0.25	-3.72	0.04	0.49	-0.20	-2.54	0.00	0.02
β_{time}	0.50	16.39	0.36	7.53	0.46	9.12	0.38	8.44
β_{cong}	1.82	7.21			1.71	4.99		
s_1					0.10	3.13	0.12	4.10
s_2	1.06	18.09	1.01	22.58	1.07	21.56	1.01	30.06
s_3	1.66	9.27	1.23	11.34	1.65	9.71	1.23	11.11
Model III								
	Car		PT		Car		PT	
	estimate	t test	estimate	t test	estimate	t test	estimate	t test
β_0	-1.58	-3.67	-2.32	-8.66	-1.61	-3.68	-2.33	-8.67
β_{age}	-0.31	-5.03	-0.16	-3.82	-0.31	-4.95	-0.16	-3.59
β_{inc}	0.55	2.97	0.88	7.64	0.54	2.89	0.88	7.66
β_{ninc}	0.87	2.40	1.34	5.43	0.86	2.39	1.36	5.44
β_{female}	-0.20	-1.49	0.05	0.57	-0.22	-1.58	0.05	0.54
β_{time}	0.46	5.26	0.29	5.24	0.46	5.20	0.29	5.26
β_{cong}	2.14	4.21			2.13	4.20		
s_1					0.08	0.49	0.07	0.60
s_2	1.18	15.67	1.02	23.22	1.18	15.59	1.02	22.97
m	1.18	20.28	1.88	29.44	1.18	20.28	1.88	29.45

The dependent variable is choosing car for both segments in the probit model. Remember that the selection only uses socioeconomic variables so the parameters are differences in the effect of a variable on utility of car and PT.

The estimation of the probit is based on all 1216 individuals. All estimates are significant at the 5 % level except for γ_{child} which comes very close. One sees that γ_{age}, γ_{area2}, γ_{cars}, γ_{carsin}, γ_{grp2}, γ_{grp3}, γ_{lic}, γ_{lug}, $\gamma_{weekend}$, and γ_{occup} all are positive. This is expected for γ_{age}, γ_{area2}, γ_{cars}, γ_{carsin}, γ_{lic}, and γ_{lug}. For the others it is less obvious but also reasonable, e.g., the effect of γ_{occup} is possibly derived from higher income and car ownership. The other variables have negative signs. This is expected for γ_{area}, γ_{carno}, and γ_{female}. Of the remaining the most surprising is that γ_{hinc} has a negative sign. A possible explanation is that high-income individuals tend to work in the Copenhagen area where PT service is better.

The VTT estimation is based on the equation above. The choice of parameterising the cost coefficient is based on initial estimation suggesting higher significance of the explanatory variables and better fit. Furthermore, estimation without explanatory variables showed larger variance in the cost coefficient than in the time coefficient. From the estimation of the car VTT equation we get the expected signs on the parameters, i.e., VTT rises with income, time, and congestion, it is lower for females and declines with age. The most surprising result is the insignificant β_{inc}.[6] For the PT segment we get different sizes for the estimates but the same signs. Except for β_{female} all parameters are significant in the PT segment. This is more in line with what one would expect.

The simultaneous model consists of the selection and VTT equation connected through correlation as described in Section 3.2.3 otherwise the models are unchanged. The results for the car segment resemble the sequential model. The signs are the same and the same parameters are seen to be significant with the only exception being that γ_{child} is now significant. So β_{inc} is still insignificant. The correlation, s_1, is seen to be significant and positive. For the PT segment all signs are the same as for the sequential model and the only change is that γ_{child} becomes significant. The correlation is not significant but keeps the positive sign. The sign is positive for both segments. Since the dependent variable in the selection model is an indicator for car the positive sign is equivalent to positive correlation between higher VTT and choosing car for both segments.

There were problems with the convergence of the simultaneous models above. Using 500, 1000, 1500, and 2000 Halton draws, the t test on the correlations changed between 1 and 3. Therefore we will not put too much emphasis on this model in the discussion and conclusion later.

Model II estimation results

Model II uses the specification

$$V_{nt} \quad = \quad e^{-\beta_0-\beta_{age}x_{age}-\cdots-\beta_{cong}x_{cong}+s_1v_1+s_2v_2}e^{\beta_T+s_3v_3}\Delta C_{nt}+e^{\beta_T+s_3v_3}\Delta T_{nt}$$

[6]This could be due to low variation of the variable in the data but as the same coefficient is significant with the other specification it is probably just a sign of a poor specification.

in the VTT model. The model is specified in inverse VTT space instead of VTT space for the same reasons that the cost coefficient is parameterised in Model I.

The probit results for the sequential model are the same as for Model I. The VTT estimation for the car segment is seen to outperform Model I based on log-likelihood. This is an indication that Model II is better than Model I at describing the data. Since the models are non-nested we cannot perform a formal likelihood ratio test.[7] Concerning the signs we get the expected ones and all coefficients are significant. In general the estimates are more significant and it is very comforting that income is now significant.

For the PT segment we get similar results. Model II is again seen to clearly outperform Model I based on log-likelihood. All signs are the expected ones and except for β_{female} all estimates are more significant.

The two simultaneous models repeat the pattern from the sequential models. The parameters are all significant with expected signs for the car segment and for the PT segment only γ_{child} and β_{female} are insignificant. Both models are seen to outperform Model I. Now the correlation in both models is significant. This confirms that it is more appropriate to model correlation between the VTT and selection directly as in Model II. Again we see that the sign is positive for both segments with the same interpretation as before that VTT is higher conditional on an individual being a car user.

Model III estimation results

Model III uses the specification

$$V_{nt} \quad = \quad -ln(-\Delta C_{nt}/\Delta T_{nt}) + \beta_{age}x_{age} + \cdots + \beta_{cong}x_{cong} + s_1 v_1 + s_2 v_2$$

in the VTT model. For the results of the probit models the same comments apply that were mentioned for Model I. Looking at the VTT specification in Equation 3.4 the coefficient on the first variable is set to 1. Therefore it is possible to estimate the scale of the logistic error. The scale is reported as m in the results. The two sequential VTT estimations are seen to give the same signs as Model II. The only difference is that β_{female} is now insignificant for both segments and in general all coefficients are less significant. Compared to Model II the car model is seen to have a lower log-likelihood whereas the PT model has a higher one. In this comparison it should be noted that Model III has one mixed variable less. Again one has to remember that the models are non-nested.

The simultaneous models correspond to the sequential. In both segments the correlation is seen to be insignificant with a positive sign.

[7]Non-nested tests could be performed based on the $\overline{\rho}^2$ given in Table 3.4.

Table 3.7: VTT FOR AVERAGE INDIVIDUAL FROM MODEL I IN DKK PER MINUTE

Model/segment	Car		PT	
$E(VTT_s\|x)$	1.54	(0.21)	0.97	(0.07)
$E(VTT\|x)$	1.39	(0.19)	1.06	(0.10)
$E(VTT\|x,Y=1)$	1.79	(0.27)	1.30	(0.29)
$E(VTT\|x,Y=0)$	1.23	(0.22)	0.97	(0.07)

3.3.4 VTT estimation results

The mean VTT is simulated for each of the 3 models. The results are seen in Table 3.7-3.9 with standard deviation of the mean in parenthesis. All means are evaluated at the mean value of the explanatory variables, $\bar{x}$, in the sample, except for congestion that is set to zero to allow for comparison across modes. For all models we have evaluated the 4 different means described in Section 3.2.4. The first mean is the mean calculated from the model not taking selection into account, i.e., $E(VTT_s)$. The second mean is the population mean calculated from the model with selection. The third mean is calculated conditional on the average individual being a car user, i.e., it is the second mean conditional on $Y = 1$. This is possible since the model taking selection into account can condition on this choice. The fourth mean corresponds to the third mean but instead we condition on using PT.

Looking at Table 3.7 we see that the standard mean $E(VTT_s|x)$ in car is much higher than in PT. The pattern is similar for the model with selection. Again $E(VTT|x)$ is higher in car, but looking at the two means one sees that they are now closer. They have both moved toward one another. So part of the gap between VTT in car and PT has been explained by the selection.

The value of $E(VTT|x,Y = 1)$ for car should be compared with $E(VTT_s|x)$ for car since they represent the same mean for the two models. They are somewhat different, but the large standard deviations make it impossible to say anything about bias. In the same way $E(VTT|x,Y = 0)$ for PT should be compared to $E(VTT_s|x)$ in PT. They are seen to be very similar as both equal 0.97.

For Model II the remarks are similar for the PT segment. The model with selection has higher mean. Furthermore, the mean VTT_s is similar to mean VTT conditional on $Y = 0$. They are 0.83 and 0.84. For the car segment it is seen that the mean with and without selection are close. They are 1.07 and 1.08. This is somewhat surprising since the correlation is significant.

For Model III we see that the effects are small which corresponds to the fact that the correlation is insignificant. The patterns though are similar to the patterns for Model II while the standard deviations are higher.

We can draw one immediate conclusions: In the standard case, VTT_s, one would conclude that the VTT in car is higher than in PT. This conclusion is not so obvious when taking selection into account for the models with significant correlation.

Table 3.8: VTT FOR AVERAGE INDIVIDUAL FROM MODEL II IN DKK PER MINUTE

Model/segment	Car		PT	
$E(VTT_s \mid x)$	1.08	(0.07)	0.83	(0.04)
$E(VTT \mid x)$	1.07	(0.06)	0.89	(0.04)
$E(VTT \mid x, Y=1)$	1.21	(0.07)	1.02	(0.07)
$E(VTT \mid x, Y=0)$	1.02	(0.06)	0.84	(0.04)

Table 3.9: VTT FOR AVERAGE INDIVIDUAL FROM MODEL III IN DKK PER MINUTE

Model/segment	Car		PT	
$E(VTT_s \mid x)$	0.93	(0.11)	0.70	(0.04)
$E(VTT \mid x)$	0.90	(0.14)	0.72	(0.05)
$E(VTT \mid x, Y=1)$	0.97	(0.14)	0.79	(0.14)
$E(VTT \mid x, Y=0)$	0.87	(0.19)	0.69	(0.04)

3.3.5 Discussion

Three model specifications have been estimated. All three gave reasonable parameter estimates with the exception of the income parameter in Model I for car which turned out insignificant. Based on log-likelihood Model II and Model III both outperform Model I. Therefore the remaining discussion will concentrate on these two models. The correlation has the same sign in both models, and it is significant in Model II. The sign shows that there is a positive correlation between choosing car and having a higher VTT.

The main difference between the two model types is that the logistic error is multiplicative in Model III and additive in Model II relative to the VTT. This means that in Model III the logistic error affects the choice behaviour relatively depending on the VTT of the individual. For Model II the logistic error becomes dominant for small values of ΔT and ΔC and disappears for large values. A second difference is that the scale is lognormal in Model II whereas it is constant in Model III.

These model differences aside it is still sound to conclude based on the estimation that on our sample Model II and Model III should be preferred to Model I. Hence we are left to conclude that either we discard the standard structure with additive errors or we have to acknowledge the significant self-selection.

Of course, it is also possible to do both. Since both models depend on our instruments, we are left with the concern whether we have used appropriate instruments. As mentioned in Section 3.3.2 a good instrument must have a large impact on the selection of mode. From the estimates, car ownership, travel group, licence, trip frequency and occupation have a high impact on selection. This supports that we actually have good instruments. But since they do not result in stronger correlation we have to look critically at the selection model.[8] The assumption that the error term is normal is very restrictive. Furthermore it is very probable that car ownership and distance should be treated as endogenous. Hence a more flexible selection model could be worth pursuing. The above also highlights that besides careful design of the SP choices in future VTT experiments, it would be fruitful to design background questions in the questionnaire with the search for appropriate instruments in mind.

3.4 Summary and conclusions

This paper has presented a discrete choice model that investigates the effect of self-selection on VTT estimates through the unobserved heterogeneity in the VTT model. The model is an addition both to the literature on mixed logit models and on self-selection in the way it incorporates the self-selection into a mixed logit model.

The model has lead to reasonable VTT estimates with significant correlation in half of the models. This partial evidence of correlation should be seen as an indication that more work is needed. Moreover we would expect the possible effects of self-selection to be larger in a sample where the income differences are larger between modes. An important question is if we have used appropriate instruments and whether they can be found. This is a challenge for future SP questionnaire designs.

There are several ways to continue the present investigation. One immediate extension would be to estimate everything jointly for the car and PT segment as opposed to separately for each segment.

A second extension would be the methodological challenge to relax the assumption of joint normality of log VTT and the error in the selection equation. This has been done in the case of labour supply models, see e.g., Vella (1998).

Instead of extensions of the model a different approach to the central problem of the effect of self-selection on VTT estimates could be to carefully design an across-mode SP experiment. In such an experiment it would be possible to compare VTT for PT passengers and car users in car directly. The SP experiment has to ensure that both reference trips are carefully described to capture any mode bias in the sample not related to VTT.

The main conclusion from this paper is that the sample selection affects VTT estimates and that the effect can alter the final output from the model. A second conclusion is that self-selection should be looked at more seriously in the transportation field especially now that researches have moved away from the robust multinomial logit model. So more research is needed to investigate if it is just a sample and/or model-specific selection effect we have found.

[8] One consideration could be to introduce level-of-service variables. While this could improve the model, it would contradict the condition that the variables in the selection model should not be mode specific.

Acknowledgment
The authors would like to thank two anonymous referees for valuable comments.

References

Axhausen, K. W., König, A., Abay, G., Bates, J. J. & Bierlaire, M. (2004), *Swiss Value of Travel Time Savings*, paper presented at the European Transport Conference, Strasbourg. [28]

DeSerpa, A. (1971), 'A theory of the economics of time', *The Economic Journal* **81**, 828–846. [28, 35]

Doornik, J. (2001), *Ox: An Object-Oriented Matrix Language*, Timberlake Consultants Press, London. [37]

Eklof, J. & Karlsson, S. (1999), Testing and correcting for sample selection bias in discrete choice contingent valuation studies. SEE/EFI Working papers Series in Economics and Finance, No. 171. [29]

Fosgerau, M. (2006), 'Investigating the distribution of the value of travel time savings', *Transportation Research Part B: Methodological* **40**(8), 688–707. [31, 36]

Fosgerau, M. (2007), 'Using nonparametrics to specify a model to measure the value of travel time', *Transportation Research Part A* **41**, 842–856. [32]

Fosgerau, M., Hjorth, K. & Lyk-Jensen, S. V. (2006), *An integrated approach to the estimation of the value of travel time*, European Transport Conference, Strasbourg, 2006. [34]

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47(1)**, 153–162. [29]

Hensher, D. A. (2001), 'Measurement of the valuation of travel time savings', *Journal of Transport Economics and Policy* **35**(1), 71–98. [28]

Hensher, D. A. & Greene, W. H. (2003), 'The mixed logit model: The state of practice', *Transportation* **30**, 133–176. [34]

Manski, C. & Lerman, S. (1977), 'The estimation of choice probabilities from choice based samples', *Econometrica* **45(8)**, 1977–1988. [29]

Sillano, M. & Ortúzar, J. de D. (2004), 'Willingness-to-pay estimation from mixed logit models: some new evidence', *Environment & Planning A* **37**(3), 525–550. [28]

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY, USA. [28, 37]

Train, K. & Weeks, M. (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 1, pp. 1–16. [32]

Train, K. & Wilson, W. (2007), ‘Estimation on stated-preference experiments constructed from revealed-preference choices’, *Transportation Research B, forthcoming* . [28, 29]

Vella, F. (1992), ‘Simple tests for sample selection bias in censored and discrete choice models’, *Journal of Applied Econometrics* **7 (4)**, 413–421. [29]

Vella, F. (1998), ‘Estimating models with sample selection bias: A survey’, *The Journal of Human Resources* **33(1)**, 88–126. [44]

Wardman, M. (2004), ‘Public transport values of time’, *Transport policy* **11**, 363–377. [28]

Chapter 4

Estimation of correlated value of travel time in public transport

Stefan L. Mabit
Centre for Traffic and Transport, Technical University of Denmark

Otto A. Nielsen
Centre for Traffic and Transport, Technical University of Denmark

Abstract

Any journey by public transport has several aspects, e.g., access, egress, waiting, transfer and in-vehicle time in various modes. It is likely that willingness to pay for these aspects is somehow related. This paper investigates different correlation structures between the willingness to pay for the aspects. We apply mixed logit models to stated-preference data on public transport route choice to infer the willingness-to-pay indicators. We enlarge the models to examine how the correlation structures are affected by the inclusion of background variables in the willingness to pay. The allowance of nonstandard correlation patterns is seen to improve the models significantly, whilst it does not affect the evaluation of mean willingness to pay. The models including background variables confirm these results. This shows that even conditional on background variables it is important to allow for correlation in discrete choice models.

4.1 Introduction

Discrete choice analysis is one of the core building blocks of transport modelling. In recent years the mixed logit model has become a powerful tool in explaining discrete choices (Hensher & Greene 2003). An important application is the estimation of willingness to pay (WTP) for various goods. This could be WTP for travel time also known as the value of travel time (VTT) or other WTP, e.g., WTP for headway. The majority of mixed logit models assume that the distributions of the coefficients are independent or perfectly correlated. This assumption may not hold in models with various time components, e.g., access time, waiting time, and in-vehicle time distributed on various public transport modes in a transit network. For a recent review of issues related to WTP in public transport, see Wardman (2004).

Intuitively it seems plausible to expect the different WTP distributions to be correlated in a given population even when conditional on background variables. This would be the case if individuals with tighter time schedules have high unexplained WTP for different time aspects of travel with public transport. A second example is if individuals have high WTP for less walking time and seat availability in a way that cannot be explained by background variables. The intuition is supported by theory on VTT as explained in Section 4.2.2. Therefore it is important to investigate whether there is correlation and if this affects the evaluation of WTP.

