UPPSALA
UNIVERSITET

Working Paper 2005:5
Department of Economics

Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables

Maria Vredin Johansson, Tobias Heldt and Per Johansson

Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2005:5
February 2005
ISSN 0284-2904

Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables

Maria Vredin Johansson, Tobias Heldt and Per Johansson

Papers in the Working Paper Series are published
on internet in PDF formats.
Download from http://www.nek.uu.se
or from S-WoPEC http://swopec.hhs.se/uunewp/

Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables*

Maria Vredin Johansson†, Tobias Heldt‡and Per Johansson§

Abstract

In a travel mode choice context, we use survey data to construct and test the significance of five individual specific latent variables - environmental preferences, safety, comfort, convenience and flexibility - postulated to be important for modal choice. Whereas the construction of the safety and environmental preference variables is based on behavioural indicator variables, the construction of the comfort, convenience and flexibility variables is based on attitudinal indicator variables. Our main findings are that the latent variables enriched discrete choice model outperforms the traditional discrete choice model and that the construct reliability of the "attitudinal" latent variables is higher than that of the "behavioural" latent variables. Important for the choice of travel mode are modal travel time and cost and the individual's preferences for flexibility and comfort as well as her environmental preferences.

Key words: Modal Choice, Latent Variable, Discrete Choice Model, Modal Safety

JEL Classifications: C35, R41

*The authors thank seminar participants at the Department of Economics at Uppsala university and the Swedish National Road and Transport Research Institute in Borlänge for useful suggestions. Financial support for Maria Vredin Johansson from The Swedish National Road Administration and The Wallenberg Foundation is gratefully acknowledged. The paper is also published as VTI notat 6A-2004. The usual disclaimer applies.

†Department of Economics, Uppsala university, P. O. Box 513, SE-751 20 Uppsala and Swedish National Road and Transport Research Institute, P. O. Box 760, 781 27 Borlänge, Sweden. Email: maria.vredin-johansson@nek.uu.se

‡Department of Economics, Uppsala university, P. O. Box 513, SE-751 20 Uppsala, Sweden and Department of Economics and Society, Dalarna university, SE-781 88 Borlänge, Sweden. Email: tobias.heldt@nek.uu.se

§Department of Economics, Uppsala university and Institute for Labour Market Policy Evaluation, P. O. Box 513, SE-751 20 Uppsala, Sweden. Email: per.johansson@ifau.uu.se

1 Introduction

In designing a socially desirable and environmentally sustainable transport system in line with people's preferences, transport planners must increase their understanding of the hierarchy of preferences that drive individuals' choice of transport. Understanding modal choice is important since it affects how efficiently we can travel, how much urban space is devoted to transport functions as well as the range of alternatives available to the traveller (Ortúzar and Willumsen, 1999, ch. 6).

In the empirical literature on travel mode choice, most choice models use modal attributes to explain choice. Individual specific variables are also often included to control for individual differences in preferences and unobservable modal attributes. This paper specifically addresses the problem of unobservable, or latent, preferences in modal choice models. The overriding purpose is to examine whether constructions of latent variables, mirroring the individual's preferences, are able to provide insights into the individual's decision making "black box" and, thus, to help to set priorities in governmental policy and decision making.

In recent attempts to gain insight into the decision making process of the individual, traditional choice models have been enriched with constructions of latent variables (McFadden, 1986; Morikawa and Sasaki, 1998; Ben-Akiva et al., 1999; Pendleton and Shonkwiler, 2001; Morikawa et al., 2002; Ashok et al., 2002). For example, Morikawa and Sasaki (1998) and Morikawa et al. (2002) include modal comfort and convenience in their analyses of modal choice. In their applications, the latent variables are measured and modelled through attitudes (attitudinal indicator variables) towards the chosen and an alternative travel mode.

In this paper, we model five latent variables and a maximum of three alternative travel modes. We use *individual* specific, not mode specific, latent variables to explain choice which means that we do not construct latent variables for nonchosen modes. Since the individual's opinion of nonchosen modes could be influenced by the individual's chosen mode, there is a risk of endogeneity when constructing latent variables for nonchosen modes.

Through a survey in a commuter context, data are collected on the respondent's modal choice and on the attitudinal and behavioural indicator variables that are used to construct environmental preferences and preferences for safety, flexibility, comfort and convenience.[1] The construction of the safety and environmental preference variables is based on behavioural indicator variables and the construction of the comfort, convenience and flexibility variables is based on attitudinal indicator variables. Thus, we are able to compare the explanatory power of constructions based on either type of indicator variables. Whereas inclusion of preferences for comfort,

[1] The questionnaire is available from the authors upon request.

convenience and flexibility needs little explanation, there are several reasons for our interest in safety and environmental preferences.

Preferences for safety are interesting mainly because reduced casualties is a major benefit of road infrastructure projects. In the cost benefit analyses (CBA) of the Swedish National Road Administration (SNRA), the value of increased safety represents roughly a third of all monetized benefits from infrastructural projects (Naturvårdsverket, 2003). The value of statistical life presently applied is derived from a Swedish contingent valuation (CV) study (SIKA, 2002; Persson et al., 1998).[2] Since CV studies can only uncover *stated* preferences, the resulting value can always be criticized for being hypothetical (e.g. Diamond and Hausman, 1995). Furthermore, several CV studies have revealed people's difficulties in understanding and valuing risk changes (Hammitt and Graham, 1999; Smith and Desvousges, 1987; Jones-Lee et al., 1985). Thus, the value of statistical life from CV surveys may be questioned. Since our survey is based on *revealed* preferences, we hope to shed light on whether preferences for safety are important in a real mode choice situation.

Proenvironmental preferences are of interest because there is an increased interest in incorporating environmental impacts in cost benefit analyses and the SNRA decision making. Because conversion to an environmentally sustainable transport system will, by necessity, affect peoples' choice of transport we find it interesting to gain increased knowledge about the importance of environmental aspects in peoples' choice of travel mode. Previous research has, however, shown little support for environmental criteria being of importance in travel mode choices (Daniels and Hensher, 2000; Vredin Johansson, 1999).

We estimate the individual's preferences in a latent variable model and include predictions of the latent variables in a discrete choice model for modal choice (multinomial probit with varying choice sets). On several accounts our "latent variables enriched" choice model outperforms a traditional choice model and provides insights into the importance of unobservable individual specific variables in modal choice. Whereas environmental preferences, comfort and flexibility are significant for modal choice, convenience and safety are insignificant.

In the following section we discuss attitudinal and behavioural indicator variables, in section three we describe the data and the data collection process, in section four, we present the model and, in section five, the estimation results are given. Finally, we discuss the findings in a concluding section.

[2] The value of statistical life presently applied is equal to SEK 17.5 million per road casualty. In CBAs, the fundamental value judgement is that human preferences should be sovereign (Pearce, 1998). Thus, to elicit human preferences for nonmarket goods, hypothetical markets, mimicing real markets, have to be constructed.

2 Attitudinal and Behavioural Indicator Variables

Research in the area of attitudes and behaviour (e.g. Oskamp et al., 1991; Ajzen and Fishbein, 1980, ch. 2) has shown that there may be a considerable discrepancy between attitudes and behaviour, especially when the attitudes are only distantly related to the behaviour in question. For example, predicting a single behaviour like paper recycling from a measure of an individual's general environmental attitudes may be very difficult. Research has, however, also shown that behaviours are often correlated so that an individual with, say, a environmental "personality trait"[3] performs more environmental behaviours than an individual without such a trait (Ajzen and Fishbein, 1980, ch. 7). We are, therefore, interested in exploring whether manifested behaviour in other areas of everyday life can help us better understand the driving forces behind modal choice. A hypothesis we test is whether someone who uses safety gear when driving, boating and biking[4] is more likely to choose a safer mode than a less safety orientated individual. Another hypothesis we test is whether someone who recycles glass, paper, batteries and metal is more likely to choose an environmentally friendly mode than someone who does not. Thus, we explore whether there exist patterns in behaviour that may be explained by different personality traits, like safety orientation and environmental orientation.

We apply two different methods when constructing the latent variables: for construction of the latent variables comfort, convenience and flexibility, we use *attitudinal* indicator variables[5] and for the safety and environmental preference variables, we use *behavioural* indicator variables. An advantage with behavioural indicator variables is that they are exogenous to the individual's modal choice. When latent variables are constructed from attitudinal indicator variables the individual's attitudes could be affected by the chosen mode (the individual "rationalizes" his/her choice) causing the latent variable construction to be endogenously determined.

The assumption of complementarity between recycling behaviours and the choice of an environmentally friendly mode could, of course, be challenged. Previous empirical work has given three tentative reasons why some environmental behaviours are performed while others are not. First, environmental behaviours are often only performed when they are easy to perform (Stern and Oskamp, 1987). When behaving environmentally is perceived as cumbersome, costly, inconvenient and ineffective or when oth-

[3] A personality trait is defined as a predispostion to perform a certain category of behaviours, e.g. altruistic behaviours (Ajzen and Fishbein, 1980, ch. 7). Behavioural categories, which can not be directly observed, are inferred from single behaviours that are assumed to be part of the general behavioural category.

[4] Using bike helmets when biking is not mandatory in Sweden.