In this paper we impose three correlation structures on WTP aspects relevant to travel by public transport. The first is the standard structure with independent coefficients. This model does capture some correlation as the variation in the cost coefficient induces correlation among WTPs. The second structure is motivated by the fact that the resource value of time (see Section 4.2.2) affects all aspects of time, e.g., in the case of travel by public transport it is both part of in-vehicle time, access time, egress time, and waiting time. The third structure generalises the second structure in that it allows for correlation between the cost and the time aspect coefficients. This could be induced by a random scale often found in panel data.

The data used for the analysis consist of 6 unlabelled stated-preference (SP) experiments for the public transport modes regional train, city train, and bus covering the Copenhagen region (Nielsen et al. 2001). From the data, the WTP for in-vehicle time, access and egress time, waiting time, and other components of a travel with public transport are estimated. This is done with mixed logit models, using various lognormal distributions on the coefficients (dependent on the specific model being tested).

The paper confirms that correlation structures cannot be neglected and in the cases where a statistical test is possible the correlation is seen to be significant. The models including background variables confirm the findings concerning correlations. The model includes background variables both in coefficients and in the scale. This inclusion raises a second issue, namely, the problem of heteroscedasticity. The results show that what could be attributed to deterministic heterogeneity to some degree disappears when allowing for heteroscedasticity in the models.

For all models the mean WTP is calculated. This is mainly done as a validation check. Correlation is seen not to affect mean VTT. It is important to note that even if correlation does not affect the mean VTT, it may still have great effect on forecasting or assignment.

The paper is organised as follows. Section 4.2 describes the theoretical background on discrete choice models and WTP estimation. Section 4.3 presents the data, the model specifications examined, and the estimation results. The final Section 4.4 discusses the results and presents the main conclusions.

4.2 Theoretical background

4.2.1 Random utility theory and mixed logit models

In transportation research the most common way to estimate VTT and other WTP has been through the use of random utility maximising (RUM) models; in recent years especially mixed logit models.

A discrete choice between a finite set of alternatives, $i \in C$, is described by RUM models as maximisation over latent utilities.[1] This paper adopts an additive form of the utilities, i.e.,

(4.1) $$U_i = V_i + \varepsilon_i,$$

where ε_i is a stochastic part independent of V_i and V_i represents a conditional indirect utility function (CIUF) depending on attributes of alternative i together with background information on the individual making the choice. For the linear-in-parameters mixed logit model, Equation 4.1 becomes

(4.2) $$U_i = \beta' x_i + \varepsilon_i,$$

where ε_i is standard Gumbel distributed independently over alternatives and β is a vector of taste parameters following specific distributions. If all β's follow a fixed distribution the mixed logit model becomes a multinomial logit (MNL) model. See Train (2003) for general references on mixed logit models. The mixed logit model is good at describing panel data. In panel data, each individual makes several choices. This is represented in the model through common taste parameters for each individual over different choice situations.

Suppose we have a route choice between public transport alternatives. Then a simple specification could be

(4.3) $$U_i = \beta_c c_i + \beta_{inv} inv_i + \beta_{wait} wait_i + \varepsilon_i,$$

where c_i is the cost, inv_i is the in-vehicle time, $wait_i$ is the waiting time for alternative i, and β_c, β_{inv}, β_{wait} could be assumed to follow independent lognormal distributions. A parameter β is lognormal if $ln(\beta) \sim N(\mu, \sigma^2)$. For such a variable we use the notation $\beta = exp(\mu + \sigma * u)$, where u is standard normal.

In general, data and theory together should decide which distribution to assume. An application of mixed logit models with normal and lognormal coefficients is seen in Revelt & Train (1998). Both distributions have been criticised. The normal distribution is unrealistic because it allows for negative VTT in the left tail and also very high VTT in the right tail. The lognormal avoids the problem with negative VTT but it seems to magnify the problem with very high VTT. This has caused the mean VTT to become very high under lognormal assumptions in some applications.

[1] In this section we leave out of notation that the utility depends on individual n.

For both distributions it is possible to incorporate correlation between coefficients. This is rarely applied in practice mainly due to longer run time of estimation software and problems with convergence.

Both MNL and mixed logit can be nested to include different datasets. For the purpose of estimating based on different data sources Equation 4.2 becomes

$$U_i = s_k \beta' x_i + \varepsilon_i, \tag{4.4}$$

where k denotes the different datasets and s_k is a dataset-specific positive scale.

4.2.2 Theory on the value of travel time

The microeconomic framework of DeSerpa (1971) and others (see Jara-Diaz 1998) views VTT as the sum of the resource value of time and the direct disutility of travel. In a typical discrete choice model, it is not possible to estimate these two elements separately. The resource value of time is part of the value of all time aspects, e.g., in-vehicle time and waiting time. So unless a model explains this common part through the use of background variables it could induce a correlation structure between the different time aspects. It also seems plausible that the WTP for shorter headway and waiting time could be correlated in ways hard to capture through background variables. An important implication of the framework is that VTT is non-negative.

The general expression for VTT given a discrete choice model including time and cost as attributes is:

$$VTT = \frac{\partial V}{\partial T} / \frac{\partial V}{\partial C}, \tag{4.5}$$

where V corresponds to V_i in Equation 4.1, see Bates (1987).

One can deduce this fraction from the marginal utilities with respect to cost and time. For the model in Equation 4.3 one gets for in-vehicle time:

$$VTT(\beta) = \beta_{inv} / \beta_c. \tag{4.6}$$

In mixed logit models with lognormal distributions on the coefficients we get:

$$VTT(\beta) = exp(\beta^0_{inv} - \beta^0_c + \sigma_{inv} * u_1 - \sigma_c * u_2), \tag{4.7}$$

where β^0_{inv}, β^0_c, σ_{inv}, and σ_c are the parameters of the lognormal distributions, e.g., $\beta_{inv} = exp(\beta^0_{inv} + \sigma_{inv} * u_1)$ with $u_1 \sim N(0,1)$. In this case, VTT also follows a lognormal distribution.

Estimation using maximum likelihood gives the asymptotic simultaneous normal distribution that the parameters follow. Hensher & Greene (2003) recommend to use this asymptotic distribution to evaluate the estimation results. Below the procedure is exemplified for in-vehicle time.

Table 4.1: OBSERVATIONS DIVIDED OVER DATASETS

Dataset / purpose	Commute	Other
rb1 (reg. and bus)	381	671
rb2 (reg. and bus)	386	680
b1 (bus)	126	299
b2 (bus)	124	296
s1 (s-train)	735	377
s2 (s-train)	731	378

From the simultaneous distribution one gets a draw β^r including β^r_{inv} and β^r_c. For the VTT in Equation 4.7 this gives

$$VTT(\beta_r) = exp(\beta^r_{inv} - \beta^r_c + \sigma^r_{inv} * u_1 - \sigma^r_c * u_2). \tag{4.8}$$

This is repeated many times and $E(VTT) = \mu_{inv}$ is calculated as the mean over these repetitions.

Above we have illustrated the simulation for lognormal distributions on the coefficients. The reason for not allowing a normal distribution on the coefficients is for the cost coefficient that it would conflict with the assumption that V represents a CIUF. A normal distribution would imply that some individuals have positive cost coefficients. This is in conflict with the fact that utility should decrease with rising cost.

4.3 Data, specification, and estimation

4.3.1 Data

Data originate from a survey done for the Copenhagen-Ringsted Model dating from 1998, see Nielsen et al. (2001). We used part of these data coming from 6 surveys: rb1 and rb2 (regional train and bus), b1 and b2 (bus), s1 and s2 (s-train; the city train in Copenhagen). They all contain unlabelled binary SP choices between common modes. The rb1 and rb2 sets have choices between regional trains or buses, the s1 and s2 have choices between s-trains, and the b1 and b2 have choices between buses. The data consist of 5,287 observations. It is an unbalanced panel where each individual makes between 8 and 16 choices.

Some observations were removed resulting in 5,184 observations used for estimation. The observations were removed because of no choice or missing variables. The remaining observations were divided as seen in Table 4.1 between purposes and datasets. The attributes included in the different datasets are the following: cost, in-vehicle time (inv), access/egress (aux), delay (del), headway (head), number of interchanges (int), seat, wait/interchange time (wtint). Table 4.2 gives the average for the attributes in each dataset together with the range in parenthesis. The time unit is minutes and the cost unit is DKK.[2] Transfers are measured in number of changes and seat does not have a unit since it is a dummy variable.

[2] One Euro is 7.5 DKK (301007).

Table 4.2: MEAN ATTRIBUTES WITH RANGE IN PARENTHESIS FOR EACH DATASET

	Cost (DKK)	Inv (min)	Aux (min)	Delay (min)	Head (min)	Wtint (min)	Int (scale)	Seat (scale)
rb1	32 (1-142)	61 (6-332)	18 (6-60)	n.a.	54 (5-175)	15 (6-56)	n.a.	n.a.
rb2	33 (1-156)	58 (6-280)	n.a.	7 (1-140)	n.a.	n.a.	1.2 (0-5)	0.34 (0-1)
b1	6 (0-26)	23 (4-66)	10 (2-20)	n.a.	n.a.	10 (4-104)	n.a.	n.a.
b2	6 (0-26)	40 (6-146)	n.a.	8 (0-49)	n.a.	n.a.	n.a.	0.34 (0-1)
s1	10 (1-50)	24 (3-100)	19 (4-122)	n.a.	n.a.	9 (4-43)	n.a.	n.a.
s2	11 (1-50)	49 (13-166)	n.a.	9 (0-44)	n.a.	n.a.	n.a.	0.35 (0-1)

Table 4.3: BACKGROUND VARIABLES

Mode Variable	N	Age (scale)	Inc (scale)	Time (min)
Bus	1627 (104)	2.5	2.5	37
Re train	1336 (85)	2.2	3.2	65.1
S train	2221 (142)	2.4	2.8	24.2

The data also include the reference travel time (time) for an original journey taken in real life by each respondent. Other background variables available are age and income (inc). Age is a discrete variable with values from 1 to 4.[3] Inc is a discrete variable from 1 to 6.[4] Time is measured in minutes. The mean of the background variables for each mode is given in Table 4.3.[5] Here N is the number of observations with the number of individuals in parenthesis. The highest difference for the three modes is with respect to time. Regional train passengers are seen to have a much higher time and slightly higher income.

4.3.2 Specification

We present two MNL models and six mixed logit models. These are motivated by the theory discussed in Section 4.2 and preliminary estimation. Even though the MNL models are actually nested to deal with different data sources and the mixed logit models likewise they will be referred to as MNL and mixed logit models in the following. The nesting corresponds to scaling of each dataset, only. So in reality we just use standard mixed logit models. The eight models

[3]Here $age = 1$ if $age \leq 25$, $age = 2$ if $25 < age \leq 40$, $age = 3$ if $40 < age \leq 65$, and , $age = 4$ if $65 < age$.

[4]Here $income = k$ denotes that income is between $k-1$ and k times 100,000 DKK except for $income = 6$ which denotes income above 500,000 DKK.

[5]The average of income is only calculated over the persons reporting income (3.5 % have missing income).

Table 4.4: MODELS

	Attributes only	With background variables
MNL	MNLa	MNLb
Mixed logit	MixIa	MixIb
Mixed logit w. correlated time	MixIIa	MixIIb
Mixed logit w. correlated time and cost	MixIIIa	MixIIIb

reported are seen in Table 4.4. The first model type is a MNL model. The model is used to check if the parameter estimates have the expected signs and reasonable sizes. Furthermore, the MNL model can be used as a benchmark for the mixed models. The remaining three model types are all mixed logit models that have the MNL model as special case (if the random coefficients are not significant). These are estimated to compare the effect of differing correlation structures; especially to compare models allowing for correlation with the standard model having independently distributed parameters. The ”a models” include no background variables and the ”b models” include time, income, and age. In the following we describe the models in more detail.

MNL models: This is a MNL model with linear specification and scales to deal with different data sources. We assume that utility of alternative $i = 1,2$ for individual $n = 1,\ldots,N$ in choice situation $t = 1,\ldots,T_n$ is given by:

$$U_{itn} = V_{itn}s_n + \varepsilon_{itn}, \tag{4.9}$$

where ε_{itn} are independently standard Gumbel distributed, s_n is a scale parameter depending on data source and reference time for the individual traveller, and

$$\begin{aligned} V_{itn} &= \beta_c^n c_{itn} + \beta_{inv} inv_{itn} \\ &+ \beta_{aux} aux_{itn} + \beta_{del} del_{itn} \\ &+ \beta_{head} head_{itn} + \beta_{wtint} wtint_{itn} \\ &+ \beta_{int}^{nm} int_{itn} + \beta_{seat}^{nmp} seat_{itn}, \end{aligned}$$

where c_{itn} is the cost of alternative i at time t for individual n

inv_{itn} is the in-vehicle time of alternative i at time t for individual n

aux_{itn} is the access and egress time at time t for individual n

del_{itn} is the delay at the destination at time t for individual n

$head_{itn}$ is the headway at time t for individual n

$wtint_{itn}$ is the wait and interchange time at time t for individual n

int_{itn} is the number of interchanges at time t for individual n

$seat_{itn}$ is an indicator for not having a seat at time t for individual n

$\beta_c^n = -exp(\beta_c^0 + \beta_{inc} ln(inc_n) + \beta_{age} 1(age_n > 2))$ is the cost coefficient

$\beta_{inv} = -exp(\beta_{inv}^0)$ is the coefficient on in-vehicle time

$\beta_{aux}, \beta_{del}, \beta_{head}, \beta_{wtint}$ are coefficients using a similar parameterisation as β_{inv}

$\beta_{int}^{nm} = -exp(\beta_{int}^0 + \beta_{intr} * 1(m = re) + \beta_{timei} ln(time_n))$ and

$\beta_{seat}^{nmp} = -exp(\beta_{seat}^0 + \beta_{seatr} * 1(m = re) + \beta_{times} ln(time_n) + \beta_{purps} purp_n)$

Above $1(m = re)$ is an indicator function for mode equal to regional train and $purp_n$ is a dummy for non-commuters. The reason why we parameterise the coefficients in the above way is that it makes the parameter estimates useable as starting values in the mixed estimation. In model MNLa, the coefficients on inc, age, timei and times are fixed to zero.

MixI is a panel mixed logit model, where we assume that all coefficients on cost and times follow independent lognormal distributions. Therefore we have

$$
\begin{aligned}
V_{itn} &= \beta_c^n c_{itn} + \beta_{inv}^n inv_{itn} \\
&+ \beta_{aux}^n aux_{itn} + \beta_{del}^n del_{itn} \\
&+ \beta_{head}^n head_{itn} + \beta_{wtint}^n wtint_{itn} \\
&+ \beta_{int}^{nm} int_{itn} + \beta_{seat}^{nm} seat_{itn},
\end{aligned}
$$

where $\beta_c^n = -exp(\beta_c^0 + \sigma_c u + \beta_{inc} ln(inc_n) + \beta_{age} 1(age_n > 2))$

is a lognormal cost coefficient

$\beta_{inv}^n = -exp(\beta_{inv}^0 + \sigma_{inv} u_n)$ is a lognormal coefficient on in-vehicle time

$\beta_{aux}^n, \beta_{del}^n, \beta_{head}^n, \beta_{wtint}^n$ are lognormal coefficients similar to β_{inv}^n

$\beta_{int}^{nm}, \beta_{seat}^{nm}$ are fixed as in the MNL model

Again model MixIa has coefficients on background variables restricted to zero. The coefficients are chosen to be lognormally distributed to fulfil the requirement that VTT must be non-negative. The reason why the last two coefficients are kept fixed is that mixing all coefficients even though possible in theory, can cause convergence and identification problems in estimation. The two coefficients were chosen based on the results of testing of various alternative model formulations.

MixII is a second mixed logit specification. It incorporates the theoretically motivated connection between the different aspects of VTT. We choose to investigate a multiplicative structure. To allow for correlation, the time aspect with the smallest mean value is taken as common and the remaining time aspects are estimated relative to this aspect. Based on the MNL estimation and the estimation of MixI, it was found to be the coefficient on headway that should be taken as common.[6] The specification becomes:

$$
\begin{aligned}
&(4.10) \qquad V_{itn} &=& \beta_c^n c_{itn} + \beta_{head}^n \cdot (\beta_{inv}^n inv_{itn} \\
&(4.11) \qquad &+& \beta_{aux}^n aux_{itn} + \beta_{del}^n del_{itn} \\
&(4.12) \qquad &+& head_{itn} + \beta_{wtint}^n wtint_{itn}) \\
&(4.13) \qquad &+& \beta_{int}^{nm} int_{itn} + \beta_{seat}^{nm} seat_{itn},
\end{aligned}
$$

where the coefficients are the same as in MixI. Again model MixIIa has coefficients on background variables restricted to zero. Because of the fact that a product of lognormal distributions is again lognormal the various marginal utilities follow lognormal distributions.

MixIII is the same as MixII except that we enlarge the correlation structure. This is done by allowing the headway coefficient and the cost coefficient to be correlated. The specification is the same as for MixII except that the correlation between the cost and headway coefficient is unrestricted, i.e., $corr(\beta_c^n, \beta_{head}^n = \rho)$. MixII is then the special case $\rho = 0$.

The scale has the following specification:

$$(4.14) \qquad s_n = d^{k(n)} exp(\beta_t time_n + \beta_{purp} purp_n)$$

where $d^{k(n)}$[7] is a scale for each dataset (one is set to 1 for identification), β_t shows the effect of time on the scale and β_{purp} shows the effect of having a non-commute purpose on scale. The reason for keeping non-commuters in the sample is that mixed logit estimation requires a large sample.

4.3.3 Estimation

Here we discuss the estimation of the eight models mentioned in Section 4.3.2. The models were estimated using Biogeme (Bierlaire 2005). The model fits are reported in Table 4.5. Complete estimation results are given in Tables 4.7 and 4.8 in the appendix. A limit to the present estimation is the size of the mode-specific datasets.

MNL

The MNLa estimation was based on the attributes only. This estimation gave significant estimates for all coefficients. The MNL model also gave expected (negative) signs on the coefficients. As a check on data, alternative specific constants were also estimated. For unlabelled

[6] Since the mean and variance are dependent for lognormal distributions this procedure could be problematic if headway does not have the lowest variance.