[5] Attitudes are defined as the individual's subjective importance of the different items. We are aware that "attitudes" and "preferences" may be defined differently in psychology but hope that the definitions used here are clear enough and cause no semantic confusion.

ers, who are similarly expected to behave environmentally, are perceived as not doing so, individuals can not be expected to behave environmentally (Oskamp et al., 1991). For instance, Krantz Lindgren (2001) shows in interviews with "green" car drivers (individuals who drive regularly but recognize motorism's environmentally adverse effects)[6] that the perceived advantage of driving is large and that the perceived effect of reducing one's own car use is too small to ameliorate the environmental problems caused by motorism. Second, there might be *compensation* in environmental behaviours so that environmental behaviours are substitutes instead of complements.[7] Environmental compensation could result if people with environmental preferences net their feelings of guilt for using car with increased environmental behaviours in other areas of life, like composting and recycling. Some empirical support for this strand of reasoning can also be found in Krantz Lindgren (2001), where a compensation argument is used as an excuse for using car although awareness about the car's adverse environmental effects is high. Third, individuals may receive a "warm glow" (Andreoni, 1989) from recycling, implying that recycling and the choice of an environmentally friendly mode are altogether different behaviours.[8]

3 Data

A survey of commuters between Stockholm and Uppsala was conducted in September-October 2001. There are approximately 19,000 commutes between these cities situated 72 kilometers apart (Länsstyrelsen Uppsala län, 2002). The majority of the commuters (approximately 81 percent) travel between their home in Uppsala and their work in Stockholm. Essentially, there are only three different modes realistic for the commuter; car, train and bus. The distance is well served by both trains and buses. For instance, in the morning peak hours there are trains from Uppsala every 10 minutes and buses every 20 minutes. With train the commute takes about 40 minutes and costs SEK 36 (cheapest fare 2001)[9] and, with bus, the travel time

[6] For an average car with average work trip occupancy level (1.3 persons at peak hours, Naturvårdsverket, 1996), we believe it fair to say that work trip motorism has adverse environmental consequences. There could potentially be a few individuals in our sample whose work trips are performed in ethanol driven cars with full occupancy levels. In such cases car is likely to be a more environmental friendly alternative than a diesel driven low occupancy bus.

[7] The term"risk compensation" is well-known in transport research. For example, people tend to increase speed when the road is, or is perceived to be, safer and vice versa so that the overall perceived risk level is kept approximately constant.

[8] Warm glow is defined as a positive feeling of satisfaction from doing something desirable from society's perspective, similar to the moral satisfaction individuals receive from charitable contributions. Kahneman and Knetsch (1992) has coined the term "purchase of moral satisfaction" for warm glow generating behaviours.

[9] SEK 1 was approximately equal to € 0.11 in 2001 (www.riksbank.se).

is about an hour and costs SEK 29 (cheapest fare 2001). The rationale for choosing this particular commute was to minimize the likelihood of restrictions on the individuals' choice sets and, since Stockholm and Uppsala are situated in the most urbanized area of Sweden, there are few places where a transition between private to public modes could so easily be made.

The survey was conducted by Statistics Sweden (SCB). Altogether, 4,000 respondents, aged between 18 and 64 years, were contacted through a mail survey with two reminders. The sampling frame consisted of a matching of two registers, the total population register (actuality September 2001) and the employment register (actuality November 1999). Since the employment register was of less actuality, almost 21 percent of the individuals contacted were presently not commuting. Disregarding these cases, the overall response rate was 55 percent (number of responses, $n = 1,708$).

The sample consists to 67 percent of men. The average sample age is 43 years and the average sample household pretax monthly income is SEK 43,100. The proportion of respondents having house tenure is 49 percent and the proportion of respondents with children (18 years or younger), is 47 percent. In our sample, 900 respondents (54 percent) use car for commuting, whereas 516 respondents (31 percent) and 158 respondents (nine percent) use train and bus, respectively. The mean travel time is 58 minutes and the mean travel cost is SEK 73.[10] Most respondents (66 percent) do not have to change modes during the commute. Furthermore, 41 percent of the respondents had no alternative travel modes and are, therefore, excluded from the analysis. 39 percent of the respondents had one alternative travel mode and 21 percent had two alternative travel modes. Our analytic sample consists of $n = 811$. We find no significant differences between the total sample ($n = 1,708$) and the analytic sample ($n = 811$) regarding socioeconomic variables and commute characteristics. In the analytic sample, 50 percent use car, 38 percent train and 12 percent bus. 68 percent of the analytic sample have one alternative travel mode (i.e. a choice set size equal to two) and 32 percent have two alternative travel modes (i.e. a choice set size equal to three). For descriptive statistics, see Table A1 in Appendix A.

Table A2 in Appendix A gives descriptive statistics for the analytic sample stratified by chosen mode. There are significant differences in modal travel times where the travel time of train is significantly longer than that of bus which, in turn, has significantly longer travel time than car. Similarly we find that the travel cost of car is significantly higher than the travel

[10] In a few cases adjustment of the stated travel cost had to be made. If the mode was train and the stated travel cost was equal to or exceeded SEK 200 (admittedly an arbitrary value, but exceeding the one way fare between Stockholm and Uppsala) the travel cost was set to SEK 70. If the travel mode was bus and the stated travel cost was equal to or exceeding SEK 200, the travel cost was set to SEK 50. There are many possible reasons for respondents to give erroneous travel costs. In several cases it was quite obvious that the price of a monthly ticket had been stated.

cost of train and that the travel cost of train is significantly higher than that of bus. Furthermore, women choose car to a significantly lesser extent than they choose train and bus, respondents with children in the household choose car over train and bus and respondents with higher incomes choose car and train over bus. These findings seem natural, except for the travel time hierarchy of train and bus.

Apart from socioeconomic questions and questions regarding the respondent's habitual and alternative modes of travel and their respective times and costs, the survey contained behavioural and attitudinal questions intended to measure the latent variables postulated to be important for the individual's modal choice.[11] The behavioural questions addressed transport related safety behaviours, like questions about the use of safety gear like seat belts and bike helmets, and questions about the individual's consumer and recycling habits. The behavioural questions were scored on five-point scales from *never* to *always*. The attitudinal questions addressed issues related to modal comfort, convenience and flexibility. These were also scored on five-point scales from *not important at all* to *very important*.[12]

The attitudinal and behavioural questions resulted in ordinal data that were used in a latent variable, "multiple indicators, multiple causes" (MIMIC), model[13] to construct the latent variables postulated to be important for modal choice. Each of the latent variables in the MIMIC model is constructed from three to five observable ordinal indicator variables.

4 Model and Estimation

Traditionally, modal choice models include objective modal attributes, like travel time and travel cost. A real life complication is that individual heterogeneity, such as different preferences for e.g. safety, comfort, flexibility *et cetera*, also effects the choice of mode. In traditional choice models this heterogeneity is assumed, at least partially, to be controlled for by individual specific variables. Such blunt controls may potentially be improved upon by including measures of preferences directly in the choice model.

Whereas previous transport related applications has included modal comfort and convenience (Morikawa et al., 2002; Morikawa and Sasaki, 1998), we extend the list of included latent variables with environmental

[11] When designing the attitudinal and behavioural questions, we were influenced by Drottz-Sjöberg (1997) in which a series of questions about the frequency of proenvironmental behaviour was used as indicators of proenvironmental orientation. In the literature, there exist several other means for measuring proenvironmental orientation, e.g. the New Environmental Paradigm (NEP) scale based on attitudinal questions (Dunlap and van Liere, 1978; Dunlap et al., 2000).

[12] This type of scale is called a "semantic differential scale" (Ajzen and Fishbein, 1980, ch. 2).

[13] A MIMIC model is a confirmatory factor analytic model with explanatory variables (causes) (cf. Bollen, 1989, Ch. 8).

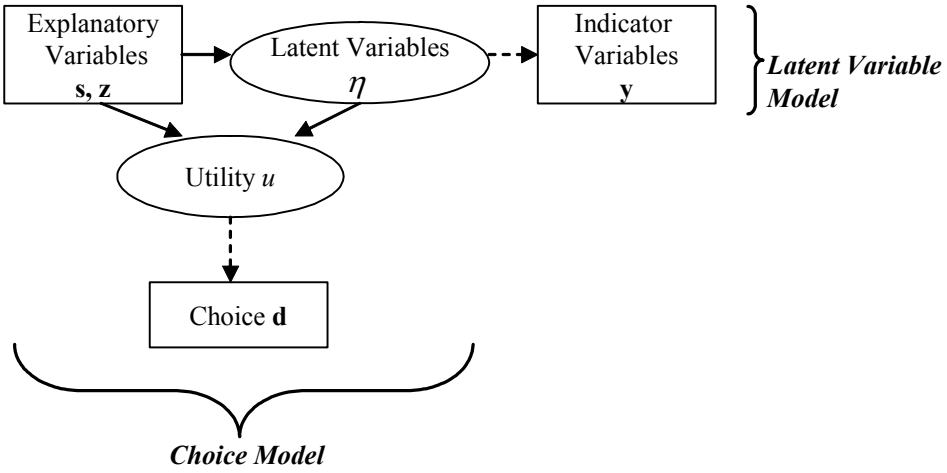

Figure 1: *Integrated Choice and Latent Variable Model (Ben-Akiva et al., 1999, p.195)*

preferences and individual preferences for flexibility and safety. Altogether we include five different latent variables in the choice model.

The framework for modelling and estimation, adapted from Morikawa et al. (2002), consists of a latent variable model (MIMIC) and a discrete choice model. Both these models consist of structural and measurement equations.[14] Figure 1, adapted from Ben-Akiva et al. (1999), gives a schematic picture of the modelling framework, where ellipses represent unobservable variables and rectangles observable variables. Dashed arrows represent measurement equations while solid arrows represent the structural equations. The latent variable model describes the relationships between the latent variables and their indicators and causes, while the discrete choice model explains modal choice. The complete, integrated choice and latent variable model explicitly incorporates latent variables in the choice process. The estimation is performed in two steps where the latent variable model is estimated first and then the discrete choice model is estimated. Although the estimation could be performed simultaneously, it is less cumbersome to estimate the model sequentially.

The specification of the "multiple indicator part" (MI) of the MIMIC model was assisted by exploratory and confirmatory factor analyses performed in the LISREL software (Jöreskog and Sörbom, 1993). The resulting latent variable model presented here is, thus, the result of a search process involving both the unconditional search of relations between indicators and latent variables as well as several direct tests of postulated relationships (for

[14] The measurement equations are also *structural,* in the sense that they describe structural relationships (Bollen, 1989, p.11).

exact model equations, see Appendix B).[15]

There are several ways of formulating discrete modal choice models, each emphasizing different aspects of modal choice (cf. Jara-Diaz and Videla, 1989; Train and McFadden, 1978; DeSerpa, 1971; Becker, 1965). The model used here is based on the fairly general disaggregate choice model by Jara-Diaz (1998) and Jara-Diaz and Videla (1989).

Generally, the conditional indirect utility u_{ij} for mode j ($j \in J$) for individual i is given by

$$u_{ij} = u(Y_i - p_j, \mathbf{w}_{ij}) + \nu_{ij},$$

where Y_i is the individual's income, p_j is the travel cost of mode j,[16] $\mathbf{w}_{ij}$ is modal attributes and individual characteristics and ν_{ij} is a random disturbance. Thus, the random utility is composed of a systematic term, which is a function of both latent and observable variables and a random disturbance, ν_{ij}.

In the empirical application we assume linear specifications of the conditional indirect utility function and of the latent variable functions. Suppressing individual indexation, the utility of travel mode j is

$$u_j = \mathbf{a}_j'\mathbf{s} + \mathbf{b}'\mathbf{z}_j + \mathbf{c}_j'\boldsymbol{\eta} + \nu_j, \tag{1}$$

where $\mathbf{z}_j$ is a vector of observable mode specific attributes (including travel cost), $\mathbf{s}$ a vector of observable individual specific attributes and $\boldsymbol{\eta}$ is a vector of individual specific latent variables. The structural relations to the latent variables are modelled as

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}. \tag{2}$$

The measurement equations are

$$d = \begin{cases} j \text{ if } u_j \geq u_k;\ \forall k \in J \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

and

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}. \tag{4}$$

In these equations, $\mathbf{y}$ is a vector of 20 observable indicator variables of $\boldsymbol{\eta}$, $\mathbf{x}$ is a vector of six exogenous observable variables that cause $\boldsymbol{\eta}$ ($\mathbf{x}$ may or may not be a part of $\mathbf{s}$), $\mathbf{a}_j$, $\mathbf{b}$ and $\mathbf{c}_j$ are vectors of unknown parameters to be estimated and $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are matrices of unknown parameters to be

[15] After some trial models, the full MIMIC model was also, at the outset, estimated in LISREL with the WLS estimator ($\chi^2[df = 274] = 2541.7$; $RMSEA = 0.07$; $NNFI = 0.94$; $CFI = 0.95$). Further information about the MIMIC model and the estimation method is given below and in Appendices B and C.

[16] The budget constraint is $Y_i = \mathbf{G} + p_j$, where $\mathbf{G}$ is a $K \times 1$ column vector of consumed continuous goods and where Y_i and p_j are normalised by the price of $\mathbf{G}$.

estimated and $\boldsymbol{\nu}= (\nu_1, ..., \nu_J)$, $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}$ are measurement errors independent of $\mathbf{s}$, $\mathbf{z}_j$ and $\mathbf{x}$ (see Appendix B).

Equations (1) and (3) form a discrete regression model, while equations (2) and (4) constitute the MIMIC model.

5 Results

5.1 The Latent Variable Model (MIMIC)

Based on the results from the factor analytic LISREL models (not reported), we postulate the existence of a safety personality trait and a environmental personality trait. While the safety personality trait is indicated by the respondent's propensity to use safety gear when biking, boating and driving ($y_6 - y_9$), the environmental personality trait is indicated by the respondent's composting and recycling habits ($y_1 - y_5$).[17]

The multiple indicator part of the model is a confirmatory factor analytical model specified such that we have five indicators for environmental preferences (η_{env}), four indicators for safety (η_{safe}), comfort (η_{comf}) and convenience (η_{conv}) and three for flexibility (η_{flex}). The multiple causes part of the model is given by

$$\eta_{li} = \boldsymbol{\gamma}_{l1}\text{WOMAN}_i + \boldsymbol{\gamma}_{l2}\text{AGE}_i + \boldsymbol{\gamma}_{l3}\text{INCOME}_i + \boldsymbol{\gamma}_{l4}\text{KID}_i +$$
$$+ \boldsymbol{\gamma}_{l5}\text{HOUSE}_i + \boldsymbol{\gamma}_{l6}\text{EDUCATION}_i + \boldsymbol{\zeta}_i,\ l = env, safe, comf, conv, flex$$

That is, the causes for the individual's latent preferences are the individual's age (years), income, gender (equal to one if woman), the presence of children in the household (equal to one if there are persons younger than 18 years in the household), education (in years) and house tenure.

Results from the first step maximum likelihood estimation are given in Tables 1 and 2.[18]

Evidently, all factor loadings in the measurement equations are positive and significant, which means that all indicators contribute to the construction of the latent preferences. Cronbach alpha values for the multiple indicator (MI) part of the MIMIC model are $\alpha_{\eta_{env}} = 0.73$, $\alpha_{\eta_{safe}} = 0.41$,

[17] All of these items are recycled without refunds and recycling is not mandatory. Collection points for recycling of glass and paper are abundant in Sweden and most grocery shops supply recycling containers for used nickel-cadmium batteries. Even though recycling is not mandatory, misapprehension or social norms seem to promote recycling (Paulsson et al., Dagens Nyheter 030210). As shown in section 2, we are aware of the fact that people may recycle for other than environmental reasons (and use bike helmets for other than safety reasons). Our hypothesis is just that environmentalism (safety) is one, among other, motives for these behaviours.

[18] For details on the estimation, see Appendices B and D.

$\alpha_{\eta_{comf}} = 0.76$, $\alpha_{\eta_{conv}} = 0.71$ and $\alpha_{\eta_{flex}} = 0.73$.[19] According to Nunnally (1978), values of 0.70 are acceptable. Thus, $\alpha_{\eta_{safe}}$ seem to be unacceptably low. This could, however, be the result of individual heterogeneity, i.e. something that we control for in the full MIMIC model.

Whereas Ben-Akiva et al. (1999) note that it sometimes can be difficult to find good causal variables for the latent variables, this does not seem to be the case here. Because the causal variables (as well as the indicator variables) are predictors of the latent variables, we retain the statistical significant (at the individual five percent level) causes and re-estimate the MIMIC model. These are the results presented in Tables 1 and 2.

We find that women are more environmentally inclined (η_{env}) than men. This result seems logical considering the indicator variables underlying the construction of η_{env}, i.e. composting kitchen refuse and recycling of glass, paper, batteries and metal, and the fact that women to a greater extent than men perform household recycling (Bennulf and Gilljam, 1991). The significance of age as a cause for η_{env} is also consistent with a previous finding (Drottz-Sjöberg, 1997). Furthermore, higher incomes are coupled with stronger preferences for convenience (η_{conv}), potentially reflecting the fact that the opportunity cost of time losses is higher at higher incomes. A little surprising is that preferences for safety (η_{safe}) decrease with income. However, this does *not* imply that safety is a non-normal good. Considering the indicators used to construct safety preferences, this merely shows that respondents with higher incomes use safety gear and adhere to speed limits to a lesser extent than respondents with lower incomes. Finally, considering the indicators used to construct flexibility (η_{flex}) it seems natural that respondents with children have stronger preferences for flexibility.

Table A2 in Appendix A gives the model predicted mean values of the latent variables ($\hat{\eta}_k$), stratified by the chosen mode. Train users have a significantly larger mean $\hat{\eta}_{env}$ value than car and bus users. Car users have a significantly lower mean value of $\hat{\eta}_{safe}$ and $\hat{\eta}_{conv}$ than train users. Car users have a significantly lower mean value of $\hat{\eta}_{comf}$ than bus users who have a significantly lower mean value of $\hat{\eta}_{comf}$ than train users. Furthermore, car users have a higher mean value of $\hat{\eta}_{flex}$ than bus and train users. Thus, the predicted values of the latent variables are in several cases significantly different for the different modes.

[19] Cronbach's alpha assesses the reliability in the measurement of an unobserved factor (Stata Reference Manual Release 7, 2001). The alpha values given here are based on standardised indicator variables.

5.2 The Discrete Choice Model

When it comes to the parameters of the choice model, we hypothesize that the generic parameters for time and cost will be negative so that the mode's likelihood of being chosen decreases when modal cost and time increase. We also postulate that the need to use own car in work (OWN) and having a car available for the worktrip (AVAIL) will increase the probability of choosing car over train and bus.

Apart from differing times and costs, the different modes also have different objective probabilities of death and injury as well as different objective energy consumption and emissions. It is, therefore, possible, with a few additional assumptions, to objectively tell which mode is the most (least) risky as well as the most (least) environmentally friendly. For the latent variables comfort, convenience and flexibility we are unable to give objective orderings of the modes, i.e. we can not on any *objective* grounds tell which is the most comfortable mode.