[7] $k(n)$ denotes that the dataset is unique for each individual.

Table 4.5: MODEL FITS

	DoF	Log-likelihood		DoF	Log-likelihood
MNLa	17	-2784.6	MNLb	21	-2713.7
MixIa	23	-2551.5	MixIb	27	-2535.2
MixIIa	23	-2536.3	MixIIb	26	-2517.9
MixIIIa	24	-2524.9	MixIIIb	27	-2511.2

SP data these are expected to be insignificant because they signify unexplained preference for one alternative. Since the alternatives are unlabelled there should be no unexplained preference. This was confirmed as the alternative specific constants were indeed estimated to be insignificant. Hence, they were left out of the final models. For the continuous attributes a linear specification was kept. The coefficient on number of transfers was tested to be linear and significantly higher for regional train. For the seat variable there was also a significant difference between regional train and the other modes. A seat was valued higher for regional train.

The coefficients parameterising the scale are both significant. These estimates show that the variance of the error is larger for non-commuters and individuals with higher travel time.

The MNLb model included ln(time), ln(inc), and an age dummy as background variables. In the final specification, time entered in the seat and interchange coefficients. It is significant for seat while insignificant for interchange. The coefficient for income is significant with the expected sign and age is also significant. The group with unknown income was tested to have the same coefficients as the lowest income group. The model is seen to outperform MNLa significantly. The inclusion of the two time variables made the dependence of the seat and interchange coefficients on mode insignificant.

Mixed Models

Next we estimated the MixIa model. The model was estimated using MLHS draws since Halton draws are correlated for higher dimensions which could affect the investigation of correlation. Five variances on the mixed variables are significant the exception is σ_{wtint}. The model is seen to improve the MNLa model with more then 224 log-likelihood units. This clearly shows that MixIa outperforms the MNL specification.

The second mixed model, MixIIa, uses the specification in Equation 4.10. We now have that all variances are significant except for σ_{del}. As would be expected the model lowers the variance on the variables other than headway. The log-likelihood is higher than the one for MixIa. Therefore this model also outperforms the MNL model.

Since MixIa and MixIIa are non-nested the standard likelihood ratio (LR) test cannot be used in comparison between them. But as described by Horowitz (1983) the adjusted ρ^2 can be used as a good indication. He points out that a correct model very seldom has the lowest adjusted ρ^2. Based on this Ben-Akiva & Lerman (1985) present a test for non-nested models. Assuming that MixIa is the true model the probability that MixIIa has an adjusted ρ^2 which is higher by 0.004225 is less then 10^{-7}. Hence this test rejects MixIa based on the fit of MixIIa.

MixIIIa allows for correlation between the cost and the headway variable based on MixIIa. The correlation between headway and cost is significant since MixIIIa reject MixIIa in a LR test with 1 degree of freedom. From the results in Table 4.7 we get that $corr(ln(\beta_c^n), ln(\beta_{head}^n)) = \rho_{cost,head}\sigma_c = 0.38$. The sign of the correlation is surprising. In general, one would expect that individuals with higher marginal utility of income have lower marginal utility of time not the opposite. It could be the effect of a random scale.

We also tried a specification with fixed cost coefficient and the remaining coefficient following lognormal distributions. This model outperformed MNL, but performed much worse than the mixed models reported in the paper. Therefore this model was not investigated any further. That model corresponds to independently distributed WTP.

The models MixI-IIIb include background variables. The two time coefficients have the same overall influence as for the MNL models. The coefficients on income have the expected signs. Income is seen to be significant in all models whereas age become insignificant in the models with correlated coefficients. The three *b* models are seen to perform significantly better than the respective *a* models. The inclusion of background variables does not change any of the patterns concerning fit or correlation.

4.3.4 Willingness-to-pay estimates

In Table 4.6 the mean WTP is reported based on the simulation method described in Section 4.2.2. The WTP concerning time aspects are in DKK per hour. The value for transfers is per interchange and the value for seat is the penalty for not having a seat on the whole journey. The approach is repeated 1000 times and $E(VTT) = \mu_{inv}$ is calculated as the mean over these 1000 values. It is reported together with the standard deviation on the mean (Mean std), which is calculated as the square root of $1/1000\Sigma_k(\mu_{inv} - VTT(\beta^k))^2$. As a third statistic the ratio between the various means and the headway mean is reported (mean relative to headway). The reason for this is that if the cost does not affect route choice in assignment, then the relative size is more relevant than absolute size.

Looking at the WTP they have similar patterns for the four *a* models. The order is $WTP_{head} < VTT < WTP_{aux} < WTP_{wtint} < WTP_{del}$ as expected, except for the MNL model where VTT and WTP_{aux} are switched. The big difference is the size of the estimates. We see that the MNL estimates are smaller than the estimates from MixI-III which are similar. It is important to remember that the VTT in the mixed models are correlated because of the random cost coefficient and in the last two models also the headway coefficient. In the relative WTP, the correlated part drops out. Again the patterns are similar for the ratios between the variable mean and the headway mean. The models with background variables in WTP give similar results.

The fact that mean VTT is similar does not necessarily imply similar performances if the es-

Table 4.6: MEAN WTP EVALUATION IN DKK

X	MNLa E(X)	SD of E(X)	Rel. to head	MNLb E(X)	SD of E(X)	Rel. to head
Acc/Egr	26.76	2.48	2.69	29.48	2.82	2.80
Delay	53.79	4.51	5.41	55.56	3.96	5.28
Headway	9.95	1.96	1.00	10.53	2.22	1.00
Inv time	27.96	1.51	2.81	31.08	1.57	2.95
Wait/ch time	42.73	3.17	4.30	47.08	3.67	4.47
Interchanges	6.77	1.59	0.68	7.67	1.89	0.73
Seat	6.46	0.76	0.65	11.26	1.24	1.07

X	MixIa E(X)	SD of E(X)	Rel. to head	MixIb E(X)	SD of E(X)	Rel. to head
Acc/Egr	61.10	7.72	2.19	65.18	8.44	2.24
Delay	101.56	13.95	3.64	107.52	15.69	3.69
Headway	27.93	5.85	1.00	29.15	6.45	1.00
Inv time	49.02	5.64	1.76	52.92	6.68	1.82
Wait/ch time	79.45	10.85	2.84	84.90	11.55	2.91
Interchanges	9.05	2.02	0.32	12.70	3.06	0.44
Seat	11.63	1.66	0.42	15.15	2.43	0.52

X	MixIIa E(X)	SD of E(X)	Rel. to head	MixIIb E(X)	SD of E(X)	Rel. to head
Acc/Egr	65.68	8.75	2.76	72.74	11.52	2.73
Delay	94.22	12.70	3.96	104.66	18.55	3.93
Headway	23.78	4.05	1.00	26.62	5.62	1.00
Inv time	51.01	5.77	2.15	57.17	7.97	2.15
Wait/ch time	83.99	10.24	3.53	95.31	14.84	3.58
Interchanges	8.66	1.97	0.36	13.63	4.60	0.51
Seat	10.02	1.38	0.42	13.37	2.95	0.50

X	MixIIIa E(X)	SD of E(X)	Rel. to head	MixIIIb E(X)	SD of E(X)	Rel. to head
Acc/Egr	62.54	15.18	2.56	64.19	10.47	2.74
Delay	91.81	13.50	3.76	86.70	14.86	3.70
Headway	24.43	4.81	1.00	23.43	4.61	1.00
Inv time	48.04	5.74	1.97	49.77	6.66	2.12
Wait/ch time	83.59	13.52	3.42	82.91	12.37	3.54
Interchanges	9.32	2.62	0.38	10.72	3.88	0.46
Seat	11.50	2.10	0.47	10.96	2.38	0.47

timates were used in forecasting or assignment. This is similar to the fact that observed heterogeneity can affect policy analysis even if it does not change the mean estimates in a population. In a given application it is therefore necessary to capture correlation suitably. In a specific application there may not be enough data to estimate unrestricted correlation structures. One way to proceed is then to use theory as a guidance as we have done here.

4.4 Summary and concluding remarks

This paper investigates how theoretically motivated correlation structures affect the estimation of the value of travel time and other willingness-to-pay measures in public transport. We apply mixed logit models to stated-preference data that describe route choices for bus, s-train or regional train. The use of correlated coefficients is present in the literature but only little in transportation. We believe this paper to be the first to apply nonstandard correlation patterns in the context of public-transport data.

The estimation of the MNL and mixed logit models gives significant coefficients with expected signs. The mixed logit models outperform the MNL model as expected. The estimation of models with correlated coefficients shows that the standard independence assumption is unreasonable in the present investigation. Based on LR and non-nested hypothesis tests it is seen that the standard mixed model is rejected.

The evaluation of WTP is reasonable with the comment that WTP in general doubles from MNL to mixed logit models. The WTP from the mixed logit models have the general pattern $WTP_{head} < VTT < WTP_{aux} < WTP_{wtint} < WTP_{del}$. Relative to headway the values are seen to be similar for all models. This similarity in relative WTP does not imply that the models would perform equally in forecasting or policy analysis. This corresponds to the fact that even though MNL models might reproduce mean values nicely, forecasting on wrong homogeneity assumptions can be highly misleading.

The inclusion of background variables in the WTP does not affect the conclusions above. The model patterns are similar as are the evaluated WTP. This confirms the importance of allowing for correlation even if background variables are included in a model. Furthermore, the investigation shows that it is important to allow for heteroscedasticity. Otherwise the effect on scale could be mistaken for heterogeneity in WTP. In our application the time effect on VTT becomes insignificant when time is allowed to affect the scale.

The main conclusion of the article is that mixed logit models should allow for correlation instead of mechanically to use an independence assumption. It would be interesting to continue the study of correlation for each mode separately and with more background variables to see if correlation is independent of these.

Acknowledgment
The authors would like to thank Mogens Fosgerau and Ninette Pilegaard for valuable comments on an earlier draft.

References

Bates, J. (1987), 'Measuring travel time values with a discrete choice model: a note', *The Economic Journal* **97**, 493–498. [50]

Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, MA, USA. [56]

Bierlaire, M. (2005), 'An introduction to biogeme (version 1.4)', biogeme.epfl.ch. [55]

DeSerpa, A. (1971), 'A theory of the economics of time', *The Economic Journal* **81**, 828–846. [50]

Hensher, D. A. & Greene, W. H. (2003), 'The mixed logit model: The state of practice', *Transportation* **30**, 133–176. [47, 50]

Horowitz, J. (1983), 'Statistical comparison of non-nested probabilistic discrete choice models', *Transportation Science* **17(3)**, 319–350. [56]

Jara-Diaz, S. R. (1998), Time and income in travel choice: Towards a microeconomic activity-based theoretical framework, *in* T. Gärling, T. Laitia & K. Westin, eds, 'Theoretical Foundations of Travel Choice Modelling', Pergamon, Oxford, pp. 51–73. [50]

Nielsen, O. A., Hansen, C. O. & Daly, A. (2001), *A large-scale model system for the Copenhagen-Ringsted railway project*, Travel Behaviour Research, The Leading Edge, Pergamon Press, Australia, chapter 35, pp. 603–26. Edited by D. Hensher. [48, 51]

Revelt, D. & Train, K. (1998), 'Mixed logit with repeated choice: Household's choices of appliance efficiency level', *The Review of Economics and Statistics* **80**(4). [49]

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY, USA. [49]

Wardman, M. (2004), 'Public transport values of time', *Transport policy* **11**, 363–377. [48]

Appendix

Table 4.7: ESTIMATION RESULTS FOR a MODELS

	MNLa		MixIa		MixIIa		MixIIIa	
DoF	17		23		23		24	
N. of obs.	5184		5184		5184		5184	
N. of ind.	5184		331		331		331	
Final LL	**-2784.62**		**-2551.45**		**-2536.26**		**-2524.9**	
Adj. ρ^2	0.220		0.284		0.288		0.291	
Parameters	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**
β^0_{aux}	-2.297	-14.7	-1.957	-10.7	0.792	4.4	0.770	4.7
β^0_c	-1.485	-10.9	-1.275	-7.1	-1.103	-5.9	-1.142	-5.9
β^0_{del}	-1.598	-16.0	-1.355	-10.1	1.349	7.3	1.289	7.8
β^0_{head}	-3.301	-14.4	-3.054	-10.2	-2.619	-11.3	-2.626	-11.1
β^0_{int}	0.402	1.5	0.427	1.6	0.620	2.0	0.525	1.6
β_{intr}	0.477	1.9	0.532	2.2	0.583	2.3	0.609	2.4
β^0_{inv}	-2.250	-16.0	-2.107	-11.4	0.705	4.0	0.677	4.1
β_{purp}	-0.250	-3.0	-0.175	-1.6	-0.198	-1.7	-0.229	-1.8
β_{purps}	0.448	3.9						
β^0_{seat}	0.374	3.2	0.692	4.9	0.781	5.2	0.757	4.9
β_{seatr}	0.751	5.0	0.645	3.9	0.817	4.0	0.769	3.9
β_t	-0.013	-5.9	-0.014	-4.6	-0.014	-4.2	-0.012	-3.0
β^0_{wtint}	-1.827	-12.1	-1.614	-8.0	1.120	5.7	1.121	5.8
σ_{aux}			0.650	7.4	-0.672	-8.5	-0.692	-3.8
σ_{cost}			-0.975	-13.0	-0.905	-10.6	-1.023	-12.6
σ_{del}			-0.484	-5.9	-0.213	-1.4	-0.273	-2.4
σ_{head}			0.983	2.9	-0.571	-8.6	-0.639	-9.5
σ_{inv}			-0.539	-8.9	-0.353	-3.3	-0.379	-6.5
σ_{wtint}			-0.483	-1.3	0.532	3.9	0.461	3.3
$\rho_{cost,head}$							-0.376	-5.2

Table 4.8: ESTIMATION RESULTS FOR b MODELS

	MNLb		MixIb		MixIIb		MixIIIb	
DoF	21		27		26		27	
N. of obs.	5184		5184		5184		5184	
N. of ind.	5184		331		331		331	
Final LL	**-2713.72**		**-2535.19**		**-2517.87**		**-2511.2**	
Adj. ρ^2	0.239		0.287		0.292		0.294	
Parameters	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**
β_{age}	0.182	2.2	0.336	2.3	0.284	1.7	0.201	1.4
β_{aux}	-2.005	-14.4	-1.907	-10.3	0.790	4.7	0.776	5.3
β_c^0	-0.808	-5.9	-0.860	-3.7	-0.521	-2.7	-0.576	-3.1
β_{del}^0	-1.369	-14.5	-1.293	-9.9	1.325	7.9	1.279	9.0
β_{head}^0	-3.052	-13.7	-2.962	-10.6	-2.533	-11.9	-2.531	-13.6
β_{inc}	-0.497	-6.7	-0.540	-3.2	-0.667	-3.9	-0.601	-4.1
β_{int}^0	-0.213	-0.2	0.469	1.8	0.725	2.5	0.655	2.7
β_{intr}	0.345	1.5	0.397	1.6	0.437	1.6	0.471	2.1
β_{inv}^0	-1.949	-16.2	-2.022	-11.2	0.717	4.4	0.689	5.1
β_{purp}	-0.323	-3.9	-0.212	-1.9	-0.238	-2.1	-0.231	-2.0
β_{purps}	0.478	5.0						
β_{seat}^0	-2.017	-6.0	-0.695	-1.3	-0.503	-0.9	-0.495	-1.2
β_{seatr}	0.045	0.4	0.424	2.4	0.654	3.7	0.651	5.2
β_t	-0.018	-7.7	-0.015	-5.1	-0.014	-4.4	-0.012	-5.0
β_{timei}	0.260	1.0	0.085	1.5	0.084	1.4	0.076	1.6
β_{times}	0.859	8.8	0.462	2.9	0.428	2.6	0.413	3.3
β_{wtint}^0	-1.536	-11.7	-1.540	-8.0	1.142	6.2	1.140	7.6
σ_{aux}			0.677	7.3	-0.667	-8.9	-0.685	-8.8
σ_{cost}			-0.913	-8.0	-0.864	-9.6	-0.945	-12.3
σ_{del}			-0.478	-5.8	-0.278	-1.8	-0.244	-2.3
σ_{head}			0.939	3.0	-0.560	-7.2	-0.596	-11.0
σ_{inv}			-0.527	-8.4	-0.331	-3.4	-0.379	-6.1
σ_{wtint}			-0.480	-1.8	0.522	4.6	0.498	4.8
$\rho_{cost,head}$							-0.297	-3.8

Chapter 5

Correlation and willingness-to-pay indicators in transport modelling

Stefan L. Mabit
Centre for Traffic and Transport, Technical University of Denmark

Stephane Hess
Centre for Tranport Studies, Imperial College, London

Sebastian Caussade
Department of Transport Engineering, Pontificia Universidad Catolica de Chile

Abstract

The mixed logit model has become the state-of-the-art tool for the estimation of willingness-to-pay indicators. This has resulted in a significant amount of research into the choice of distribution in such models. An issue that has often been overlooked is the possible correlation between the marginal distributions. In this paper we investigate how distributional assumptions including different correlation structures affect the estimation of willingness-to-pay indicators for two stated-preference datasets. Furthermore, we examine if the correlation structures are affected by the inclusion of explanatory variables. Our results show that models allowing for correlation reject standard models in both preference space and willingness-to-pay space. The results also show that the inclusion of correlation can have an impact on the evaluation of willingness-to-pay indicators. The main conclusion of the paper is that the choice of correlation structure is as important as the choice of marginal distributions in mixed logit models.

5.1 Introduction and context

Discrete choice structures belonging to the family of random utility maximisation (RUM) models have established themselves as the preferred tool for the analysis of choice behaviour in the area of transport research. They are used across a variety of contexts, ranging from mode choice to destination choice, via the choice of departure time and the choice of route. An ever increasing number of different structures are available to modellers, ranging from the basic multinomial logit (MNL) model to more advanced structures, such as mixed logit, and other generalised extreme value (GEV) mixture models.