Considering environmental friendliness, Lenner (1993) has calculated emission equivalents per person and energy consumption equivalents per person for car, bus and train. Based on Lenner's results, we postulate that respondents with environmental preferences will choose train over bus and bus over car.[20]

On the relevant stretch of the motorway (the "E4") between Uppsala and Stockholm, car has considerable higher historical, objective, risks of death and injury compared to bus and train. Between January 1998 and January 2003, six people have been killed in car accidents and none in bus accidents (Swedish National Road Administration, pers. comm.). Despite the real number of deaths, the historical probabilities of being killed in car and bus accidents on this particular stretch of road are very small, especially considering the number of vehicles and people travelling there. Even though the *difference* in risk between bus and car is large, the baseline risks are still very small.[21] Thus, it is possible that the differences in modal safety are too small to be discernible.

For the parameters of the other latent variables ($c_{comf} - c_{flex}$), we base our hypotheses about the parameters on the indicator variables used to construct the individual preferences (see Appendix B). For comfort (c_{comf}), we postulate that individuals with preferences for comfort will choose train over bus and bus over car, since the comfort of train is larger that of bus and the comfort of bus is larger than that of car - proviso the indicator variables used for comfort. Furthermore, we hypothesize that car provides greater flexibility (c_{flex}) than bus and train (with no significant difference

[20] Lenner (1993) shows that, under Swedish conditions, electricity driven trains have lower energy consumption and produce less emissions than petrol/diesel driven buses.

[21] Anecdotal evidence based on personal communication with employees at the SNRA suggests that this stretch of the motorway is the safest in Sweden.

between the latter). We have no hypotheses about the convenience (c_{conv}) parameter.

In Table 3 the results from a multinomial probit models with and without latent variables (MNP_{LVE} and MNP_{REF}, respectively) are given.[22] A likelihood ratio test between the two models results in a test statistic of 255.3, which, with 10 degrees of freedom[23], strongly rejects the null hypothesis of the reference model without latent variables (MNP_{REF}).[24] Furthermore, the Akaike information criterion (AIC) - a means for comparing non-nested models - with number of parameters equal to the sum of the MIMIC and MNP_{LVE} parameters is smaller for the latent variables enriched model than for the reference model.

Below we first comment on the modal and socioeconomic variables, thereafter we provide a longer discussion on the latent variables, $\boldsymbol{\eta}$. Economizing on space, we will mainly comment on results that we find particularly interesting.

5.2.1 Modal and Socioeconomic Variables

Most of the common variables that are significant in the reference choice model are also significant in the latent variables enriched discrete choice model. However, there are a few exceptions. For instance, in the reference choice model, the presence of children in the household increases the likelihood of choosing car over bus. This relationship is insignificant in the latent variables enriched discrete choice model. Presumably the preferences captured by the variable KID in the reference choice model is better captured by the latent variables in the enriched discrete choice model.

Turning to the traditional modal choice variables, travel time and travel cost, we find that both are significant with the expected signs in both the reference choice model and in the latent variables enriched discrete choice model. The value of time (VOT) from the reference choice model is SEK 224, while the value of time in the latent variables enriched discrete choice model is SEK 175. The VOT is still very high compared to the official value of SEK 42 for private travels of less than 100 kilometers (SIKA, 2002), a

[22] Based on the estimates from the MIMIC model we formulate the predicted values of $\widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\Upsilon}}$ (the conditional covariance of $\boldsymbol{\eta}$) (see Appendix D for details). The discrete choice model is then estimated (employing these predicted values) using a multinomial probit ML estimator with varying choice sets. Since we include predicted values of $\boldsymbol{\eta}$ in place of the unknown values in the discrete choice model (see e.g. Murphy and Topel, 1985; Pagan, 1986), we correct the standard ML covariance matrix estimator (see equation (16) in Appendix D).

[23] This LR test is not strictly correct, since we neglect the indicators and causes used to construct the latent variables in the MIMIC model. However, it can still be taken as evidence of the benefit of using latent variables as determinants in the modal choice model.

[24] A less restrictive, random parameters probit, model in which the modal time and cost parameters were allowed to vary across the respondent was also estimated ($\ln \ell = -413.6$). The Akaike information criterion (AIC) for this model is equal to 1.07.

fact potentially explained by the higher incomes in our sample and/or by the fact that a number of respondents have to make one or more modal changes.[25]

5.2.2 Latent Variables

Turning to the latent variables, we find that two latent variables are significant at the five percent level ($c_{comf,CAR}, c_{flex,CAR}$) in the choice between car and bus. In the choice between train and bus one latent variable is significant at the five percent level ($c_{comf,TRAIN}$) while another is significant at the ten percent level ($c_{env,TRAIN}$).

Thus, preferences for comfort increase the likelihood of choosing bus over car ($c_{comf,CAR}$) and train over bus ($c_{comf,TRAIN}$). This is consistent with our hypothesis and is hardly surprising considering the indicator variables used to construct the comfort variable, i.e. the respondent's attitudes towards travelling in a non-noisy, environment with possibilities of resting, working and moving around. Preferences for flexibility increase the likelihood of choosing car over bus ($c_{flex,CAR}$) which also is consistent with our hypothesis and reasonable considering the indicators used to construct the flexibility variable: the need to shop, run errands or leave or collect children on the way to and from work. Consistent with our hypothesis, we find that environmental preferences ($c_{env,TRAIN}$) increase the likelihood of choosing train over bus.

Interesting to note is that safety, the latent variable with the lowest Cronbach alpha value in the factor analytic model, is insignificant in both the choice between car and bus and in the choice between train and bus. If the low Cronbach alpha value can not be explained by individual heterogeneity (as is done in the MIMIC model), it is possible that the indicators used are not well suited for capturing the latent variable we would like to model. For instance, if the safety variable constitutes a mixture of preferences for *personal* (security) and traffic safety, it is not surprising that the individuals' safety values (stratified by mode) are more similar than they would be if traffic safety and personal safety were independent constructions. This follows naturally from the assumption that the personal and traffic safety effects work in opposite directions, i.e. car is low on traffic safety and high on personal safety whereas public modes are high on traffic safety and low on personal safety.

[25] The average Swedish pre-tax household income in 2001 was SEK 23,506 per month (HE 20 SM 0201). The value of time during modal changes is twice the value of time when travelling (SEK 84) (SIKA, 2002). As a reference to these values, the average hourly earnings in the private sector (excluding overtime) was SEK 108 in october 2001 (AM SM 38 0201).

6 Conclusions

In a commute context, we use survey data to construct and test the significance of five individual specific latent variables postulated to be important for modal choice: environmental preferences, safety, comfort, convenience and flexibility.

On several accounts our "latent variables enriched" choice model outperforms a traditional choice model and provides insights into the importance of unobservable variables in modal choice. Our latent variables enriched choice model also turns out to be superior to a random parameters model where modal time and cost are allowed to vary.

In general, our results confirm that modal time and cost are significant for modal choice but also show that preferences for flexibility and comfort are very important.

According to expectation, environmental preferences increase the likelihood of choosing an environmentally friendly mode, train, over a less environmentally friendly mode, bus. Environmental preferences do, however, not matter in the choice between car and bus. If the government's goals for an environmentally sustainable and safe transport sector is to be achieved (Gov. Bill 1997/98:56), policy makers have to understand what prevents individuals from making environmentally sounder transport choices. Based on our results, we believe the policy challenge lies in reducing the welfare loss from behaving environmentally. Given the existing vehicle fleet, there are two possible ways (or a combination thereof) of doing this: either public modes become more "private" through, for instance, increased levels of flexibility or car becomes more expensive and cumbersome to use. In the future, fuel cell or other technology may reduce motorism's adverse environmental effects. Congestion problems are, however, likely to remain unless individuals have incentives to change from private to public modes.

Interesting to note is that preferences for safety are insignificant in the present modal choice model. This does not necessarily mean that safety considerations are unimportant in modal choice in general. Because the base line risks are very small in the commute under study here, the risks are perhaps too small to be discernible to the respondents. Furthermore, since the safety variable has low construct reliability, we may not fully measure what we intend to measure. An interesting issue for future research would be to investigate whether the form of safety (traffic safety, personal safety *et cetera*) preferences varies systematically with the trip characteristics, i.e. whether the trip is long or short, performed once or repeatedly, at work or leisure, within a city or in the countryside and so on. Should such differences be significant, the VOSL used in SNRA's cost benefit analyses should arguably be adjusted and differentiated accordingly. Differentiated VOSL which better capture individuals' preferences for safety is also desired from a policy perspective (SIKA, 2002). Thus, elicitation methods should

be designed to elicit individuals' preferences for different forms of safety under varying circumstances (e.g. trip length, trip purpose, initial risk level, geographical location).

Because the construct reliability of the attitudinal latent variables was on average higher than the construct reliability of the behavioural latent variables, a tentative conclusion is that preferences constructed from attitudinal indicators are to be preferred over preferences constructed from behavioural indicators. Because it is easier to find suitable attitudinal than behavioural indicator variables, attitudinal indicator variables may also be preferred on practical grounds. Nonetheless, an indisputable *advantage* of behavioural indicator variables over attitudinal is that they are *exogenous* to modal choice. Notwithstanding the mixed results of this pioneering survey, we still believe that a carefully constructed battery of behavioural questions have a great potential in capturing the individual's latent preferences. Future research will put our belief at test.

Table 1: The $\hat{\mathbf{\Lambda}}$ matrix of factor loadings (t-statistics in parentheses).

Indicator	η_{env}	η_{safe}	η_{comf}	η_{conv}	η_{flex}
Compost (y_1)	1				
Glass (y_2)	1.65 (15.1)				
Paper (y_3)	1.30 (14.1)				
Battery (y_4)	1.45 (14.8)				
Metal (y_5)	1.08 (13.7)				
Bikehelm (y_6)		1			
Speedlim (y_7)		1.67 (5.24)			
Lifejacket (y_8)		1.37 (7.29)			
Safebelt (y_9)		1.22 (6.42)			
Calmenv (y_{10})			1		
Rest (y_{11})			1.43 (20.7)		
Move (y_{12})			0.97 (16.9)		
Work (y_{13})			1.12 (18.3)		
Nowait (y_{14})				1	
Knowtime (y_{15})				1.79 (18.4)	
Novarian (y_{16})				1.53 (17.9)	
Noqueues (y_{17})				0.76 (12.0)	
Shop (y_{18})					1
Leavekid (y_{19})					2.88 (12.6)
Drivekid (y_{20})					2.63 (12.8)

Note: Indicator and variable definitions are given in Appendix B.

Table 2: The $\hat{\mathbf{\Gamma}}$ matrix (t-statistics in parentheses).

	WOMAN	AGE	INCOME	KID	HOUSE	EDUCATION
η_{env}	0.030 (2.36)	0.112 (7.89)	-	0.041 (3.25)	-	-
η_{safe}	0.114 (7.21)	0.083 (5.51)	-0.042 (-2.95)	0.031 (2.77)	-	0.042 (3.24)
η_{comf}	0.066 (4.40)	0.059 (3.68)	-	-	-0.099 (-6.20)	0.191 (11.40)
η_{conv}	0.102 (7.42)	-	0.053 (4.00)	-	-	-
η_{flex}	-	-0.065 (-7.76)	-	0.186 (11.80)	-	-

Note: Variable definitions are given in Appendix B.

Table 3: MNP estimations of the reference and latent variables enriched models.

	MNP_{REF}		MNP_{LVE}	
Variables/Parameters	Estimate	t-statistic	Estimate	t-statistic
TIME	-0.61	-10.66	-1.07	-6.82
COST	-0.23	-4.98	-0.51	-3.70
α_{CAR}	-2.06	-3.37	-6.77	-3.77
WOMAN_{CAR}	-0.20	-1.30	-0.69	-1.44
AGE_{CAR}	-0.00	-0.64	0.01	0.45
KID_{CAR}	0.34	2.18	0.07	0.15
EDUCATION_{CAR}	0.04	1.67	0.17	2.76
HOUSE_{CAR}	0.04	0.25	-0.01	-0.02
DCOM_{CAR}	0.05	0.57	0.15	0.72
OWN_{CAR}	1.07	3.63	2.67	3.27
AVAIL_{CAR}	1.83	8.11	4.07	5.12
$\eta_{env,CAR}$			-0.02	-0.03
$\eta_{safe,CAR}$			1.27	1.24
$\eta_{comf,CAR}$			-3.68	-5.91
$\eta_{conv,CAR}$			0.24	0.55
$\eta_{flex,CAR}$			2.46	2.84
α_{TRAIN}	-1.07	-1.76	-1.20	-0.93
WOMAN_{TRAIN}	-0.14	-0.89	-0.83	-2.20
AGE_{TRAIN}	-0.00	-0.28	-0.02	-1.13
KID_{TRAIN}	0.20	1.28	0.43	1.12
EDUCATION_{TRAIN}	0.11	4.60	0.16	3.18
HOUSE_{TRAIN}	-0.14	-0.82	-0.82	-1.47
DCOM_{TRAIN}	0.00	0.03	0.00	0.02
OWN_{TRAIN}	-0.13	-0.38	-0.37	-0.42
AVAIL_{TRAIN}	0.13	0.79	0.24	0.81
$\eta_{env,TRAIN}$			0.70	1.86
$\eta_{safe,TRAIN}$			0.72	0.76
$\eta_{comf,TRAIN}$			1.22	2.45
$\eta_{conv,TRAIN}$			0.60	1.60
$\eta_{flex,TRAIN}$			-0.16	-0.24
n	811		811	
$\ln \ell$	-453.49		-325.84	
LRI	0.33		0.52	
AIC	1.17		0.94	

Note: The likelihood ratio index, $LRI = 1 - (\ln \ell \ / \ \ln \ell_0)$. $\ln \ell_0$ is the log likelihood only with a constant term (Greene, 1993, ch.21). The Akaike information criterion, $AIC = -\frac{2}{n} \ln \ell + \frac{2p}{n}$, where p is the number of parameters and n the sample size (Amemiya, 1985).

References

Amemiya, T. (1985): *Advanced Econometrics.* Harvard University Press, Cambridge

Ajzen, I. and Fishbein, M. (1980): *Understanding Attitudes and Predicting Social Behavior.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey

AM SM 38 0201: *Konjunkturstatistik, löner för privat sektor under oktober 2001.* Statistics Sweden (2002). In Swedish with a summary in English.

Andreoni, J. (1989): Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, **97**, 1447-58

Ashok, K., Dillon, W. R. and Yuan, S. (2002): Extending Discrete Choice Models to Incorporate Attitudinal and Other Latent Variables. *Journal of Marketing Research*, **XXXIX**, 31-46

Becker, G. S. (1965): A Theory of the Allocation of Time, *The Economic Journal*, **75**, 493-517

Ben-Akiva, M., McFadden, D., Gärling, T., Gopinath, D., Walker, J., Bolduc, D., Börsch-Supan, A., Delquié, P., Larichev, O., Morikawa, T., Polydoropoulou, A., Rao, V. (1999): Extended Framework for Modeling Choice Behavior. *Marketing Letters,* **10**(3), 187-203

Bennulf, M. and Gilljam, M. (1991): *Snacka går ju - men vem handlar miljövänligt?* In Weibull, L. and Holmberg, S. (eds.) Åsikter om massmedier och samhälle. SOM undersökningen 1990, SOM Rapport 7. In Swedish.

Bollen, K. A. (1989): Structural Equations with Latent Variables. John Wiley and Sons, Inc.

Browne, M. W. (1984): Asymptotically Distribution Free Methods for the Analysis of Covariance Structures. British Journal of Mathematical and Statistical Psychology, **37**, 62-83

Browne, M. W. and Cudeck, R. (1993): Alternative Ways of Assessing Model Fit. In Bollen, K. A. and Long, J. S. (eds.): Testing Structural Equation Models. Newbury Park, CA, Sage, 136-162

Daniels, R. F. and Hensher, D. A. (2000): Valuation of Environmental Impacts of Transport Projects. The Challenge of Self-Interest Proximity. *Journal of Transport Economics and Policy*, **34**(2), 189-214

DeSerpa, A. C. (1971): A Theory of the Economics of Time. *The Economic Journal*, **81**, 828-846

Diamond, P. A. and Hausman, J. A. (1994): Contingent Valuation: Is Some Number better than No Number? *The Journal of Economic Perspectives*, **8**(4), 45-64

Drottz-Sjöberg, B-M. (1997): Attitudes, Values and Environmentally Adapted Products. Rhizikon Report No 30, Stockholm School of Economics

Dunlap, R. E. and van Liere, K. D. (1978): The New Environmental Paradigm. The Journal of Environmental Education, 9, 10-19

Dunlap, R. E., van Liere, K. D., Mertig, A. G. and Jones, R. E. (2000): Measuring Endorsement of the New Environmental Paradigm: A Revised NEP scale. *Journal of Social Issues*, **56**(3), 425-442

Gov. Bill 1997/98:56: *Transportpolitik för en hållbar utveckling.* Regeringens proposition 1997/98:56. In Swedish.

Greene, W. H. (1993): *Econometric Analysis.* 2nd edition, Macmillan, New York

Hammit, J. K. and Graham, J. D. (1999): Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? *Journal of Risk and Uncertainty*, **18**, 33-62

Hausman, J. A. and Wise, D. A. (1978): A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences. *Econometrica*, **46**(2), 403-426

HE 20 SM 0201: *Inkomster, skatter och bidrag 2000. Individ- och familjeuppgifter*, Statistics Sweden (2002). In Swedish with a summary in English.

Hu, L-T. and Bentler, P. M. (1995): Evaluating Model Fit. In Rick H. Hoyle (ed.): Structural Equation Modelling. Concepts, Issues and Applications. Sage Publications Inc., 76-99

Jara-Diaz, S. R. (1998): Time and Income in Travel Choice: Towards a Microeconomic Activity-Based Theoretical Framework. In Gärling, T., Laitila, T. and Westin, K. (eds.): *Theoretical Foundations in Travel Choice Modeling.* Pergamon

Jara-Diaz, S. R. and Videla, J. (1989): Detection of Income Effect in Mode Choice: Theory and Application. *Transportation Research*, **23B**, 393-400

Jones-Lee, M. W., Hammerton, M. and Philips, P.R. (1985): The Value of Safety: Results from a National Sample Survey. *Economic Journal*, **95**, 49-72

Jöreskog, K. G. and Sörbom, D. (1996): LISREL 8: User's Reference Guide, Scientific Software International, Inc., Chicago

Kahneman, D. and Knetsch, J. L. (1992): Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*, **22**, 57-70

Krantz Lindgren, P. (2001): *Att färdas som man lär? Om miljömedvetenhet och bilåkande.* Gidlunds förlag, Hedemora. In Swedish.

Lenner, M. (1993): *Energy Consumption and Exhaust Emissions Regarding Different Means and Modes of Transportation.* VTI Meddelande nr 718 (in Swedish with a summary in English)

Länsstyrelsen Uppsala län (2002): Fakta om Uppsala län. Mini uppslagsboken 2002-2003. In Swedish.