A main output of studies using discrete choice models is the computation of willingness-to-pay (WTP) indicators. These WTP indicators give an estimate of the readiness by a respondent to accept an increase in the cost of an alternative in return for an improvement to the alternative along some other dimension, such as travel time. This specific case, namely the trade-off between travel cost and travel time, gives rise to the most commonly used WTP measure, the valuation of travel time (VTT), with some recent discussions including Hensher (2001*a*,*b*,*c*), Lapparent & de Palma (2002), and Sillano & Ortúzar (2004). While the vast majority of discussions looking at the computation of WTP measures have focussed solely on the case of VTT, it is important to stress that various other WTP measures are also of interest in the context of transport research. As such, policy planners may be interested in travellers' willingness to pay for increases in frequency, improvements in on-time performance or reductions in schedule delay. As the referenced studies indicate a main tool to WTP estimation is the mixed logit model.

With the ever increasing use of the mixed logit model, an important area for research is the issue of the distributional assumptions made in such models, with some recent discussions being given by Hensher & Greene (2003) and Hess et al. (2005). Many applications of the mixed logit model rely exclusively on the normal distribution. It is also a common assumption that the distributions of individual coefficients are independent, which is clearly not always appropriate, especially in the case of stated-preference (SP) data, where, as pointed out by Train & Weeks (2005), it is reasonable to expect data to be affected by a random scale, which would induce correlated coefficients in a mixed logit model. Two examples of the relatively low number of applications looking into the correlation between randomly distributed coefficients are given by Revelt & Train (1998) and Huber & Train (2001).

From a policy perspective, the interest clearly lies in the distribution of the WTP indicators rather than the distribution of individual coefficients. With this in mind, and by noting that the computation of WTP indicators on the basis of individual randomly distributed marginal utility coefficients is a non-trivial task (cf. Section 5.2), an appealing alternative comes in the form of models working directly in WTP space, as discussed for example by Train & Weeks (2005), hereafter T&W. In WTP space, random distributions are specified directly at the level of WTP indicators rather than at the level of individual coefficients. Applications of this approach are given by Cameron & James (1987) in the field of economics and Fosgerau (2007) has applied it to VTT data, although not making use of RUM models with additive errors. However, even when working in WTP space, the choice of correlation structure is still a crucial issue, as the assumption of independent WTP indicators is unrealistic. This is supported by the empirical results in this paper.

The objectives of this paper are twofold. First, the paper emphasises that when allowing for random variations in the individual coefficients used in the computation of WTP indicators, we also need to make a decision on the correlation structure in place between these coefficients. Second, we apply specifications in preference space and WTP space to both route choice and mode choice data, hence looking at different context similar to T&W. In addition to T&W, we introduce explanatory variables in the models to investigate their effect in the various models.

Our results show that independently of whether we work in preference space or in WTP space, it is important to allow for correlation, not only due to better model fit, but also as this can have a large impact on the evaluation of WTP indicators. As such, the finding and the contribution of this paper is that the choice of a correlation structure should be an integral part of any distributional assumptions when there are two or more random coefficients.

The remainder of this paper is organised as follows. Section 5.2 presents the modelling methodology used in this paper, with results of the two analyses presented in Section 5.3. Finally, Section 5.4 summarises the findings of the paper.

5.2 Methodology

Two different modelling approaches are used in this paper, working in preference space and in WTP space. Focussing on the case of the VTT, we have in preference space a utility function (for alternative i) given by:

$$V_i = \cdots + \beta_{TT}\, TT_i + \beta_{TC}\, TC_i + \ldots, \tag{5.1}$$

where we work with a linear-in-attributes specification of travel cost, TC_i, and travel time, TT_i.

Under the assumption that all effects of travel time and travel cost are captured in the observed part of utility V, the VTT measure is simply computed as:

$$VTT = \frac{\partial V / \partial TT}{\partial V / \partial TC}. \tag{5.2}$$

The computation of the VTT on the basis of this formula is straightforward in fixed coefficients models, such as MNL and nested logit. However, complications arise once we work in a framework with random taste heterogeneity.

Therefore the issue of the distributional assumptions also have a major impact when looking beyond individual taste coefficients, i.e., when the interest lies in the WTP indicators computed on the basis of such coefficients. In general, a WTP indicator computed on the basis of two random taste coefficients does not itself follow a known statistical distribution, such that typically, the ratio between the two coefficients needs to be simulated. Here, severe problems may arise with extreme values, as discussed for example by Hensher & Greene (2003) and Hess et al. (2005). Additionally, any correlation between individual taste coefficients potentially has a significant impact on the WTP indicators.

Given the possible problems with ratios of two random coefficients, it can be preferable to work in WTP space as opposed to preference space, an approach discussed recently by Train & Weeks (2005). Here, the VTT is estimated directly as a coefficient, as opposed to being based on the ratio of two separately estimated coefficients. As such, Equation 5.1 is rewritten as:

(5.3) $$V_i = \cdots + \beta_S \beta_W TT + \beta_S TC + \ldots,$$

where $\beta_W = \frac{\beta_{TT}}{\beta_{TC}}$ is an estimate of the VTT that is obtained directly from the data and β_{TC} is rewritten as β_S to act as a scale. This approach is no different from working in preference space when all coefficients are non-random except for a possible influence on confidence intervals of the estimates. When working with random coefficients, the approach has potential advantages as issues with the ratio do no longer arise. However, correlation between the two parameters needs to be taken into account.

As highlighted above, the correlation between individual random taste coefficients potentially plays a role both in preference space and in WTP space. Nevertheless, most applications assume independence between any random parameters in mixed logit models. However, from a theoretical point of view it is straightforward to allow for correlation as illustrated in Revelt & Train (1998) as long as the distributions are derived from normals. In fact, it can be argued that the correlation between coefficients should be seen as a natural part of the distributional assumptions made in the specification of a mixed logit model. As such, a model that has two coefficients distributed independently with a normal distribution clearly makes a different distributional assumption to a model specified with two correlated normals. The specification of correlation should be based on theory, intuition and empirical evidence. Another approach is to use a heuristic method as outlined in Sørensen (2003).

Before closing the theoretical discussions, we will clarify a point indirectly presented in Train & Weeks (2005). They argue that random scale in a model induces correlation in a preference space model. Here we argue that this correlation is positive. A mixed logit model describing choice $i \in C$ for individual n at time t with random scale can be written

(5.4) $$U_{int} = s_n \beta_{TT_n} (TT_{int}) + s_n \beta_{TC_n} (TC_{int}) + \varepsilon_{int},$$

where s_n, β_{TC_n} and β_{TT_n} follow distributions in the population and ε_{int} is Gumbel distributed with constant scale. Assume that β_{TC_n} and β_{TT_n} are both independent of s_n and each other. Then the covariance between the coefficients on cost and time in the model is given by

(5.5) $$Cov(\beta_{TT_n} s_n, \beta_{TC_n} s_n) = \mu_{TT}\, \mu_{TC}\, Cov(s_n, s_n) = \mu_{TT}\, \mu_{TC}\, Var(s_n),$$

where μ_{TT}, μ_{TC} are the respective means of β_{TC_n} and β_{TT_n}. The first equality only works because of our independence assumptions. Based on economic theory it is reasonable to assume that μ_{TT} and μ_{TC} are both negative, therefore a random scale induces positive correlation in this case. Since this is the limiting case for the correlation between β_{TC_n} and β_{TT_n} going toward zero it is possible to obtain positive correlation between the coefficients on cost, $s_n\beta_{TC_n}$, and time, $s_n\beta_{TT_n}$, even in the case where β_{TC_n} and β_{TT_n} are negatively correlated.

5.3 Empirical applications

Two empirical analyses were conducted for this paper, using SP data collected in Denmark and in Switzerland. The two studies are described in more detail in Section 5.3.1 and Section 5.3.2. Before proceeding with a description of the two studies, a few additional points need to be addressed.

As highlighted in Section 5.2, one of the main issues in the specification of a mixed logit model is the choice of distribution for random parameters. In all models, we relied on a lognormal distribution for the cost coefficient, in conjunction with a sign change for the attribute. This choice is motivated by utility theory where we have $\partial V/\partial TC < 0$ and from the framework of DeSerpa where $\partial V/\partial TT \leq 0$. So within this economic foundation only a nonnegative VTT is theoretically consistent. The papers by Jara-Diaz (2000) and Mackie et al. (2001) provide excellent overviews of the development in economic theory to support this. The points made by Hess et al. (2005) in terms of not imposing fixed constraints on a distribution were addressed in preliminary tests that estimated offset parameters for the distribution. Here, no evidence was found to suggest that the bound of the distribution should not be at zero. A separate issue with the lognormal distribution is the long tail to one side. However, with the focus of the present analysis being on the effects of correlation and the difference between preference and WTP space, the use of the lognormal should be acceptable. For the time coefficients in preference space and WTP coefficients in WTP space, models were estimated with both normal and lognormal distributions. This leaves us with the scale parameter in the WTP space models. A scale is by nature positive. So for similar reasons used when choosing β_{TC} to be lognormally distributed in the preference space models we also assume that the scale coefficient in the WTP space models is lognormal.

In models estimated on SP data, it is important to recognise the repeated choice nature of the data. For the random coefficients, this was taken into account by carrying out the integration at the level of individual respondents as opposed to individual choice situations, hence assuming that tastes vary across respondents, but not within respondents (cf. Train 2003). We also allowed for an additional individual-specific effect that is shared across alternatives and is specific to a given respondent. Here, we used an approach based on an error components specification that is discussed in more detail by Hess (2007). This effect was insignificant in the Danish data and it was therefore left out of the modelling.[1]

All estimations reported here were carried out using BIOGEME (Bierlaire 2005).

5.3.1 Analysis on Danish data

The first analysis made use of the data collected for the DATIV study carried out in Denmark in 2004 (cf. Fosgerau et al. 2006). We used a sample of 477 car commuters, each answering binary SP route choice questions referenced around a recent car trip. The alternatives were described

[1] This seems to be a reasonable result for unlabelled choices.

Table 5.1: Statistics on explanatory variables

variable	mean	min	max
TT_1 (min)	29.3	1	566
TT_2 (min)	29.3	1	566
TC_1 (DKK)	37.2	0	622.5
TC_2 (DKK)	37.1	0	585
x_{cong} (share)	1.10	1	1.49
x_{inc} (scale)	3.9	0	11
x_{noinc} (scale)	0.07	0	1
t_{ref} (hr)	0.54	0.08	5

by 3 attributes; cost, free flow time, and congested time, where the latter had a fixed ratio to the total time. Based on this x_{cong} is defined as 1 plus the congestion share. For each respondent, there was 8 observations per respondent, we thus obtained a final sample of $3,816$ observations. The attributes together with the additional variables income, x_{inc}, and reference travel time, t_{ref}, are described in Table 5.1.[2]

Model specification

Table 5.2 gives an overview of the models estimated on the DATIV data.

Table 5.2: MODELS ESTIMATED ON DANISH DATA

spec/model	mnl	log	logwtp	norm	normwtp
basic	1	2	3	4	5
corr.	-	6	7	8	9
ev.	10	11	12	13	14
ev. corr.	-	15	16	17	18

The first row of models are all basic models with either the specification in preference space

$$U_i = \delta_i + \beta_{TT}\left(\frac{x_{cng}}{\bar{x}_{cng}}\right)^{\lambda_{cng}} TT_i + \beta_{TC} TC_i + \varepsilon_i$$

or WTP space

$$U_i = \delta_i + \beta_W\left(\frac{x_{cng}}{\bar{x}_{cng}}\right)^{\lambda_{cng}} \beta_S TT_i + \beta_S TC_i + \varepsilon_i.$$

Here, δ_i is a constant for alternative i, with TT_i and TC_i giving the time and cost attributes for the alternative. The variable x_{cng} gives the rate of congestion for the current respondent, with $\bar{x}_{cng}$ giving the mean for this attribute in the sample. The parameter λ_{cng} gives the elasticity of the travel time sensitivity to the degree of congestion, where a positive estimate for λ_{cng} indicates

[2]See Section 3.3.1 for a description of the income variable.

that the travel time sensitivity rises with increasing congestion.

For the four mixed logit models in the first row, the random distributions are assumed to be independent across coefficients. In the two log models of Table 5.2, both coefficients follow a lognormal distribution, while in the two norm models, a normal distribution is used for the time coefficient (preference space) and the VTT (WTP space), with a lognormal distribution being used for the cost coefficient (preference space) and the scale (WTP space).

The difference between the models in the first and second rows is that the models in the latter group allow for correlation between the random distributions.

In the third and fourth rows, the models are expanded to include further explanatory variables, namely x_{inc}, and t_{ref}, giving the income for the respondent, and the travel time for the reference trip. The specification has been used in other studies, but we have no knowledge of studies including factors on all attributes to allow for the explanatory variables to explain the scale. A continuous interaction specification is used again, where, with

$$f(x,t;\lambda^k) = \left(\frac{x_{cng}}{\bar{x}_{cng}}\right)^{\lambda^k_{cng}} \left(\frac{x_{inc}}{\bar{x}_{inc}}\right)^{\lambda^k_{inc}} \left(\frac{t_{ref}}{\bar{t}_{ref}}\right)^{\lambda^k_{tref}} \tag{5.6}$$

we have:

$$U_i = \delta_i + \beta_{TT} f(x,t;\lambda^{TT})\, TT_i + \beta_{TC} f(x,t;\lambda^{TC})\, TC_i + \varepsilon_i. \tag{5.7}$$

In the WTP space models the explanatory variables enter in the following way

$$U_i = \delta_i + \beta_W f(x,t;\lambda^W) f(x,t;\lambda^S) \beta_S\, TT_i + \beta_S f(x,t;\lambda^S)\, TC_i + \varepsilon_i. \tag{5.8}$$

For respondents not providing income information, an interaction with a dummy variable was included (λ_{ninc}).

Model results

The estimation of the mixed logit models made use of 600 MLHS draws (cf. Hess et al. 2006), where no change in stability was observed beyond 300 draws.

Using the numbering from Table 5.2, the model fits are summarised in Table 5.3, with detailed estimation results for all models in Tables 5.10 and 5.11.[3] The results show that the four basic mixed logit models (models $2-5$) all outperform the MNL model, where the models making use of two lognormals outperform the models combining a normal distribution with a lognormal distribution. In the former group, working in WTP space leads to better fit than working in preference space, while the converse is the case in the models combining a normal and lognormal distribution. The relative ranking of models is maintained when introducing correlation between

[3]The notation used is $\beta_W = \beta_{TT} = \beta_{time} + \sigma_{time} u$ and $\beta_S = \beta_{TC} = exp(\beta_{cost} + \sigma_{cost} u)$ with $u \sim N(0,1)$.

random coefficients, where this leads to significant gains in model fit across models.[4] Note that model 6 and 7 are equal, which is to be expected since they are theoretical equivalent. Introducing additional explanatory variables (models 10-18) leads to significantly better model fit across all specifications, where again, the relative rankings are maintained.[5] A summary of the correlation between random terms is given in Table 5.4. Here, the results for models 6 to 9 and models 15 to 18 show that, when working in preference space, the correlation between the time and cost coefficients turned out to be positive, while, when working in WTP space, the correlation between the VTT and the scale parameter was negative. The discussion at the end of Section 5.2 is crucial in the interpretation of these observations. Models 10 to 18 include explanatory variables that parameterise the VTT. Table 5.13 shows the estimates for these coefficients. For the models in preference space the elasticity of VTT is found as the difference between individual elasticities, e.g, $\lambda_{inc}^{TT} - \lambda_{inc}^{TC}$, and for the models in WTP space the elasticity of VTT is found as a single coefficient, e.g., λ_{inc}^{W}. The results show rising VTT with increases in income and travel time. The scale coefficient, e.g., λ_{inc}^{S}, shows the effect on the scale. The inclusion of two coefficients for each explanatory variable shows that there is significant heteroscedasticity. The travel time and the unknown income dummy are seen to have a significant effect on the scale, whereas income and congestion do not affect the scale significantly. It is reasonable that scale increases with time, i.e, that the variance of the error becomes larger. The significance of the no income dummy could indicate that individuals not giving income information pay less attention to the SP task.

Table 5.3: MODEL FIT RESULTS ON DATIV DATA

Model	par.	Final LL	Adj ρ^2	Model	par.	Final LL	Adj ρ^2
1	4	-2,494.79	0.0553	10	11	-2,384.44	0.0944
2	6	-2,144.76	0.1869	11	13	-2,082.90	0.2076
3	6	-2,112.03	0.1992	12	13	-2,068.15	0.2132
4	6	-2,158.84	0.1816	13	13	-2,090.03	0.2049
5	6	-2,162.00	0.1804	14	13	-2,115.06	0.1955
6	7	-2,094.29	0.2056	15	14	-2,051.38	0.2192
7	7	-2,094.29	0.2056	16	14	-2,051.38	0.2192
8	7	-2,116.35	0.1972	17	14	-2,065.44	0.2138
9	7	-2,157.68	0.1816	18	14	-2,112.25	0.1961

Table 5.4: ESTIMATES OF CORRELATION

Pref. space model	6	8	15	17
corr.	0.53	0.31	0.51	0.37
WTP space model	7	9	16	18
corr.	-0.05	-0.08	-0.08	-0.04

[4]The coefficient ρ reported corresponds to the off-diagonal element in the Choleski factorisation of the underlying normal distribution.

[5]Here significant gains signify that the models 1-9 are all rejected in LR tests with 7 degrees of freedom.