McFadden, D. (1986): The Choice Theory Approach to Market Research. *Marketing Science*, **5**(4), 275-297

Morikawa, T., Ben-Akiva, M. and McFadden, D. (2002): Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data. Econometric Models in Marketing Advances in Econometrics: A Research Annual, **16**, 27-53, Elsevier Science Ltd.

Morikawa, T. and Sasaki, K. (1998): Discrete Choice Models with Latent Variables Using Subjective Data. In Ortúzar, J de D., Hensher, D. A. and Jara-Diaz, S.: Travel Behaviour Research: Updating the State of Play, 435-455, Pergamon, Oxford

Murphy, K. and Topel, R. (1985): Estimation and Inference in Two Step Econometric Models. *Journal of Business and Economic Statistics*, **3**, 370-379.

Naturvårdsverket (1996): *Biff och Bil? Om Hushållens Miljöval*, Report 4542. In Swedish.

Naturvårdsverket (2003): *Värdering av tid, olyckor och miljö vid väginvesteringar. Kartläggning och modellbeskrivning.* Rapport 5270. In Swedish.

Nunnally, J. C. (1978): *Psychometric Theory.* Second edition. McGraw-Hill, New York

Ortúzar, J. de D. and Willumsen, L. G. (1999): Modelling Transport. Second edition. John Wiley and Sons, New York

Oskamp, S., Harrington, M. J., Edwards, T. C., Sherwood, D. L., Okuda, S. M. and Swanson D. C. (1991): Factors Influencing Household Recycling Behavior. *Environment and Behavior*, **23**(4), 494-519

Pagan, A. (1986): Two Stage and Related Estimators and Their Applications. *Review of Economic Studies*, **LIII**, 517-538

Paulsson, V., Norrby, S., Mellbin, K-G, Selberg, P., Löfstedt, L.: *Bättre för miljön att inte sopsortera.* Dagens Nyheter 030210. In Swedish.

Pearce, D. (1998): *Cost-Benefit Analysis and Environmental Policy.* Oxford Review of Economic Policy, **14**(4): 84-100

Pendleton, L. H. and Shonkwiler, J. S. (2001): Valuing Bundled Attributes: A Latent Characteristics Approach. *Land Economics*, **77**(1): 118-129

Persson, U., Nilsson, K., Hjalte, K. and Norinder, A. (1998): *Beräkning av Vägverkets riskvärden. En kombination av "contingent valuation"-skattningar och uppmätta hälsoförluster hos vägtrafikskadade personer behandlade vid fyra sjukhus.* The Swedish Institute of Health Economics (IHE), mimeo.

Satorra, A. and Bentler, P. M. (1988): *Scaling corrections for chi-square statistics in covariance structure analysis.* Proceedings of the Business and Economic Statistics Section of the American Statistical Association, **36**, 308-313.

SIKA (2002): *Översyn av samhällsekonomiska metoder och kalkylvärden på transportområdet - ASEK.* SIKA Rapport 2002:4. In Swedish.

Review of cost benefit calculation. Methods and valuations in the transport sector. SIKA.

Smith, V. K. and Desvousges, W. (1987): An Empirical Analysis of the Economic Value of Risk Changes. *Journal of Political Economy*, **95**, 89-114

Stata Reference Manual Release 7 (2001), Stata Statistical Software, Stata Press, College Station

Stern, P. C. and Oskamp, S. (1987): Managing Scarce Environmental Resources. In Stokols D. and I. Altman (eds.): Handbook of Environmental Psychology, vol. 2, 1043-1088. Wiley and Sons, New York

Train, K.A. and McFadden D. (1978): The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models. *Transportation Research*, **12**(5), 349-353

Vredin Johansson, M. (1999): Using Modal Perceptions to Determine Work Trip Travel Mode. In Vredin Johansson, M.: *Economics Without Markets: Four Papers on the Contingent Valuation and Stated Preference Methods*. Umeå Economic Studies No.517. Umeå University Press

West, S. G., Finch, J. F. and Curran, P. J. (1995): Structural Equation Models with Nonnormal Variables. Problems and Remedies. In Hoyle, R. H. (ed.): *Structural Equation Modelling. Concepts, Issues and Applications*. Sage

Appendix A: Descriptive Statistics

Table A1: Descriptive statistics: total and analytic sample.
Means ($\hat{\mu}$), standard errors (SE) and number of observations (n).

	Total sample			Analytic sample		
Variable	$\hat{\mu}$	SE	n	$\hat{\mu}$	SE	n
Gender (Woman=1)	0.33	0.01	1,706	0.35	0.02	811
Age (years)	43.20	0.26	1,706	42.77	0.37	811
Education (years)	14.52	0.09	1,697	14,91	0.13	811
Household income (SEK)	43,100	438	1,688	44,994	630	811
Children	0.47	0.01	1,678	0.49	0.02	811
Travel time (minutes)[a]	57.57	0.55	1,683	59.39	0.79	811
Travel cost (SEK)[a]	73.45	3.07	1,686	72.62	1.72	811
Commuting days per week	4.58	0.02	1,690	4.56	0.03	811
House tenure	0.49	0.01	1,683	0.45	0.02	811

[a] Mean travel time and cost are given for the chosen modes.

Table A2: Descriptive statistics: analytic sample stratified by mode. Means ($\hat{\mu}$), standard errors (SE) and number of observations (n).

	Car ($n = 406$)		Train ($n = 309$)		Bus ($n = 96$)	
Variable	$\hat{\mu}$	SE	$\hat{\mu}$	SE	$\hat{\mu}$	SE
Gender (Woman=1)	0.29	0.02	0.40	0.03	0.45	0.05
Age (years)	42.67	0.50	43.35	0.61	41.32	1.16
Education (years)	14.30	0.18	16.17	0.19	13.47	0.35
Household income (SEK)	46,022	903	46,545	997	35,651	1,576
Children	0.55	0.02	0.43	0.03	0.40	0.05
Travel time (minutes)	46.39	0.78	74.58	1.07	65.44	2.48
Travel cost (SEK)	95.27	2.93	55.84	0.72	30.81	1.45
Commuting days per week	4.58	0.04	4.53	0.06	4.51	0.08
House tenure	0.52	0.02	0.39	0.03	0.36	0.05
$\hat{\eta}_{env}$	-0.02	0.02	0.07	0.02	-0.07	0.05
$\hat{\eta}_{safe}$	-0.01	0.01	0.03	0.01	0.01	0.02
$\hat{\eta}_{comf}$	-0.32	0.02	0.40	0.02	0.11	0.03
$\hat{\eta}_{conv}$	-0.04	0.02	0.08	0.02	-0.05	0.05
$\hat{\eta}_{flex}$	0.07	0.02	-0.07	0.02	-0.08	0.03

Appendix B: Variables and Equations

Table B1: Latent and model variables.

Variable	Definition
η_{env}	Environmental preferences (latent variable).
η_{safe}	Safety (latent variable).
η_{comf}	Comfort (latent variable).
η_{conv}	Convenience (latent variable).
η_{flex}	Flexibility (latent variable).
WOMAN	Dummy variable for the gender of the respondent with one for female respondents.
AGE	The age of the respondent in years.
INCOME	The income of the respondent in SEK.
KID	Dummy variable with value one if the respondent's household includes children (persons younger than 19 years).
HOUSE	Dummy variable with value one if the respondent has house tenure.
TIME	Travel time in minutes.
COST	Travel cost in SEK.
EDUCATION	The respondent's education in years.
DCOM	Number of days the respondent commutes per week.
OWN	Dummy variable equal to one for the need to use own car in work at least one day a week.
AVAIL	Dummy variable equal to one for the availability of a car for worktrips at least one day a week.

Table B2: Indicator variables.

Indicator Variable		Definition
y_1	Compost	The respondent's habit of composting kitchen refuse.
y_2	Glass	The respondent's habit of recycling non deposit-refund glass bottles, jars et cetera.
y_3	Paper	The respondent's habit of recycling newspapers and paper.
y_4	Battery	The respondent's habit of recycling batteries.
y_5	Metal	The respondent's habit of recycling metal.
y_6	Bikehelm	The respondent's habit of wearing a bike helmet when riding a bike.
y_7	Speedlim	The respondent's habit of adhering to prevailing speedlimit when driving.
y_8	Lifejacket	The respondent's habit of using a life jacket when in smaller boats.
y_9	Safebelt	The respondent's habit of using safety belts in cars (also in the rear seats).
y_{10}	Calmenv	The respondent's appreciation of travelling in a calm, non-noisy environment.
y_{11}	Rest	The respondent's appreciation of being able to rest or read while travelling to/from work.
y_{12}	Move	The respondent's appreciation of being able to move around while travelling to/from work.
y_{13}	Work	The respondent's appreciation of being able to work while travelling to/from work.
y_{14}	Nowait	The respondent's appreciation of not having to wait for another travel mode while travelling to/from work.
y_{15}	Knowtime	The respondent's appreciation of knowing how long the daily travel time to/from work is.
y_{16}	Novarian	The respondent's appreciation of having little or no variation in her daily travel time to/from work.
y_{17}	Noqueues	The respondent's appreciation of avoiding queues and congestion while travelling to/from work.
y_{18}	Shop	The respondent's appreciation of being able to shop or run errands while travelling to/from work.
y_{19}	Leavekid	The respondent's appreciation of being able to leave/collect children at school or similar while travelling to/from work.
y_{20}	Drivekid	The respondent's appreciation of being able to give children a ride to their leisure time activities while travelling to/from work.