Table 5.5: MEAN VTT FOR DATIV IN DKK PER HOUR

spec/model	mnl	log	logwtp	norm	normwtp
basic	64.9 (0.1)	130.9 (13.0)	67.8 (3.8)	116.6 (9.0)	45.2 (1.7)
corr.		93.0 (5.0)	92.9 (5.1)	-147.1 (125.4)	57.5 (1.8)
ev.	54.9 (0.1)	107.8 (7.4)	67.0 (3.0)	98.0 (5.1)	51.6 (1.7)
ev. corr.		89.3 (4.3)	89.2 (4.2)	35.0 (28.7)	60.4 (1.7)

Table 5.6: Statistics on explanatory variables

variable	mean	min	max
TT_t (min)	174.9	35	511
TT_{sm} (min)	87.9	12	333
TT_c (min)	138.9	0	416
TC_t (CHF)	100.4	9	576
TC_{sm} (CHF)	121.4	11	768
TC_c (CHF)	88.9	0	286
HW_t (min)	69.7	30	120
HW_{sm} (min)	20.1	10	30
t_{ref} (min)	157.4	39	337.8

Even though the explanatory variables are seen to have a significant effect on the scale, the size of correlation in Table 5.4 does not seem to be affected.

VTT evaluation

The mean VTT in Table 5.5 are evaluated using simulation as described for example by Hensher & Greene (2003). The standard deviation of the mean is given in parenthesis. The evaluation is done at mean time, income, and congestion. We only comment on the VTT for models without explanatory variables as the pattern and sizes of mean VTT are similar. The mean values range from 35.0 to 130.9 DKK per hour if we neglect the strange behaviour of model 8.

Two immediate observations are that the lognormal specification gives rise to higher VTT and that the estimation in WTP space lowers the VTT. In preference space, correlation is seen to lower VTT and in WTP space the VTT rises in the models allowing for correlation.

5.3.2 Analysis on Swiss data

In the second application, we used SP data collected on a new hypothetical high-speed railway alternative in Switzerland, called the Swissmetro (see Bierlaire et al. 2001). The data were from 524 fare paying business travellers making $4,716$ mode choices. The choice set contains three alternatives, namely car, train, and Swissmetro, with alternatives described by three attributes: travel time (TT), headway (HW), and travel cost (TC), where headway is used only for train and Swissmetro. The reference travel time is retained as explanatory variable.

Model specification

We estimated the model types from Table 5.2 on the Swiss data. One difference was an added error component to test for panel effects, as described at the start of Section 5.3. With this addition the MNL specification became:

$$U_i = \delta_i + \sigma_p u_i + \beta_{TC} TC_i + \beta_{TT_i} TT_i + \beta_{HW} HW_i + \varepsilon_i, \tag{5.9}$$

where u_i is an individual-specific draw from a $N(0,1)$ distribution for each alternative and the headway, HW, is zero for the car alternative.

Similar to the MNL model, the mixed models in the first row of Table 5.2 are all basic models with either the specification in preference space

$$U_i = \delta_i + \sigma_p u_i + \beta_{TC} TC_i + \beta_{TT_i} TT_i + \beta_{HW} HW_i + \varepsilon_i, \tag{5.10}$$

or WTP space

$$U_i = \delta_i + \sigma_p u_i + \beta_S TC_i + \beta_S(\alpha_{TT_i} TT_i + \alpha_{HW} HW_i) + \varepsilon_i, \tag{5.11}$$

where β_{TC}, β_S follow lognormal distributions and $\beta_{TT}, \beta_{HW}, \alpha_{TT}, \alpha_{HW}$ follow either a lognormal or a normal distribution with the time coefficient being alternative specific. The models in the second row allow for correlation between the cost coefficient and the other random coefficients in the first row models. The models in the final two rows add reference travel time as an explanatory variable to the models from the first and second row using the elasticity formulation, see Equation 5.6. Income was tested but left out since it only had a weak effect.

Since we allow for alternative-specific time coefficients we have five random coefficients plus the panel error component. We only look at correlation between the random coefficients. Five coefficients potentially lead to 10 correlation coefficients. We restrict ourselves to 4 since it is mainly the correlation with the cost coefficient that affect the evaluation of WTP. Furthermore, estimation of all 10 might in theory be possible but in an actual application, it puts high demands on data. Even with 4 correlation coefficients we had problems with local maxima in the estimation procedure due the complexity of the models.

Model results

The estimation of the mixed logit models made use of $1,000$ MLHS draws, where no change in stability was observed beyond 500 draws. The results on model fits are given in Table 5.7 with detailed estimation results in Tables 5.12 and 5.13. The MNL model 1 is estimated with the added panel error component where the log-likelihood increases from -3375.48 based on a simple MNL model (not reported) to -2472.56. This is remarkable with just one additional parameter. The gains in fit by including this term are not as significant in the mixed logit models, where the panel effects are also captured through the specification of random taste heterogeneity. The mixed logit models 2-5 are all estimated with normal and lognormal distributions or lognormals only in preference space and in WTP space. The best model fit is obtained in model

Table 5.7: MODEL FIT RESULTS ON SWISSMETRO DATA

Model	par.	Final LL	Adj ρ^2	Model	par.	Final LL	Adj ρ^2
1	8	-2,472.56	0.5008	10	11	-2,402.17	0.5144
2	11	-2,383.68	0.5181	11	14	-2,340.40	0.5262
3	11	-2,370.12	0.5208	12	14	-2,351.11	0.5241
4	11	-2,378.36	0.5192	13	14	-2,337.53	0.5267
5	11	-2,378.64	0.5191	14	14	-2,357.68	0.5227
6	15	-2,325.19	0.5291	15	18	-2,307.49	0.5320
7	15	-2,325.19	0.5291	16	18	-2,307.49	0.5320
8	15	-2,348.17	0.5245	17	18	-2,310.58	0.5314
9	15	-2,347.87	0.5245	18	18	-2,321.89	0.5295

Table 5.8: ESTIMATES OF CORRELATION BETWEEN COST AND CAR TRAVEL TIME

Pref. space model	6	15	8	17
corr.	0.7938	-0.707	-0.6816	-0.791
WTP space model	7	16	9	18
corr.	0.9316	-0.490	0.8919	0.980

3 which uses lognormal coefficients in WTP space.

The specifications in models 2-5 were then estimated allowing for correlation between the cost and the remaining random taste coefficients. The results for the correlation between cost and car travel time in models 11 and 13, and scale and car VTT in models 12 and 14 can be seen in Table 5.8. The correlation between the cost coefficient and the other coefficients are significant for the time coefficients but insignificant for headway. The significant correlations are of the same sign for each model.

In the last nine models we included reference travel time as explanatory variable, with results summarised in Table 5.13 for the eight mixed logit models. This shows a significant effect on scale and WTP for headway, whereas the effect is only significant on VTT in model 16.

WTP evaluation

The evaluation of the mean WTP is summarised in Table 5.9, where again, simulation was used. As a general observation, the WTP for headway is much lower than the VTT. The VTT varies between modes with train in general having the highest value. For a specific mode, e.g., car, we see that the value varies between 90.9 CHF per hour in model 14 and 248.5 CHF per hour in the normal model with correlation (model 8). The values are rather high but one has to remember that the sample consist of business travellers in Switzerland and that we evaluate at the average travel time of 150 minutes.

Table 5.9: MEAN WTP FOR SWISS DATA IN CHF PER HOUR

Model	1	2	3	4	5
WTP head	24.6	57.5	48.4	33.1	29.9
VTTS car	112.3	133.3	91.4	136.5	92.0
VTTS sm	103.9	116.8	86.4	122.9	92.3
VTTS train	94.0	138.2	106.3	138.2	104.3
Model		6	7	8	9
WTP head		36.5	36.4	62.7	30.7
VTTS car		103.6	103.6	248.5	92.0
VTTS sm		132.2	132.2	301.2	93.6
VTTS train		129.7	129.7	285.5	111.7
Model	10	11	12	13	14
WTP head	22.4	71.1	49.3	34.3	31.6
VTTS car	111.5	139.4	92.7	147.4	90.9
VTTS sm	74.6	83.2	77.3	96.2	79.9
VTTS train	84.4	114.9	94.0	128.6	92.8
Model		15	16	17	18
WTP head		58.9	54.6	47.3	32.5
VTTS car		196.3	196.7	186.8	93.4
VTTS sm		242.8	241.5	187.2	62.3
VTTS train		238.4	236.8	197.6	95.5

5.3.3 Discussion of analyses

In summarising the results, we see that with the DATIV data, working in WTP space gives the best results for the log models while working in preference space is preferable for the norm models. In the Swissmetro data the picture is more blurred. The only clear conclusion is that standard models in both preference space and WTP space are rejected by models allowing for a more complex correlation pattern.

When working purely with correlated lognormal distributions, the models in preference space and WTP space are formally equivalent, although this can be hard to establish in estimation, potentially due to local maxima.

The fact that the sign on correlations are not similar in the two applications does not pose a contradiction of the discussion in Section 5.2; it merely shows that the sign is unknown prior to testing on data.

The evaluation of the WTP indicators shows that the WTP is affected by the presence of correlation. The general trend in DATIV is that including correlation lowers the mean VTT in preference space while raising it for WTP space models. The inclusion of explanatory variables is seen to be important on both datasets. It does affect the WTP indicators in the base situation and could of course affect it even more in the case of a forecast. The fact that both datasets show heteroscedasticity also indicates that this is an important issue which is not obsolete in mixed logit models even though they allow for heteroscedasticity with respect to attributes.

5.4 Summary and conclusions

In the estimation of mixed logit models the choice of distribution is important. However, existing work in transport research has mainly focused on the marginal distribution of attributes. This paper highlights that a multivariate distribution consists of correlation in addition to marginal distributions. Two approaches that represent different distributional assumptions in this sense are modelling in preference space and WTP space. Here we investigate how both specification types compare to models allowing for more general correlation structures. In the case of models with all lognormal coefficients these two approaches represent restrictive correlation patterns and they can be embedded in a common model. The effect of correlation is also investigated in models where the scale and WTP are parameterised by explanatory variables.

The results show that both WTP space and preference space are rejected by the models allowing for more complex correlation structures. Furthermore, the results show that the inclusion of correlation may have an impact on the evaluation of willingness-to-pay indicators.

The investigations underline that some correlation probably comes from a random scale when individuals evaluate SP choices. The inclusion of explanatory variables in the scale shows that especially travel time has a significant effect on the scale. This emphasises the need to use heteroscedastic models. While this has previously been discussed at the MNL level with the use of heteroscedastic logit models (e.g., Caussade et al. 2005), it seems to have been neglected in the context of mixed logit studies, where the focus has been on the distributional assumptions.

The main conclusion of the paper is that decisions in terms of correlation between random coefficients are as important as decisions in terms of the choice of marginal distributions in a mixed logit model. As the applications show, correlation may change results both related to fit and evaluation of WTP indicators.

Acknowledgment
This work was initiated during the first workshop on Applications of Discrete Choice Models organised at EPFL in Lausanne, Switzerland, September 2005. The authors would like to thank Mogens Fosgerau and Katrine Hjort for valuable comments on an earlier draft. Part of the work was carried out during visits by the second author to the Institute for Transport and Logistic Studies at the University of Sydney and the Department of Civil and Environmental Engineering at the Massachusetts Institute of Technology.

References

Bierlaire, M. (2005), 'An introduction to biogeme (version 1.4)', biogeme.epfl.ch. [67]

Bierlaire, M., Axhausen, K. W. & Abay, G. (2001), *The acceptance of modal innovation : The case of Swissmetro*, Proceedings of the 1^{st} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland. [71]

Cameron, T. & James, M. (1987), 'Efficient estimation methods for closed-ended contingent valuation survey data', *Review of Economics and Statistics* **69**, 269–276. [64]

Caussade, S., Ortúzar, J. de D., Rizzi, L. I. & Hensher, D. A. (2005), 'Assessing the influence of design dimensions on stated choice experiment estimates', *Transportation Research Part B: Methodological* **39**(7), 621–640. [75]

Fosgerau, M. (2007), 'Using nonparametrics to specify a model to measure the value of travel time', *Transportation Research Part A* **41**, 842–856. [64]

Fosgerau, M., Hjorth, K. & Lyk-Jensen, S. V. (2006), *An integrated approach to the estimation of the value of travel time*, European Transport Conference, Strasbourg, 2006. [67]

Hensher, D. A. (2001*a*), 'Measurement of the valuation of travel time savings', *Journal of Transport Economics and Policy* **35**(1), 71–98. [64]

Hensher, D. A. (2001*b*), 'The sensitivity of the valuation of travel time savings to the specification of unobserved effects', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 129–142. [64]

Hensher, D. A. (2001*c*), 'The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications', *Transportation* **28**, 101–118. [64]

Hensher, D. A. & Greene, W. H. (2003), 'The mixed logit model: The state of practice', *Transportation* **30**, 133–176. [64, 65, 71]

Hess, S. (2007), 'Posterior analysis of random taste coefficients in air travel choice behaviour modelling', *Journal of Air Transport Management, paper accepted for publication (February 2007)* . [67]

Hess, S., Bierlaire, M. & Polak, J. W. (2005), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236. [64, 65, 67]

Hess, S., Train, K. & Polak, J. W. (2006), 'On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice', *Transportation Research Part B: Methodological* **40**(2), 147–163. [69]

Huber, J. & Train, K. E. (2001), 'On the similarity of classical and bayesian estimates of individual mean partworths.', *Marketing Letters* **12 (3)**, 259–269. [64]

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. [67]

Lapparent, M. D. & de Palma, A. (2002), *Non-linearities in the valuations of time estimates*, Proceedings of the PTRC Annual Meeting, 2002, PTRC, London. [64]

Mackie, P. J., Jara-Diaz, S. R. & Fowkes, A. S. (2001), 'The value of travel time savings in evaluation', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 91–106. [67]

Revelt, D. & Train, K. (1998), 'Mixed logit with repeated choice: Household's choices of appliance efficiency level', *The Review of Economics and Statistics* **80**(4). [64, 66]

Sillano, M. & Ortúzar, J. de D. (2004), 'Willingness-to-pay estimation from mixed logit models: some new evidence', *Environment & Planning A* **37**(3), 525–550. [64]

Sørensen, M. V. (2003), Discrete Choice Models. Estimation of passenger traffic., PhD thesis, Danish Technical University, DK-2800 Kgs. Lyngby, Denmark. [66]

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY, USA. [67]

Train, K. & Weeks, M. (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 1, pp. 1–16. [64, 66]

Appendix

Table 5.10: ESTIMATION RESULTS FOR DATIV MODELS 1 - 9

Model	1		2		3		4		5	
DoF	4		6		6		6		6	
Final LL	**-2494.8**		**-2144.8**		**-2112.0**		**-2158.8**		**-2162.0**	
Adj. ρ^2	0.055		0.187		0.199		0.182		0.180	
Coefficients	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_1	0.104	3.1	0.178	4.0	0.196	4.1	0.177	4.1	0.185	4.0
λ_{cng}	2.408	6.6	2.362	4.8	3.150	68.0	1.880	3.4	2.199	53.2
β_{cost}	0.103	8.8	-0.978	-9.0	-0.674	-4.5	-1.019	-9.4	-0.954	-7.9
β_{time}	0.111	10.4	-1.347	-13.3	-0.596	-100.0	0.318	10.9	0.754	113.5
σ_{cost}			-1.264	-13.8	2.463	9.9	1.251	15.1	1.807	9.5
σ_{time}			-0.817	-14.3	-1.194	-136.7	0.198	8.7	0.956	110.7
Model			6		7		8		9	
DoF			7		7		7		7	
Final LL			**-2094.3**		**-2094.3**		**-2116.4**		**-2157.7**	
Adj. ρ^2			0.206		0.206		0.197		0.182	
Coefficients			**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_1			0.199	4.2	0.199	4.2	0.196	4.2	0.185	4.0
λ_{cng}			2.599	22.5	2.599	22.5	1.706	4.4	2.006	8.0
β_{cost}			-0.680	-5.3	-0.680	-5.3	-0.968	-9.1	-0.914	-8.0
β_{time}			-0.927	-8.9	-0.247	-3.6	0.498	13.1	0.958	12.3
σ_{cost}			2.100	13.6	2.100	13.6	1.884	23.3	1.716	10.5
σ_{time}			-0.798	-31.7	-0.798	-31.7	0.174	7.8	0.890	19.5
ρ			1.249	8.1	-0.851	-31.1	0.345	12.1	-0.236	-24.8