The first nine indicator variables ($y_1 - y_9$) are measured on five point category scales scored between *Never* and *Always*. All other indicator variables ($y_{10} - y_{20}$) are measured on five point semantic differential scales with the end-anchors *Not important at all* and *Very important*.

Equation (2):

$$\boldsymbol{\eta} = \Gamma \mathbf{x} + \boldsymbol{\zeta}$$

$$\begin{bmatrix} \eta_{env} \\ \eta_{safe} \\ \eta_{comf} \\ \eta_{conv} \\ \eta_{flex} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & \gamma_{26} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & \gamma_{46} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} & \gamma_{56} \end{bmatrix} \begin{bmatrix} \text{WOMAN} \\ \text{AGE} \\ \text{INCOME} \\ \text{KID} \\ \text{EDUCATION} \\ \text{HOUSE} \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{bmatrix}$$

Equation (4):

$$\mathbf{y} = \Lambda \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{19} \\ y_{20} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 \\ \lambda_{41} & 0 & 0 & 0 & 0 \\ \lambda_{51} & & & & \\ & 1 & & & \\ 0 & \lambda_{72} & 0 & 0 & 0 \\ 0 & \lambda_{82} & 0 & 0 & 0 \\ 0 & \lambda_{92} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_{113} & 0 & 0 \\ 0 & 0 & \lambda_{123} & 0 & 0 \\ 0 & 0 & \lambda_{133} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_{154} & 0 \\ 0 & 0 & 0 & \lambda_{164} & 0 \\ 0 & 0 & 0 & \lambda_{174} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_{195} \\ 0 & 0 & 0 & 0 & \lambda_{205} \end{bmatrix} \begin{bmatrix} \eta_{env} \\ \eta_{safe} \\ \eta_{comf} \\ \eta_{conv} \\ \eta_{flex} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \\ \varepsilon_{19} \\ \varepsilon_{20} \end{bmatrix}$$

Appendix C: MIMIC Model Estimation

Estimation of a structural equation latent variable model minimizes the difference between the sample covariance matrix, $\mathbf{S}$, and the covariance matrix, Σ. The elements of Σ are hypothesized to be a function of the parameter vector $\boldsymbol{\theta}$ so that $\Sigma = \Sigma(\boldsymbol{\theta})$. The parameters are estimated so that the discrepancy between $\mathbf{S}$ and the implied (by the parameters) covariance matrix $\Sigma(\hat{\boldsymbol{\theta}})$ is minimal. The discrepancy function, $F = F(\mathbf{S}, \Sigma(\boldsymbol{\theta}))$, measures the discrepancy between $\mathbf{S}$ and $\Sigma(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}$. $F_{\min}$ is the minimum value of the discrepancy function and equals zero only if $\mathbf{S} = \Sigma(\hat{\boldsymbol{\theta}})$. An indication of model fit is, therefore, given by the closeness of the $F_{\min}$ to zero (Browne and Cudeck, 1993). To test the model, the test statistic $T = (N-1)F_{\min}$ is calculated. If the model holds and is identified, T is asymptotically χ^2 distributed. This test statistic, T, is often referred to as "the χ^2 test" (Hu and Bentler, 1995)[26]. However, the χ^2 test statistic for overall model fit is vulnerable to sample size and departures from multivariate normality of the variables. If sample size is small, T might not be χ^2 distributed and, if sample size is large, even a trivial model misspecification results in model rejection. Therefore, there are several supplementary fit indices available for assessing model fit (Hu and Bentler, 1995; Browne and Cudeck, 1993; Jöreskog and Sörbom, 1993).

There are several different iterative estimation methods for structural equation models; unweighted least squares (ULS), generalized least squares (GLS), maximum likelihood (ML) and others (Jöreskog and Sörbom, 1993)[27]. The most commonly used estimators, ML and GLS, assume that the measured variables are continuous and multivariate normally distributed. However, if the data are highly non-normal, ML and GLS produce inflated χ^2 values and underestimate the standard errors of the parameters (West, Finch and Curran, 1995). An alternative estimator in the case of non-normality is the asymptotically distribution free weighted least squares estimator (ADF-WLS or WLS) developed by Browne (1984). Under normality, the WLS estimator is equivalent to ML but, under non-normality, it produces asymptotically unbiased estimates of the χ^2 test statistic and the standard errors. However, since the WLS estimator requires estimates of fourth-order moments[28], the WLS is of limited practical relevance when the sample size is small (West, Finch and Curran, 1995). When the variables are non-normal and the sample size is small[29], an alternative is to use ML with a correction

[26] The χ^2 test is in fact a "badness-of-fit" measure since small values correspond to good fit and large values correspond to bad fit (Jöreskog and Sörbom, 1993).

[27] Discrepancy functions for the different estimators are given in Bollen, 1989.

[28] The fourth-order moment, kurtosis, $m_4 = E\left[(x-\mu)^4\right]$.

[29] Depending on the model's complexity, a *small* sample can consists of 1,000-5,000 cases (West, Finch and Curran, 1995).

of the χ^2 statistic. This correction, the Satorra-Bentler correction (Satorra and Bentler, 1988), re-scales the normal-theory χ^2 statistic to account for non-normality (multivariate kurtosis) and holds regardless of the distribution of the variables (Hu and Bentler, 1995). The Satorra-Bentler correction also produces robust standard errors.

In our data, the majority of the variables clearly depart from normality since they consists of ordinal indicator variables. When testing the more continuous variables[30] for normality, all proved significant kurtosis and skew. Therefore, we estimate the model with WLS.

[30] These variables are; the AGE of the respondent (a truncated variable), the INCOME of the respondent (a categorised variable) and the respondent's EDUCATION in years.

Appendix D: Full Model Estimation

The structural equations consist of the four measurement and structural equations ((1)-(4)) given in Section 3 (repeated here for convenience)

$$u_j = \mathbf{a}_j'\mathbf{s} + \mathbf{b}'\mathbf{z}_j + \mathbf{c}_j'\boldsymbol{\eta} + \nu_j, \tag{5}$$

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}. \tag{6}$$

$$d = \begin{cases} j \text{ if } u_j \geq u_k;\ \forall k \in J \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

and

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}. \tag{8}$$

$\mathbf{y}$ is a $(q \times 1)$ vector of observable indicators of $\boldsymbol{\eta}$, $\mathbf{s}$ and $\mathbf{z}_j$ are vectors of observable exogenous variables ($\mathbf{z}_j$ is mode specific, while $\mathbf{s}$ is individual specific), $\boldsymbol{\eta}$ is a $(l \times 1)$ vector of individual latent variables, $\mathbf{x}$ is a $(k \times 1)$ vector of exogenous observable variables that cause $\boldsymbol{\eta}$ ($\mathbf{x}$ may or may not be a part of $\mathbf{s}$), $\mathbf{a}_j$, $\mathbf{b}$ and $\mathbf{c}_j$ are vectors of unknown parameters to be estimated and $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are, respectively, $(l \times k)$ and $(q \times l)$ matrices of unknown parameters to be estimated and $\boldsymbol{\nu}= (\nu_1, ..., \nu_J)$, $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}$ are measurement errors independent of $\mathbf{s}$, $\mathbf{z}_j$ and $\mathbf{x}$ and

$$E(\boldsymbol{\nu}\boldsymbol{\nu}') = \Xi,\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Theta,\ E(\boldsymbol{\zeta}\boldsymbol{\zeta}') = \boldsymbol{\Psi} \text{ and } E(\boldsymbol{\nu}\boldsymbol{\varepsilon}') = E(\boldsymbol{\nu}\boldsymbol{\zeta}') = E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}') = \mathbf{0}.$$

Let $\mathbf{u} = (u_1, ...u_J)'$. Then we can write the J utilities above as

$$\mathbf{u} = \mathbf{A}\mathbf{s} + \mathbf{Z}\mathbf{b} + \mathbf{C}\boldsymbol{\eta} + \boldsymbol{\nu}, \tag{9}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ ... \\ \mathbf{a}_J' \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ ... \\ \mathbf{z}_J' \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} \mathbf{c}_1' \\ \mathbf{c}_2' \\ ... \\ \mathbf{c}_J' \end{bmatrix}$$

For identification we let $\mathbf{a}_J = \mathbf{c}_J = \mathbf{0}$.

Assume for the moment that the vector $\mathbf{q} = (\mathbf{y}', \boldsymbol{\eta}', \mathbf{u}')'$ is multivariate normal with mean $\mathbf{m}_1$ and covariance matrix $\boldsymbol{\Omega}_1$, hence

$$\mathbf{m}_1 = \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x} \\ \boldsymbol{\Gamma}\mathbf{x} \\ \mathbf{A}\mathbf{s} + \mathbf{Z}\mathbf{b} + \mathbf{C}\boldsymbol{\Gamma}\mathbf{x} \end{bmatrix} \text{ and } \boldsymbol{\Omega}_1 = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Lambda}\boldsymbol{\Psi} & \boldsymbol{\Lambda}\boldsymbol{\Psi}\mathbf{C}' \\ \boldsymbol{\Psi}\boldsymbol{\Lambda}' & \boldsymbol{\Psi} & \boldsymbol{\Psi}\mathbf{C}' \\ \mathbf{C}\boldsymbol{\Psi}\boldsymbol{\Lambda}' & \mathbf{C}\boldsymbol{\Psi} & \Xi + \mathbf{C}\boldsymbol{\Psi}\mathbf{C}' \end{bmatrix},$$

where $\boldsymbol{\Omega}_{11} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \Theta$.