Table 5.11: ESTIMATION RESULTS FOR DATIV MODELS 10 - 18

Model	10		11		12		13		14	
DoF	11		13		13		13		13	
Final LL	**-2384.4**		**-2082.9**		**-2068.2**		**-2090.0**		**-2115.1**	
Adj. ρ^2	0.094		0.208		0.213		0.205		0.195	
Coefficients	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_1	0.114	3.3	0.179	4.1	0.192	4.1	0.178	4.1	0.185	4.1
$\lambda^W_{cng}, \lambda^{TT}_{cng}$	3.760	6.3	2.454	3.4	2.799	36.6	2.313	3.6	1.681	7.3
$\lambda^S_{cng}, \lambda^{TC}_{cng}$	1.383	2.4	0.228	0.3	0.138	0.1	0.262	0.3	0.523	0.6
β_{cost}	0.139	13.2	-1.134	-11.5	-0.871	-6.3	-1.153	-11.5	-1.075	-9.4
$\lambda^W_{inc}, \lambda^{TT}_{inc}$	0.360	1.9	0.396	1.9	0.575	15.1	0.337	1.8	0.243	2.7
$\lambda^S_{inc}, \lambda^{TC}_{inc}$	-0.054	-0.3	0.030	0.1	0.159	0.5	0.027	0.1	0.010	0.0
$\lambda^W_{ninc}, \lambda^{TT}_{ninc}$	-0.718	-1.5	-1.029	-2.5	0.160	2.7	-0.937	-2.5	0.130	2.4
$\lambda^S_{ninc}, \lambda^{TC}_{ninc}$	-1.067	-2.0	-1.443	-3.1	-1.683	-2.5	-1.379	-2.9	-1.571	-3.2
β_{time}	0.127	11.5	-1.347	-13.6	-0.424	-32.4	0.303	11.5	0.861	36.7
$\lambda^W_{tref}, \lambda^{TT}_{tref}$	-0.376	-3.9	-0.570	-5.0	0.501	35.8	-0.605	-6.3	0.366	10.8
$\lambda^S_{tref}, \lambda^{TC}_{tref}$	-0.836	-8.4	-1.006	-8.7	-1.036	-6.4	-1.060	-8.7	-0.966	-7.1
σ_{cost}			-1.000	-13.6	1.850	9.7	1.029	15.4	1.389	8.6
σ_{time}			-0.747	-6.5	-1.034	-77.4	0.163	7.4	0.913	22.2
Model			15		16		17		18	
DoF			14		14		14		14	
Final LL			**-2051.4**		**-2051.4**		**-2065.4**		**-2112.3**	
Adj. ρ^2			0.219		0.219		0.214		0.196	
Coefficients			**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_1			0.194	4.1	0.194	4.1	0.193	4.2	0.185	4.1
$\lambda^W_{cng}, \lambda^{TT}_{cng}$			2.184	2.5	2.076	12.3	1.925	2.9	1.738	4.7
$\lambda^S_{cng}, \lambda^{TC}_{cng}$			0.108	0.1	0.108	0.1	-0.018	0.0	0.624	0.5
β_{cost}			-0.906	-7.7	-0.906	-7.7	-1.057	-9.9	-1.043	-7.8
$\lambda^W_{inc}, \lambda^{TT}_{inc}$			0.454	1.8	0.399	12.6	0.278	1.5	0.269	0.6
$\lambda^S_{inc}, \lambda^{TC}_{inc}$			0.055	0.2	0.055	0.2	-0.021	-0.1	0.019	0.1
$\lambda^W_{ninc}, \lambda^{TT}_{ninc}$			-1.203	-2.5	0.459	1.8	-1.029	-2.7	0.277	0.8
$\lambda^S_{ninc}, \lambda^{TC}_{ninc}$			-1.662	-3.6	-1.662	-3.6	-1.526	-3.0	-1.626	-1.6
β_{time}			-1.105	-10.0	-0.200	-3.6	0.429	11.7	1.006	4.9
$\lambda^W_{tref}, \lambda^{TT}_{tref}$			-0.598	-4.7	0.431	20.9	-0.595	-6.0	0.410	0.6
$\lambda^S_{tref}, \lambda^{TC}_{tref}$			-1.029	-8.1	-1.029	-8.1	-1.054	-7.9	-0.985	-2.6
σ_{cost}			1.728	10.4	1.728	10.4	1.538	15.9	1.374	3.5
σ_{time}			-0.871	-31.3	-0.871	-31.3	0.172	7.8	0.837	13.9
ρ			1.073	6.7	-0.655	-21.5	0.233	7.5	-0.182	-2.5

Table 5.12: ESTIMATION RESULTS FOR SWISS METRO 1 - 9

Model	1		2		3		4		5	
DoF	8		11		11		11		11	
Final LL	**-2472.6**		**-2383.7**		**-2370.1**		**-2378.4**		**-2378.6**	
Adj. ρ^2	0.501		0.518		0.521		0.519		0.519	
Coefficients	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_{car}	2.607	3.8	0.853	0.9	-0.350	-0.4	1.237	1.5	-0.178	-0.2
δ_{SM}	1.606	2.7	-0.424	-0.5	-1.352	-2.8	-0.020	0.0	-0.934	-2.3
β_{cost}	0.028	4.2	-3.240	-39.0	-3.161	-48.0	-3.264	-41.3	-3.183	-42.9
β_{head}	0.011	4.7	-5.040	-10.9	-1.949	-5.8	0.015	3.7	0.499	3.3
β_{timeC}	0.049	15.2	-2.770	-38.3	0.407	5.4	0.063	15.0	1.534	14.7
β_{timeS}	0.045	9.5	-2.908	-26.4	0.349	3.9	0.057	10.5	1.539	8.5
β_{timeT}	0.041	9.2	-2.739	-26.8	0.555	5.7	0.064	10.4	1.739	8.6
σ_p	2.519	13.5	1.773	12.1	2.008	11.0	1.715	12.4	1.868	12.1
σ_{cost}			0.750	8.0	0.633	8.0	0.779	6.7	0.648	6.6
σ_{head}			1.696	9.0	-1.853	-15.4	0.013	2.4	-0.449	-3.2
σ_{time}			-0.265	-10.8	-0.150	-8.6	-0.016	-11.2	-0.297	-12.4
Model			6		7		8		9	
DoF			15		15		15		15	
Final LL			**-2325.2**		**-2325.2**		**-2348.2**		**-2348**	
Adj. ρ^2			0.529		0.529		0.524		0.525	
Coefficients			**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_{car}			-0.383	-0.8	-0.383	-0.8	0.510	0.6	-0.624	-0.9
δ_{SM}			0.066	0.2	0.066	0.2	-0.604	-0.8	-1.185	-2.3
β_{cost}			-2.828	-30.6	-2.828	-30.6	-3.368	-28.0	-3.128	-46.0
β_{head}			-4.507	-12.1	-1.679	-4.4	0.019	5.3	0.510	4.1
β_{timeC}			-2.662	-28.4	0.166	1.8	0.065	14.8	1.533	13.8
β_{timeS}			-2.624	-27.8	0.204	2.5	0.059	11.1	1.559	9.5
β_{timeT}			-2.530	-29.0	0.298	3.3	0.071	10.5	1.861	11.9
σ_p			2.069	13.1	2.069	13.1	1.762	10.1	1.992	12.5
σ_{cost}			0.529	7.1	0.529	7.1	0.934	10.2	0.407	6.0
σ_{head}			-1.450	-8.2	-1.450	-8.2	0.014	4.1	-0.429	-3.5
σ_{time}			-0.146	-49.9	-0.146	-49.9	-0.016	-11.0	-0.306	-11.6
$\rho_{cost,head}$			0.167	0.6	-0.362	-1.3	-0.004	-1.1	0.102	0.5
$\rho_{cost,timeC}$			1.382	11.8	0.854	11.8	-0.027	-5.6	0.747	6.9
$\rho_{cost,timeS}$			1.596	14.2	1.067	16.1	-0.053	-7.9	1.236	8.3
$\rho_{cost,timeT}$			1.483	12.5	0.954	14.0	-0.035	-6.2	0.881	5.8

Table 5.13: ESTIMATION RESULTS FOR SWISS METRO 10 - 18

Model	10		11		12		13		14	
DoF	11		14		14		14		14	
Final LL	**-2402.2**		**-2340.4**		**-2351.1**		**-2337.5**		**-2357.7**	
Adj. ρ^2	0.514		0.526		0.524		0.527		0.523	
Coefficients	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_{car}	3.121	3.5	2.249	1.4	0.425	0.6	2.121	1.6	0.548	0.4
δ_{SM}	0.765	0.9	-0.730	-0.5	-1.047	-1.7	-0.632	-0.5	-0.752	-0.8
β_{cost}	0.031	7.0	-3.199	-41.0	-3.182	-51.0	-3.246	-40.0	-3.178	-41.4
β_{head}	0.011	4.9	-5.186	-11.4	-1.517	-4.2	0.016	3.2	0.526	4.0
β_{timeC}	0.057	17.1	-2.682	-42.7	0.430	6.4	0.069	16.0	1.514	12.7
β_{timeS}	0.038	8.7	-3.202	-25.7	0.245	2.8	0.045	8.9	1.332	7.0
β_{timeT}	0.043	8.6	-2.883	-23.5	0.442	4.9	0.060	8.0	1.547	6.7
σ_p	2.420	13.4	2.074	10.0	2.169	13.0	1.565	6.5	2.001	9.0
σ_{cost}			0.780	9.8	0.441	9.3	0.790	10.1	0.476	4.8
σ_{head}			1.897	8.0	-1.616	-8.8	0.012	1.9	-0.461	-3.1
σ_{time}			-0.140	-3.3	-0.080	-3.1	-0.015	-8.9	-0.231	-2.4
$\lambda^{S}_{tref}, \lambda^{TC}_{tref}$	-0.816	-3.3	-1.072	-6.7	-0.651	-5.9	-1.158	-3.8	-0.642	-5.3
$\lambda^{WHW}_{tref}, \lambda^{HW}_{tref}$	0.349	0.9	1.393	4.3	0.874	2.5	0.678	1.2	1.727	2.5
$\lambda^{WTT}_{tref}, \lambda^{TT}_{tref}$	-0.866	-9.9	-0.814	-8.2	-0.015	-0.1	-0.870	-5.7	0.006	0.0
Model			15		16		17		18	
DoF			18		18		18		18	
Final LL			**-2307.5**		**-2307.5**		**-2310.6**		**-2322**	
Adj. ρ^2			0.532		0.532		0.531		0.529	
Coefficients			**Value**	**t test**	**Value**	**t test**	**Value**	**t test**	**Value**	**t test**
δ_{car}			-0.074	-0.1	-0.074	-0.1	0.473	0.4	0.506	0.5
δ_{SM}			-1.389	-1.3	-1.389	-1.3	-1.586	-1.6	-2.069	-2.3
β_{cost}			-3.231	-43.2	-3.231	-43.2	-3.304	-39.7	-3.141	-50.0
β_{head}			-5.076	-10.3	-1.845	-3.8	0.018	3.7	0.541	2.9
β_{timeC}			-2.726	-35.6	0.505	5.1	0.070	15.0	1.558	12.5
β_{timeS}			-3.003	-23.9	0.229	1.7	0.055	7.9	1.041	6.5
β_{timeT}			-2.653	-22.1	0.578	4.0	0.073	10.1	1.593	8.2
σ_p			2.106	8.4	2.106	8.4	1.951	7.6	2.351	14.3
σ_{cost}			0.610	6.0	0.610	6.0	0.694	6.0	0.268	4.6
σ_{head}			-1.745	-7.7	-1.745	-7.7	0.009	0.9	-0.442	-2.0
σ_{time}			-0.142	-3.6	-0.142	-3.6	-0.013	-7.1	-0.056	-1.6
$\rho_{cost,head}$			0.070	0.5	-0.540	-3.3	-0.006	-1.8	0.115	1.1
$\rho_{cost,timeC}$			-0.530	-4.6	-1.140	-11.8	-0.027	-4.9	0.807	7.0
$\rho_{cost,timeS}$			-0.892	-9.7	-1.502	-18.9	-0.048	-7.1	1.366	7.7
$\rho_{cost,timeT}$			-0.622	-5.6	-1.232	-13.0	-0.029	-6.5	0.882	6.0
$\lambda^{S}_{tref}, \lambda^{TC}_{tref}$			-1.105	-8.1	-1.105	-8.1	-1.068	-7.6	-0.592	-4.9
$\lambda^{WHW}_{tref}, \lambda^{HW}_{tref}$			0.877	1.5	1.982	3.5	0.811	2.4	1.372	2.0
$\lambda^{WTT}_{tref}, \lambda^{TT}_{tref}$			-0.757	-6.3	0.348	3.2	-0.718	-5.7	-0.114	-1.0

Chapter 6

Studies of correlated willingness to pay for public transport

Stefan L. Mabit
Centre for Traffic and Transport, Technical University of Denmark

Abstract

Much research on the estimation of willingness-to-pay measures is based on mixed logit models. In recent studies, it has been shown that models with several mixing dimensions should allow for correlation. This paper investigates how to specify models to capture correlation between different willingness-to-pay measures for public transport. It is done for several modes and the distributional assumptions are tested in the final models.

The estimations show significant correlation for all modes. A main contributor to the correlation is a random scale. The correlation affects the evaluation of mean willingness to pay significantly for some modes. A mixed logit model with partial correlation structure including background variables is developed for each mode. These show that the partial correlation structure can capture the correlation and that the background variables explain part of the correlation. The principal conclusion of the paper is that correlation structures should be included in modelling but that the most reasonable structure might be neither the simplest nor the most complex.

6.1 Introduction

When public transport is part of the transport system many willingness-to-pay (WTP) indicators, such as WTP for access or waiting time, are important. Reflecting this, much research has focused on the measurement of WTP indicators in transportation. In the context of public transport a review is given by Wardman (2004).

The estimation of the value of travel time (VTT) is especially important. In travel demand

and assignment models, it plays a central role, and in appraisal it is the main driver as travel time savings can account for up till 80 % of the benefits (Mackie et al. 2001).

During the last decade it has become common to apply mixed logit models to estimate WTP indicators. This is due to the large range of behavioural aspects the model can imitate. The cost of the flexibility is that many choices have to be made in the specification of the model. An example is the choice of marginal distributions (see, e.g., Hensher 2001, Hess et al. 2005, Fosgerau 2006). The evidence shows that model results are highly dependent on the chosen distributions.

Two aspects of the mixed logit model that have seen little attention in applications are the possibility of taste correlation and distributional tests. Concerning taste correlation the inclusion of correlation is described in Revelt & Train (1998) and applied with bayesian procedures in Huber & Train (2001) and Train & Sonnier (2005). Models including taste correlation are applied to the estimation of public transport WTP by Mabit & Nielsen (2006). But due to a small dataset they only investigated a limited correlation structure based on theory. Other studies such as Train & Weeks (2005) investigated the correlation structures indirectly while Mabit et al. (2006) investigated the relationship between distributional assumptions and correlation structures.

The second issue is the testing of the distributional assumptions. This subject has had limited attention given its importance. Exceptions are McFadden & Train (2000) who address tests to test for mixing, Fosgerau (2007) who uses a nonparametric approach to compare distributions against a general alternative, and Fosgerau & Bierlaire (2007) who develop a distributional test based on a method of sieves. An application of the latter test to WTP estimation is given in Fosgerau et al. (2006) where they limit themselves to a one-dimensional distribution.

This paper investigates how to specify models to capture correlation between WTP measures for public transport. A mixed logit model is estimated for each public transport mode in the Copenhagen region and takes into account the many dimensions of a public transport journey: access and egress, cost, headway, number of interchanges, in-vehicle time, and waiting time. As part of the investigation, the paper illustrates how to test the distributional assumptions in mixed logit models.

We find significant correlation for all modes which implies that correlation between willingness-to-pay measures is not adequately described in a model assuming independent coefficients. A main contributor to the correlation is a random scale. The correlation is seen to change the evaluation of willingness to pay for some modes. A mixed logit model with partial correlation structure including background variables is developed for each mode.[1] These show that the partial correlation structure can capture the correlation and that the background variables explain part of the correlation. The random scale is also partly explained by the background variables. The principal conclusion is that correlation structures should be included in modelling. Then the estimation can decide whether independent distributions are reasonable, or if a partial or full correlation structure is necessary.

[1] Partial correlation structures are also developed in Daly & Zachary (1975).

The paper is organised as follows: Section 2 presents the mixed logit model and describes the tests applied in this paper. In Section 3 we describe the application and discuss the estimation results. Section 4 concludes.

6.2 Methodology

6.2.1 The mixed logit model and taste correlation

Suppose an individual chooses between J alternatives $j \in C$. In a mixed logit model with linear-in-parameters specification, the choice is modelled through utilities for each alternative

$$U_j = \beta' x_j + \varepsilon_j, \qquad (6.1)$$

where β are parameters that follow specified distributions, x_j are explanatory variables, and ε_j are independent Gumbel distributed error terms.

Assuming the specification in Equation 6.1, the researcher must specify the joint distribution for the random parameters in β. A flexible assumption is that β is derived from independent normal distributions through a monotone transformation (Train & Sonnier 2005), i.e., $\beta = h(u)$ where $u \sim N(\mu, \Sigma)$ and h is a vector of monotone transformations of the marginal distributions. In general the covariance matrix Σ may be described by a Choleski factorisation, i.e., L such that $\Sigma = LL'$ where L is a lower triangular matrix. In case of a model with independent coefficients the Choleski factorisation is a diagonal matrix (matrix b in Figure 6.1).

$$a)\begin{bmatrix}0 & 0 & 0\\ 0 & s_2 & 0\\ 0 & 0 & s_3\end{bmatrix} \quad b)\begin{bmatrix}s_1 & 0 & 0\\ 0 & s_2 & 0\\ 0 & 0 & s_3\end{bmatrix} \quad c)\begin{bmatrix}s_1 & 0 & 0\\ s_{12} & s_2 & 0\\ s_{13} & 0 & s_3\end{bmatrix} \quad d)\begin{bmatrix}s_1 & 0 & 0\\ s_{12} & s_2 & 0\\ s_{13} & s_{23} & s_3\end{bmatrix}$$

Figure 6.1: Four different Choleski matrices

The Choleski matrix determines the covariance of the marginal utility coefficients. What really matters is the covariance among the WTP indicators. Suppose that our model includes a cost coefficient, β_c, in our mixed logit model in Equation 6.1. Then WTP for the k'th attribute is given by β_k/β_c. If the cost coefficient is fixed then the covariance between coefficients and the covariance between WTP measures are proportional. If the cost coefficient is random then the covariance of the WTP will depend on the specific distributions. Suppose that the first coefficient is the cost coefficient. Then Choleski matrix a) would restrict WTP for attributes two and three to be independent and the cost coefficient to be fixed; Choleski matrix b) allows a restricted correlation were attributes two and three are correlated only through the cost coefficient; Choleski matrix c) represents a less restricted correlation pattern that allows for a random scale;[2] and Choleski matrix d) allows for any correlation pattern. This illustrates how the choice of Choleski matrix imposes strict restrictions on the correlation pattern in WTPs that a model may describe.

In an application it is necessary to think both about identification and computational issues. The identification is not different from the general identification of the random tastes. These are identified as discussed in Revelt & Train (1998) by the variation in attributes. Therefore it is only a question of data size and variation in the attributes whether correlation is empirically identified. The computational aspect of adding correlation is that these extra parameters will demand longer run times as any additional parameter in MSL estimation. On the other hand the distribution becomes more realistic without adding an extra dimension of integration.

6.2.2 Tests

In this section we will discuss how to test the support and the marginal distributions in a mixed logit model.

In many contexts it is natural that a mixed coefficient has a restricted support, e.g., it is natural to assume that the travel time coefficient has non-positive support. If the coefficient follows a distribution with density $f(\beta)$ then the support assumption may be tested by allowing the distribution to be shifted as $\delta + f(\beta)$. The hypothesis $\delta = 0$ is then tested in a likelihood ratio (LR) test.

The question of marginal distributions has led to much research with no definite conclusion only that the choice is influential and context-dependent. Little guidance exists except that the distributions should comply with theory. Some exceptions concerning methods on how to choose distributions are Hensher & Greene (2003) who advocate a jackknife procedure and Fosgerau (2007) who uses nonparametric methods.

Fosgerau & Bierlaire (2007) develop a test for the marginal distribution in the mixed logit model. The intuition is that in the model we assume some distribution f with cumulative distribution function (CDF) F whereas the true distribution is g with CDF G. Since both CDFs have range within the unit interval we have

$$G(w) = Q(F(w))$$

for some function Q and

$$g(w) = q(F(w))f(w)$$

with $q = Q'$. If our model is correctly specified q will be unity. In their seminonparametric (SNP) test, $q = 1$ is tested by replacing $q(F(w))$ with a Legendre approximation. Fosgerau & Bierlaire (2007) recommend to use 3 terms in the approximation.

[2]As a random scale is assumed away in a mixed logit model it enters the model as a common part of each coefficient. Therefore a model with random scale must allow for some correlation between any pair of marginal distributions. One way to do this is by letting one variable be correlated to all the remaining.

Table 6.1: STATISTICS ON THE ATTRIBUTES

	Train		Bus		City train	
no. of obs.	3455		7751		2731	
variable	mean	range	mean	range	mean	range
Δx_c (Euro)	-0.2	-93 - 93	-0.1	-33 - 35	0.0	-22 - 22
Δx_{ae} (min)	0.1	-105 - 105	-0.1	-89 - 80	-0.7	-79 - 96
Δx_h (min)	-0.5	-105 - 105	0.5	-105 - 105	-0.7	-100 - 90
Δx_{int} (scale)	0.0	-2 - 2	0.0	-2 - 2	0.0	-2 - 2
Δx_{inv} (min)	-0.1	-244 - 276	-0.1	-108 - 144	0.2	-46 - 46
Δx_w (min)	-0.2	-15 - 15	0.1	-15 - 15	0.0	-15 - 15

Table 6.2: DESCRIPTIVE STATISTICS FOR BACKGROUND VARIABLES

variable	Train	Bus	City train
no. of indiv.	523	1148	401
$\bar{s}_{inc}$ (Euro)	20508	16755	22861
$\bar{s}_{time}$ (min)	94.7	43.4	35.1

6.3 Data and estimation

In this section we present an application of mixed logit models with correlated coefficients to the estimation of public transport WTP measures. First, we present the data and the empirical model. Second, we present results for the train mode and a summary for the other modes. Third, we discuss similarities and differences across the three modes.

6.3.1 Data

The data were collected in 2004 as part of the newest Danish VTT study (cf. Fosgerau et al. 2006). These data are part of DATIV 2. The data consist of unlabelled binary choices of public transport. The modes are train, bus, s-train and metro. Here we have pooled s-train and metro into city train based on initial estimation. An observation corresponds to a route choice between two lines each with the same unobserved characteristics as a reference trip. The dataset is a panel of stated-preference (SP) choices where each individual makes up to eight choices.

The explanatory variables consist of three groups: design variables (attributes), trip characteristics, and person characteristics. We will denote the design variables as attributes, x, and the latter two groups as background variables, s. The attributes are travel cost, x_c, access/egress time, x_{ae}, headway, x_h, number of interchanges, x_{int}, in-vehicle time, x_{inv}, and waiting time, x_w. They are summarised in Table 6.1 where $\Delta x_c = x_{c1} - x_{c2}$, etc. In-vehicle time and cost are seen to have larger range for train than bus and larger for bus than city train. The other variables are similar across the modes. The units are Euro and minutes. The background variables used are annual net income, s_{inc}, and travel time of the reference trip, s_{time}. These are summarised in Table 6.2.

As a check on data we computed the Pearson Correlation Matrix for the eight variables described above. In general, all correlations were below 0.1 except for the correlation between Δx_{int} and Δx_w that was around 0.8 for all modes. The only other exception was a correlation of 0.15 between s_{inc} and s_{time} in the train segment.

6.3.2 Model formulation

All models are based on the generic model presented below. It includes the design variables travel cost, x_c, access/egress time, x_{ae}, headway, x_h, number of interchanges, x_{int}, in-vehicle time, x_{inv}, waiting time, x_w, and the background variables $s = (s_{inc}, s_{time})$. Since we have binary choices we only need to specify $\Delta U = U_1 - U_2$:

$$\begin{aligned}\Delta U \quad = \quad & \alpha + \beta_c f(s;\lambda^c,\lambda^{sc})\,\Delta x_c \\ & + \beta_{ae} f(s;\lambda^{ae},\lambda^{sc})\,\Delta x_{ae} \\ & + \beta_h f(s;\lambda^h,\lambda^{sc})\,\Delta x_h \\ & + \beta_{int} f(s;\lambda^{int},\lambda^{sc})\,\Delta x_{int} \\ & + \beta_{inv} f(s;\lambda^{inv},\lambda^{sc})\,\Delta x_{inv} \\ & + \beta_w f(s;\lambda^w,\lambda^{sc})\,\Delta x_w \\ & + \varepsilon,\end{aligned}$$

where α is an alternative specific constant[3], the λs are vectors of coefficients, the βs are either fixed in the binary logit (BL) models or lognormally distributed, i.e., $\beta = exp(\beta^0 + \Sigma^{\frac{1}{2}} u)$ with $u \sim N(0,I)$, and ε is logistically distributed.[4] The background variables are excluded when $f(s,\lambda) = 1$. Otherwise, they are included using a continuous interaction specification

$$f(s;\lambda^k,\lambda^{sc}) = \left(\frac{s_{inc}}{\bar{s}_{inc}}\right)^{\lambda^k_{inc}+\lambda^{sc}_{inc}} \left(\frac{s_{time}}{\bar{s}_{time}}\right)^{\lambda^k_{time}+\lambda^{sc}_{time}}, \tag{6.2}$$

where k signifies the six attributes, $\lambda^k = (\lambda^k_{inc}, \lambda^k_{time})$ and $\lambda^{sc} = (\lambda^{sc}_{inc}, \lambda^{sc}_{time})$ can be interpreted as the effect of time and income on the scale. For identification λ^{cost} is fixed to zero. This gives 12 λ coefficients. This causes the λs to be direct estimates of the elasticity of WTP_k with respect to income or time. Models without background variables correspond to $\lambda = 0$.

The lognormal distribution is chosen based on the fact that coefficients are naturally restricted to one half axis, see, e.g., McFadden (2000) and Jara-Diaz (2000). Even though wrong signs on these coefficients contradict theory, they may reasonably be interpreted as the effect of a point mass at zero. We will test the adequacy of the assumption of lognormal parameters. Thus the assumption is not as crucial as when left untested.

To develop a model that adequately captures taste correlation for each mode we estimate the six model types in Table 6.3. The intuition behind the models are the following. First, we esti-

[3]Even though alternative specific constants do not have the same motivation in unlabelled experiments as in labelled experiments it is still important to allow for these as they are sometimes significant.

[4]Here $exp()$ signifies that each marginal distribution is transformed by the exponential function.

Table 6.3: MODELS

BL: Binary logit specification based on attributes only
ML1: BL with independent lognormally distributed coefficients
ML2: ML1 with a full covariance matrix
ML3: ML1 with a partial covariance matrix
ML4: ML3 with background variables added
ML5: ML4 without insignificant parameters

mate a BL model as a base model to check signs and significance. Second, we estimate a mixed logit model (ML1) with independent lognormal distributions as this is what is normally done. Third, we estimate ML2 based on an unrestricted Choleski matrix and ML3 that reduces the number of parameters in the Choleski matrix without a significant loss in fit. In ML4 we add background variables to the specification. This is done to see if they can explain the correlation. The final model, ML5, is a reduction of ML4 where superfluous parameters are removed. We estimate the models with MSL in Biogeme (Bierlaire 2005) using 1000 MLHS draws (Hess et al. 2006) in the final models.

6.3.3 Train estimation

The results on model fit for the six train models are summarised in Table 6.4. Detailed estimation results for all models are given in Table 6.10. The MNL model has all coefficients significant with the expected negative signs. In a preliminary estimation the mixing on waiting time was insignificant therefore this was left out of ML1. The ML1 model has significant mixing for all attributes except waiting time. The two models with correlated coefficients (ML2 and ML3) are seen to outperform model ML1.[5] Although ML3 has better fit than ML2 it cannot reject it in a LR test as the χ^2_6 distribution has its 95%-quantile at 12.6.

Next all the 12 λ coefficients in Equation 6.2 were added. Based on tests the ML4 adds six parameters. These improve the fit very significantly. Four of the parameters explain unobserved heterogeneity in the random coefficients while the other two explain unobserved heterogeneity in the scale. Based on tests all λ^k_{inc} are restricted to equal one coefficient λ_{inc}. Likewise λ^{ae}_{time}, λ^{h}_{time}, and λ^{inv}_{time} are restricted to equal λ_{time}. In the models, λ_{time}, λ^{int}_{time} and λ_{inc} are significant and positive while λ^{w}_{time} is significant and negative. Both λ^{sc}_{time} and λ^{sc}_{inc} are significant and negative.

The most remarkable difference between ML3 and ML4 is a large drop in σ_{cost} from -2.19 to -1.29. This indicates that our background variables explain a large part of unobserved heterogeneity in the cost coefficient while other variables need to be introduced if we want a similar result for the other attributes. In the final ML5 model, we removed two insignificant Choleski factors.

[5] The parameterisation of the correlation matrix is not the standard Choleski factor, e.g., $cov(ln(\beta_c), ln(\beta_{ae})) = \rho_{ae}\sigma_c^2$. The corresponding Choleski factor would be $\rho_{ae}\sigma_c$.

Table 6.4: MODEL FITS FOR TRAIN MODELS

	Train models	
	DoF	LL
BL	7	-1968.7
ML1	12	-1665.5
ML2	27	-1624.0
ML3	21	-1627.1
ML4	27	-1557.7
ML5	25	-1559.7

Table 6.5: TESTS FOR TRAIN MODELS

	ML4 tests	
	DoF	LL
ML5	25	-1559.7
snp-cost	28	-1557.6
snp-ea	28	-1559.6
snp-hdw	28	-1555.6
snp-int	28	-1558.7
snp-inv	28	-1556.8

Table 6.6: MEAN WTP FOR TRAIN IN EURO PER HOUR

Model	BL	ML1	ML3	ML5
WTP_{ae}	16.1	25.5	23.8	16.8
	(13.0 - 19.9)	(17.8 - 36.3)	(18.8 - 29.8)	(12.8 - 22.0)
WTP_{h}	5.3	9.2	7.7	5.0
	(4.2 - 6.6)	(6.3 - 13.2)	(6.5 - 9.1)	(3.8 - 6.7)
WTP_{int}	4.9	8.0	9.4	8.3
	(3.2 - 6.9)	(5.6 - 11.5)	(6.3 - 13.5)	(4.7-17.2)
WTP_{inv}	12.5	21.8	15.2	11.3
	(10.2 - 15.3)	(15.7 - 30.2)	(12.4 - 18.6)	(9.3 - 13.7)
WTP_{w}	13.2	22.8	6.6	3.8
	(1.0 - 25.5)	(12.1 - 39.5)	(5.5 - 7.9)	(2.3-6.2)

The distributional assumptions were tested by the SNP method described in Section 6.2.2. Table 6.5 shows the results. The snp-cost model apply the SNP test to the cost coefficient, the snp-ea models apply it to the access-egress coefficient, etc. These models are all compared to ML5 in LR tests, e.g., the difference between ML5 and snp-cost is 2.1 therefore 4.2 is compared to the 95% quantile in the χ^2_3 distribution, which is 7.8. These tests show that only the distribution on headway is rejected. This might be removed by a different distribution but another reason could be that the remaining specification of headway is insufficient. Other research on the same data have shown that headway is weighted different for short and long headways. We also tested the support for the lognormal coefficients. This test could not reject the bound to be at zero.

Table 6.6 shows the mean WTP[6] evaluated for the models BL, ML1, ML3, and ML5 with 95%-confidence intervals of the mean estimate in parenthesis. We leave out ML2 and ML4 as they are similar to ML3 and ML5. The ML1 and ML3 models reject the low values of the BL model but ML5 resembles the BL evaluation and rejects the other mixed models. In addition, the two models allowing for correlation reject the high values of the ML1 model. For ML3 and ML5 the WTP for waiting is unreasonably low.

The mean correlation between $WTP_{ae}, WTP_h, WTP_{int}, WTP_{inv}$, and WTP_{wait} are seen in Figure 6.2 for ML3. It is found as the mean over 100 simulations of the correlation each with 10,000 draws. The correlations are all positive and significant except for $corr(WTP_{int}, WTP_w)$. The overall impression is that ML1 (not reported) overestimates the correlation between the WTPs. This said it is only in case of the four correlations concerning WTP_{wait} that the differences are significant.

$$\begin{bmatrix} 1 & . & . & . & . \\ 0.41 & 1 & . & . & . \\ 0.21 & 0.10 & 1 & . & . \\ 0.69 & 0.48 & 0.23 & 1 & . \\ 0.48 & 0.24 & 0.01 & 0.32 & 1 \end{bmatrix}$$

Figure 6.2: Mean correlation for train

6.3.4 Bus and city train estimation

Table 6.7 shows the estimation results for the twelve bus and city train models. Detailed estimation results for the models are given in Tables 6.11 and 6.12. The patterns resemble the train results. The only difference is that for the bus mode all six variances are significant in the mixed models. And that only eight additional Choleski parameters are needed for city train while eleven are needed for bus in the ML3 models with partial correlation compared to the ML1 models.

When background variables are added the two modes differ. For bus all λ^k_{inc} are reduced to one coefficient λ_{inc} while λ^{ae}_{time} and λ^h_{time} are reduced to λ_{time}, as are λ^{inv}_{time} and λ^{int}_{time} to λ^{in}_{time}. Only λ^{in}_{time} and λ_{inc} are significant and positive. Both λ^{sc}_{time} and λ^{sc}_{inc} are significant and negative. The drop in the unexplained variance of the cost coefficient is also present for the bus segment but less than for train. For the city train models we have the same six coefficients as for train. Here λ^{sc}_{time}, λ_{time}, and λ^{int}_{time} are significant while λ_{inc} is only significant in some models. The coefficients are positive except for λ^{sc}_{time}, which is negative. The drop in the unexplained variance of the cost coefficient is also present for the city train segment but much less than for train.

[6]The WTP is evaluated based on a simulation using the asymptotic distribution of the estimated coefficients.

The distributions in the ML5 models were tested. For bus only the distribution on cost is slightly rejected. For city train the distributions on access/egress and cost are rejected. The support test did not reject the bus model while it was significant for city train mainly due to a positive shift of the interchange coefficient.

Table 6.8 and 6.9 show the mean WTP evaluated for the different models with 95%-confidence intervals of the mean estimate in parenthesis. The mixed models reject the low values of the BL model. For bus, the models allowing for correlation each reject three of the ML1 values. For city train, the two models allowing for correlation do not reject the ML1 values except for the WTP for interchange.

The correlations between $WTP_{ae}, WTP_h, WTP_{int}, WTP_{inv}$, and WTP_{wait} are positive and significant for the ML3 models for both modes as seen in Figure 6.3.

$$\begin{bmatrix} 1 & . & . & . & . \\ 0.25 & 1 & . & . & . \\ 0.22 & 0.10 & 1 & . & . \\ 0.62 & 0.30 & 0.20 & 1 & . \\ 0.63 & 0.33 & 0.12 & 0.51 & 1 \end{bmatrix} \begin{bmatrix} 1 & . & . & . & . \\ 0.50 & 1 & . & . & . \\ 0.18 & 0.08 & 1 & . & . \\ 0.46 & 0.33 & 0.04 & 1 & . \\ 0.59 & 0.43 & 0.05 & 0.88 & 1 \end{bmatrix}$$

Figure 6.3: Mean correlation for bus(left) and city train

6.3.5 Discussion

Across all modes we find significant taste correlation between the mixed coefficients. All coefficients are positively correlated with the cost coefficient which indicates that a random scale is present in the data. This is also supported by the models with background variables where s_{time} has a highly significant effect on the scale. We have that the variance of the random error term is larger for longer journeys which is reasonable.

The correlation structures across modes show differences but also some common features. All correlations are significant, except one in the train matrix, and positive. In general, WTP_{ae} is most positively correlated with the other WTP measures while WTP_{int} is least correlated with the others.

The effect of background variables on the variance of the cost coefficient is largest for train but present for bus and city train as well. Since this variance induces correlation between WTP measures the results show that some unobserved correlation disappears when background variables are introduced into the utility specification. This leads to the thought that the introduction of correlation might not be worth the trouble. This concern is reasonable and addresses that as much variance and correlation as possible should be explained by explanatory variables. However, it should be noted that many applied models either do not have access to detailed background

Table 6.7: MODEL FITS FOR BUS AND CITY TRAIN

	Bus models		City train models	
	DoF	LL	DoF	LL
BL	7	-4248.1	7	-1375.6
ML1	13	-3729.4	12	-1210.4
ML2	28	-3679.0	27	-1194.8
ML3	24	-3683.3	20	-1198.9
ML4	30	-3634.0	25	-1185.6
ML5	26	-3635.8	22	-1187.1
snp-cost	29	-3631.4	25	-1180.0
snp-ea	29	-3632.0	25	-1183.1
snp-hdw	29	-3635.4	25	-1184.4
snp-int	29	-3633.6	25	-1183.8
snp-inv	29	-3634.5	25	-1184.4
snp-wait	29	-3635.3		

Table 6.8: MEAN WTP FOR BUS IN EURO PER HOUR

Model	BL	ML1	ML3	ML5
WTP_{ae}	4.8	10.1	8.0	7.7
	(4.1 - 5.5)	(7.7 - 13.2)	(6.3 - 10.1)	(6.1-9.7)
WTP_h	1.2	3.9	3.0	2.8
	(1.1 - 1.5)	(2.8 - 5.4)	(2.2 - 3.9)	(2.1-3.9)
WTP_{int}	0.7	1.4	1.2	1.3
	(0.5 - 0.9)	(1.0 - 1.9)	(0.8 - 1.9)	(0.9-1.8)
WTP_{inv}	2.9	5.9	4.6	4.2
	(2.4 - 3.5)	(4.4 - 7.7)	(3.7 - 5.9)	(3.3-5.6)
WTP_w	5.7	11.5	12.4	8.8
	(4.3 - 7.3)	(8.0 - 16.5)	(8.0 - 19.6)	(5.9-13.2)

Table 6.9: MEAN WTP FOR CITY TRAIN IN EURO PER HOUR

Model	BL	ML1	ML3	ML5
WTP_{ae}	5.1	10.8	10.9	10.3
	(4.3 - 6.2)	(7.6 - 15.1)	(8.2 - 14.1)	(6.8-16.0)
WTP_h	2.8	6.7	6.2	6.6
	(2.2 - 3.6)	(4.6 - 9.7)	(4.5 - 8.5)	(4.3-10.2)
WTP_{int}	0.5	1.1	0.8	0.9
	(0.3 - 0.8)	(0.6 - 1.9)	(0.6 - 1.1)	(0.5-1.8)
WTP_{inv}	3.9	7.5	8.8	8.0
	(2.9 - 5.1)	(5.1 - 10.9)	(5.4 - 15.0)	(4.9-13.1)
WTP_w	7.5	13.9	14.3	14.0
	(5.5 - 9.7)	(9.5 - 20.0)	(9.9 - 19.8)	(8.7-22.4)

variables or that many models work at an aggregated level where the variables will not explain the correlation to the same degree as here. Second, we observe that we still have unexplained correlation left after the introduction of background variables.

We have applied the SNP test to the final models. It rejects a few of the distributions. Most worrying is the rejection of the cost coefficient in the bus and city train case since this distribution affects all WTP measures. Ideally one should work with the specification until none of the tests reject the distributions. We have not done this as the focus of the paper is on correlations.

6.4 Summary and concluding remarks

One of the main applications of discrete choice models in transportation research is the estimation of VTT and other relevant WTP measures. The present paper has focused on the estimation of WTP measures for public transport. We have investigated correlation structures between WTP measures for different public transport modes in search of a structure that captures the essential correlation without adding too many coefficients. Finally, we tested the distributional assumptions in our models.

The results confirm a significant taste correlation and WTP correlation. For all modes we develop a mixed logit model with a partial correlation structure including background variables. These models show that both observed heterogeneity and correlation are present. So even though their explanatory powers partly overlap they are both necessary to explain choices. Ultimately it would be desirable to remove correlation through observed heterogeneity.

The main conclusion of the paper is that taste correlation should be taken into account in mixed logit estimation. Since the amount of correlation that is empirically identifiable depends on data size it may be difficult to investigate all possible correlation structures in small datasets. This study then can serve to indicate the correlation that is important in the context of public transport WTP measures.

It is clear from the results that the background variables used in this study lowers the unobserved heterogeneity in the cost coefficient. The question remains if other background variables can explain heterogeneity in other coefficients as well as the remaining correlation.

Acknowledgment
The author would like to thank Mogens Fosgerau for valuable comments on an earlier draft.

References

Bierlaire, M. (2005), 'An introduction to biogeme (version 1.4)', biogeme.epfl.ch. [89]

Daly, A. & Zachary, S. (1975), Commuters' values of time, Technical report, Local Government Operational Research Unit. Report No. T 55. [84]

Fosgerau, M. (2006), 'Investigating the distribution of the value of travel time savings', *Transportation Research Part B: Methodological* **40**(8), 688–707. [84]

Fosgerau, M. (2007), 'Using nonparametrics to specify a model to measure the value of travel time', *Transportation Research Part A* **41**, 842–856. [84, 86]

Fosgerau, M. & Bierlaire, M. (2007), 'A practical test for the choice of mixing distribution in discrete choice models', *Transportation Research Part B* **41 (7)**, 784–794. [84, 86]

Fosgerau, M., Hjorth, K. & Lyk-Jensen, S. V. (2006), *An integrated approach to the estimation of the value of travel time*, European Transport Conference, Strasbourg, 2006. [84, 87]

Hensher, D. A. (2001), 'The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications', *Transportation* **28**, 101–118. [84]

Hensher, D. A. & Greene, W. H. (2003), 'The mixed logit model: The state of practice', *Transportation* **30**, 133–176. [86]

Hess, S., Bierlaire, M. & Polak, J. W. (2005), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236. [84]

Hess, S., Train, K. & Polak, J. W. (2006), 'On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice', *Transportation Research Part B: Methodological* **40**(2), 147–163. [89]

Huber, J. & Train, K. E. (2001), 'On the similarity of classical and bayesian estimates of individual mean partworths.', *Marketing Letters* **12 (3)**, 259–269. [84]

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. [88]

Mabit, S. L., Caussade, S. & Hess., S. (2006), *Representation of taste heterogeneity in willingness to pay indicators using parameterisation in willingness to pay space*, European Transport Conference, Strasbourg, 2006. [84]

Mabit, S. L. & Nielsen, O. A. (2006), *The effect of correlated Value of Travel Time Savings in public transport assignment*, European Transport Conference, Strasbourg, 2006. [84]

Mackie, P. J., Jara-Diaz, S. R. & Fowkes, A. S. (2001), 'The value of travel time savings in evaluation', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 91–106. [84]

McFadden, D. (2000), *Disaggregate Behavioral Travel Demand's RUM Side*, IATBR Conference. [88]

McFadden, D. & Train, K. E. (2000), 'Mixed mnl models for discrete response', *Journal of Applied Econometrics* **15 (5)**, 447–70. [84]

Revelt, D. & Train, K. (1998), 'Mixed logit with repeated choice: Household's choices of appliance efficiency level', *The Review of Economics and Statistics* **80**(4). [84, 86]

Train, K. & Sonnier, G. (2005), Mixed logit with bounded distributions of correlated partworths, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 7, pp. 117–134. [84, 85]

Train, K. & Weeks, M. (2005), Discrete Choice Models in Preference Space and Willingness-to-Pay Space, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 1, pp. 1–16. [84]

Wardman, M. (2004), 'Public transport values of time', *Transport policy* **11**, 363–377. [83]

Appendix

Table 6.10: ESTIMATION RESULTS FOR TRAIN MODELS

	BL		ML1		ML2		ML3		ML4		ML5	
DoF	7		12		27		21		27		25	
Obs.	3455		3455		3455		3455		3455		3455	
Ind.	3455		523		523		523		523		523	
Final LL	**-1968.7**		**-1665.5**		**-1624.0**		**-1627.1**		**-1557.7**		**-1559.7**	
Adj. ρ^2	0.175		0.300		0.311		0.312		0.338		0.338	
Coefficients	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**
α	0.08	2.1	0.10	1.5	0.15	1.8	0.14	1.7	0.12	1.5	0.11	1.4
β_{cost}	-0.03	-12.0	-2.85	-25.1	-1.98	-10.2	-2.13	-10.7	-1.97	-11.3	-2.14	-16.5
β_{ae}	-0.01	-9.8	-2.71	-22.9	-2.07	-10.0	-2.20	-11.0	-2.59	-13.6	-2.76	-17.9
β_{hdw}	-0.01	-11.3	-4.21	-28.9	-3.59	-16.6	-3.69	-19.1	-3.93	-19.9	-4.12	-24.2
β_{int}	-0.51	-6.2	-0.64	-2.9	0.13	0.5	-0.08	-0.3	-0.20	-0.8	-0.37	-1.3
β_{inv}	-0.02	-11.0	-2.94	-23.1	-2.08	-10.0	-2.21	-11.5	-2.66	-14.1	-2.85	-19.8
β_{wait}	-0.02	-2.1			-2.50	-11.0	-2.81	-12.8	-3.83	-9.2	-3.98	-9.9
σ_{cost}			-2.70	-11.1	-2.38	-8.7	-2.19	-8.5	-1.29	-6.3	-1.20	-6.1
σ_{ae}			-0.75	-7.4	-0.94	-34.6	-0.97	-13.9	-0.82	-9.2	-0.80	-9.7
σ_{hdw}			-1.41	-14.0	1.28	38.1	1.27	69.4	1.05	25.5	1.04	17.1
σ_{int}			1.11	13.9	-1.25	-15.9	-1.44	-12.2	-1.10	-9.4	-0.93	-1.7
σ_{inv}			-1.42	-13.5	-0.59	-26.9	-0.56	-20.4	-0.52	-13.0	-0.53	-5.8
ρ_{ae}			-0.67	-8.2	0.45	7.7	0.46	7.2	0.27	2.4	0.21	1.4
ρ_{hdw}					0.65	15.6	0.58	11.6	0.72	13.5	0.68	9.0
ρ_{hdwae}					0.90	30.8	0.81	15.3	0.95	7.9	0.97	6.3
ρ_{int}					0.32	3.5	0.30	3.6	0.23	1.8		
ρ_{intae}					0.08	0.7						
ρ_{inthdw}					-0.12	-1.2						
ρ_{inv}					0.59	12.5	0.58	11.7	0.60	9.5	0.55	6.9
ρ_{invae}					0.46	27.0	0.52	10.7	0.70	5.9	0.72	5.6
ρ_{invhdw}					0.20	10.0	0.19	11.4	0.09	1.4		
ρ_{invint}					0.19	10.3						
ρ_{wait}					0.95	58.8	0.97	85.8	0.72	8.5	0.85	10.1
ρ_{waitae}					0.92	31.1	1.02	17.0	-0.89	-5.7	-1.00	-5.3
$\rho_{waithdw}$					-0.13	-16.9						
$\rho_{waitint}$					0.39	17.6						
$\rho_{waitinv}$					-0.27	-15.2						
λ_{inc}									0.58	8.5	0.60	7.5
λ_{time}									0.71	11.8	0.70	12.0
λ_{time}^{int}									1.31	6.8	1.34	7.6
λ_{time}^{w}									-1.26	-4.4	-1.15	-5.0
λ_{inc}^{sc}									-0.45	-3.1	-0.44	-3.1
λ_{time}^{sc}									-1.27	-9.3	-1.30	-10.2

Table 6.11: ESTIMATION RESULTS FOR BUS MODELS

	BL		ML1		ML2		ML3		ML4		ML5	
DoF	7		13		28		24		30		26	
Obs.	7751		7751		7751		7751		7751		7751	
Ind.	7751		1148		1148		1148		1148		1148	
Final LL	**-4248.1**		**-3729.4**		**-3679.0**		**-3683.3**		**-3634.0**		**-3635.8**	
Adj. ρ^2	0.208		0.303		0.310		0.310		0.318		0.318	
Coefficients	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**
α	0.10	4.0	0.15	3.8	0.18	4.0	0.18	3.9	0.15	3.6	0.15	3.6
β_{cost}	-0.09	-12.9	-1.58	-21.0	-1.16	-11.6	-1.19	-12.0	-1.16	-11.2	-1.21	-12.3
β_{ae}	-0.06	-17.6	-2.40	-30.8	-1.94	-19.0	-1.97	-18.7	-2.17	-19.8	-2.23	-21.7
β_{hdw}	-0.01	-18.6	-3.84	-39.5	-3.45	-29.9	-3.48	-30.4	-3.70	-30.8	-3.75	-32.8
β_{int}	-0.48	-8.5	-0.39	-2.6	-0.05	-0.3	-0.21	-1.1	-0.33	-2.1	-0.40	-2.7
β_{inv}	-0.03	-10.8	-2.81	-28.5	-2.32	-18.9	-2.40	-19.2	-2.59	-19.4	-2.67	-20.2
β_{wait}	-0.07	-8.8	-2.39	-13.3	-2.29	-14.7	-2.13	-13.3	-2.43	-15.1	-2.48	-15.2
σ_{cost}			-0.77	-9.0	1.60	11.6	1.61	11.7	1.28	10.7	1.24	13.1
σ_{ae}			1.21	15.9	0.94	32.4	0.92	26.3	0.81	15.1	0.82	14.6
σ_{hdw}			1.25	17.0	-1.40	-45.3	-1.37	-26.7	-1.31	-20.9	-1.26	-16.2
σ_{int}			-0.88	-6.8	-1.01	-12.9	-1.10	-6.8	-0.91	-7.7	-0.88	-5.8
σ_{inv}			-0.57	-5.7	0.56	12.3	0.67	10.1	0.45	3.7	0.44	2.8
σ_{wait}			-0.88	-3.6	0.88	25.9	0.69	8.8	0.76	8.8	0.78	10.1
ρ_{ae}					0.54	13.1	0.49	11.2	0.35	5.3	0.33	5.3
ρ_{hdw}					0.61	18.9	0.55	14.0	0.42	6.9	0.38	5.3
ρ_{hdwae}					0.45	13.2	0.44	14.5	0.46	6.3	0.51	5.3
ρ_{int}					0.31	5.1	0.35	5.0	0.14	1.5		
ρ_{intae}					0.18	2.0						
ρ_{inthdw}					0.26	7.5						
ρ_{inv}					0.58	14.4	0.57	14.1	0.40	4.0	0.37	4.1
ρ_{invae}					0.55	24.2	0.60	16.5	0.46	3.3	0.48	3.1
ρ_{invhdw}					0.17	8.6	0.13	4.3	0.12	1.4		
ρ_{invint}					0.32	9.2						
ρ_{wait}					0.58	12.2	0.43	7.9	0.33	4.7	0.35	2.9
ρ_{waitae}					1.15	19.9	1.11	7.1	1.07	11.7	1.05	11.6
$\rho_{waithdw}$					0.25	10.4	0.32	11.9	0.31	8.4	0.27	4.4
$\rho_{waitint}$					-0.14	-2.5						
$\rho_{waitinv}$					0.24	3.7	0.20	2.1	0.29	2.0		
λ_{inc}									0.32	4.7	0.33	5.0
λ_{time}									0.03	0.3		
λ_{time}^{in}									0.33	2.7	0.32	2.7
λ_{time}^{w}									-0.23	-1.7	-0.22	-1.5
λ_{inc}^{sc}									-0.25	-2.6	-0.25	-2.8
λ_{time}^{sc}									-0.81	-7.8	-0.80	-8.5

Table 6.12: ESTIMATION RESULTS FOR CITY TRAIN MODELS

	BL		ML1		ML2		ML3		ML4		ML5	
DoF	7		12		27		20		25		22	
Obs.	2731		2731		2731		2731		2731		2731	
Ind.	2731		401		401		401		401		401	
Final LL	**-1375.6**		**-1210.4**		**-1194.8**		**-1198.9**		**-1185.6**		**-1187.1**	
Adj. ρ^2	0.270		0.354		0.355		0.356		0.360		0.361	
Coefficients	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**	**estimate**	**t test**
α	0.09	2.1	0.15	2.0	0.20	2.3	0.19	2.3	0.17	2.1	0.15	1.9
β_{cost}	-0.07	-15.2	-1.54	-13.8	-1.04	-4.6	-1.13	-4.3	-2.02	-7.9	-2.20	-10.6
β_{ae}	-0.11	-9.8	-2.11	-23.1	-1.60	-7.5	-1.64	-6.7	-2.35	-8.9	-2.50	-13.2
β_{hdw}	-0.04	-9.9	-2.66	-18.2	-2.07	-8.5	-2.16	-7.7	-2.90	-10.1	-3.08	-13.9
β_{int}	-0.42	-4.2	-1.12	-2.7	-0.84	-2.6	-0.80	-1.9	-0.32	-0.9	-0.30	-0.6
β_{inv}	-0.06	-7.2	-2.55	-11.6	-2.03	-8.2	-2.08	-7.2	-2.77	-8.8	-2.94	-11.5
β_{wait}	-0.11	-7.4	-1.63	-14.0	-1.19	-5.7	-1.30	-5.1	-2.33	-8.4	-2.52	-10.4
σ_{cost}			-1.11	-11.0	-1.83	-6.5	-1.83	-5.2	-1.41	-6.2	-1.20	-7.2
σ_{ae}			-0.66	-8.4	-0.77	-15.6	-0.72	-14.4	-0.78	-12.3	-0.76	-7.3
σ_{hdw}			-0.77	-7.5	-0.85	-10.9	-0.87	-11.0	-0.79	-7.8	-0.82	-6.9
σ_{int}			-1.42	-8.6	-1.34	-21.8	-1.48	-12.6	-1.08	-9.2	-0.88	-4.1
σ_{inv}			0.74	3.1	0.72	4.6	0.70	4.6	0.63	4.7	0.63	3.4
ρ_{ae}					0.38	4.3	0.40	3.6	0.35	3.3	0.19	2.1
ρ_{hdw}					0.55	8.0	0.54	6.7	0.48	4.7	0.24	2.3
ρ_{hdwae}					0.57	8.1	0.42	3.8	0.68	8.2	0.66	4.5
ρ_{int}					0.85	17.7	0.83	17.6	0.73	10.7	0.52	2.8
ρ_{intae}					1.15	16.6	1.07	12.3	1.08	6.8	0.79	4.2
ρ_{inthdw}					0.05	1.5						
ρ_{inv}					0.38	3.5	0.32	2.5	0.19	1.2		
ρ_{invae}					-0.06	-0.4						
ρ_{invhdw}					0.40	1.6						
ρ_{invint}					-0.12	-1.6						
ρ_{wait}					0.42	4.9	0.37	3.2	0.28	2.3		
ρ_{waitae}					0.32	3.4						
$\rho_{waithdw}$					0.42	2.5						
$\rho_{waitint}$					0.12	1.7						
$\rho_{waitinv}$					0.69	9.5	0.68	3.9				
λ_{inc}									0.78	3.6	0.50	1.4
λ_{time}									0.12	1.6	0.08	0.8
λ_{time}^{int}									0.36	2.6	0.38	2.1
λ_{time}^{w}									1.63	6.8	1.49	1.7
λ_{inc}^{sc}									-0.09	-0.6		
λ_{time}^{sc}									-0.97	-5.1	-0.94	-4.9