Let $\boldsymbol{\phi}$ be the vector of parameters given in $\mathbf{m}_1$ and $\boldsymbol{\Omega}_1$ above and let $\mathbf{d}$ be an indicator variable taking value one in row j if $d = j$ then the likelihood for $\boldsymbol{\phi}$, for a sample of n individuals, is

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{n} \mathbf{d}_i \ln \Pr(d_i = j, \mathbf{y}_i | \mathbf{x}_i, \mathbf{s}_i, \mathbf{z}_{ij}), \tag{10}$$

where

$$\Pr(d_i = 1, \mathbf{y}_i|\mathbf{x}_i, \mathbf{s}_i, \mathbf{z}_{ij}) = \int_{\eta_{env}} \dots \int_{\eta_{flex}} \left[\begin{array}{c} \int_{-\infty}^{\infty} \int_{-\infty}^{u_{i1}} .... \int_{-\infty}^{u_{i1}} f(u_{i1}, ..., u_{iJ}) du_{i1} \\ , ..., du_{iJ} \end{array} \right] d\boldsymbol{\eta}_i,$$

where f is a J variate normal density. Maximum likelihood estimation of $\boldsymbol{\phi}$ is difficult and we do not pursue this here, instead we use a two step estimator (see e.g. Murphy and Topel, 1985; Pagan, 1986) for the model parameters.

Conditional on $\mathbf{y}$ the distribution of the unobservables $\mathbf{q}_2 = (\boldsymbol{\eta}', \mathbf{u}')'$ is multivariate normal with mean $\mathbf{m}_2 = (E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x})', E(\mathbf{u}|\mathbf{y}, \mathbf{x}, \mathbf{Z}, \mathbf{s})')$ and covariance matrix

$$\boldsymbol{\Omega}_2 = \left[\begin{array}{cc} \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon}\mathbf{C}' \\ \mathbf{C}\boldsymbol{\Upsilon} & \boldsymbol{\Xi} + \mathbf{C}\boldsymbol{\Upsilon}\mathbf{C}' \end{array} \right].$$

where

$$\boldsymbol{\Upsilon} = \boldsymbol{\Psi} - \boldsymbol{\Psi}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_{11}^{-1}\Lambda\boldsymbol{\Psi}. \tag{11}$$

Here

$$E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x}) = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\Psi}\boldsymbol{\Lambda}'\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y} - \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x}) \tag{12}$$

and $E(\mathbf{u}|\mathbf{y}, \mathbf{x}, \mathbf{Z}, \mathbf{s}) = \mathbf{A}\mathbf{s} + \mathbf{Z}\mathbf{b} + \mathbf{C}E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x})$ and hence we can write the utility functions (9) above as

$$\mathbf{u} = \mathbf{A}\mathbf{s} + \mathbf{Z}\mathbf{b} + \mathbf{C}E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x}) + \boldsymbol{\vartheta} \tag{13}$$

where $\boldsymbol{\vartheta} = \mathbf{C}\mathbf{e} + \boldsymbol{\nu}$ and $\mathbf{e} = \boldsymbol{\eta} - E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x})$. Thus $\text{Var}(\boldsymbol{\vartheta}) = \boldsymbol{\Xi} + \mathbf{C}\boldsymbol{\Upsilon}\mathbf{C}'$**.**

Divide the parameter vector $\boldsymbol{\phi}$ into a parameter vector $\boldsymbol{\phi}_1$ for the MIMIC model (i.e. Equations (2) and (4)) and a parameter vector $\boldsymbol{\phi}_2$ describing the discrete choice model, thus $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2')'$. Now for a given value of $\boldsymbol{\phi}_1 = \widehat{\boldsymbol{\phi}}_1$ the log-likelihood for $\boldsymbol{\phi}_2$, under random sampling, is

$$\ell_2(\boldsymbol{\phi}_2; \widehat{\boldsymbol{\phi}}_1) = \sum_{i=1} \mathbf{d}_i \ln \Pr(d_i = j|\mathbf{x}_i, \mathbf{y}_i, \mathbf{s}_i, \mathbf{z}_{ij}), \tag{14}$$

where

$$\Pr(d_i = 1|\mathbf{x}_i, \mathbf{y}_i, \mathbf{s}_{i1}) = \int_{-\infty}^{\infty} \int_{-\infty}^{u_{i1}} .... \int_{-\infty}^{u_{i1}} f(u_{i1}, ..., u_{iJ}) du_{i1} ... du_{iJ}, \tag{15}$$

$$u_{ij} = \mathbf{a}_j'\mathbf{s} + \mathbf{b}'\mathbf{z}_j + \mathbf{c}_j' E(\boldsymbol{\eta}|\mathbf{y}, \mathbf{x}, \widehat{\boldsymbol{\phi}}_{\mathbf{1}}) + \vartheta_{ij}$$

Observe that $\vartheta_{ij} = \nu_{ij} + \mathbf{c}_j'\mathbf{e}$ and hence

$$E(\vartheta_{ij}^2) = \sigma_j^2 + \mathbf{c}_j'\boldsymbol{\Upsilon}\mathbf{c}_j \text{ and } E(\vartheta_{ij}\vartheta_{ik}) = \sigma_{jk} + \mathbf{c}_j'\boldsymbol{\Upsilon}\mathbf{c}_k, \ k \neq j,$$

where $\sigma_j^2 = E(\nu_{ij}^2)$ i.e. the j:*th* diagonal element of Ξ and $\sigma_{jk} = E(\nu_{ij}\nu_{ik})$.

If $\widehat{\boldsymbol{\phi}}_1$ is the maximum likelihood estimator of the MIMIC model then the maximum likelihood estimator based on maximizing (14) is (see e.g. Pagan,

1986) a consistent estimator asymptotically normal and with covariance matrix

$$\mathbf{V}_2(\widehat{\boldsymbol{\phi}}_2)= \mathbf{V}_2 + \mathbf{V}_2[\mathbf{H}\mathbf{V}_1\mathbf{H}' - \mathbf{L}\mathbf{V}_1\mathbf{H}' - \mathbf{H}\mathbf{V}_1\mathbf{L}']\mathbf{V}_2, \tag{16}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are the asymptotic covariance matrices of respectively, $\widehat{\boldsymbol{\phi}}_1$ and $\widehat{\boldsymbol{\phi}}_2$ conditional on $\widehat{\boldsymbol{\phi}}_1$, $\mathbf{H} = E((\partial\ell_2/\partial\boldsymbol{\phi}_2)(\partial\ell_2/\partial\boldsymbol{\phi}_1'))$ and $\mathbf{L} = E((\partial\ell_2/\partial\boldsymbol{\phi}_2)(\partial\ell_1/\partial\boldsymbol{\phi}_1'))$. Here

$$\ell_1 = -\frac{n}{2}\ln|\boldsymbol{\Omega}_{11}| - \sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x}_i)'\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_i - \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x}_i). \tag{17}$$

The asymptotic covariance matrix $\mathbf{V}_2(\widehat{\boldsymbol{\phi}}_2)$ is estimated using the outer product of the gradients at the maximum for $\mathbf{H}$ and $\mathbf{L}$ while $\mathbf{V}_1$ and $\mathbf{V}_2$ are estimated using the Hessian matrix at the maximum.

In our application individuals have varying choice sets. However, the maximum choice set is three (car, train and bus). Based on the ML estimates from the MIMIC model (i.e. maximization of equation 17) we formulate the predicted values of the conditional means (12) and variance (11), hence

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Gamma}}\mathbf{x} + \widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}_{11}^{-1}(\mathbf{y} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Gamma}}\mathbf{x}) \tag{18}$$

and

$$\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Psi}} - \widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}_{11}^{-1}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Psi}}. \tag{19}$$

The choice probability (15) for an individual with a choice set of 3 is now given as

$$\Pr(d_i = 1|\mathbf{x}_i, \mathbf{y}_i, \mathbf{s}_{i1}) = \int_{-\infty}^{\Delta q_{i21}} \int_{-\infty}^{\Delta q_{i31}} f(\Delta\vartheta_{i21}, \Delta\vartheta_{i31}, \boldsymbol{\Sigma}) d(\Delta\vartheta_{i21}) d(\Delta\vartheta_{i31})$$

where $\Delta\vartheta_{i21} = (\vartheta_{i2} - \vartheta_{i1})$, $\Delta\vartheta_{i31} = (\vartheta_{i3} - \vartheta_{i1})$, $\Delta q_{i21} = (\mathbf{a}_2' - \mathbf{a}_1')\mathbf{s} + \mathbf{b}'(\mathbf{z}_2 - \mathbf{z}_1) + (\mathbf{c}_2' - \mathbf{c}_1')\widehat{\boldsymbol{\eta}}$, $\Delta q_{i31} = (\mathbf{a}_3' - \mathbf{a}_1')\mathbf{s} + \mathbf{b}'(\mathbf{z}_3 - \mathbf{z}_1) + (\mathbf{c}_3' - \mathbf{c}_1')\widehat{\boldsymbol{\eta}}$ and $f(\cdot)$ is the bivariate normal density function with covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \varkappa_{11} & \varkappa_{12} \\ \varkappa_{12} & \varkappa_{22} \end{bmatrix},$$

where $\varkappa_{11} = \sigma_2^2 + \mathbf{c}_2'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_2 + \sigma_1^2 + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_1 - 2(\sigma_{12} + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_2)$, $\varkappa_{12} = \sigma_1^2 + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_1 - (\sigma_{13} + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_3) - (\sigma_{12} + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_2) + (\sigma_{23} + \mathbf{c}_3'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_3)$ and $\varkappa_{22} = \sigma_3^2 + \mathbf{c}_3'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_2 + \sigma_1^2 + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_1 - 2(\sigma_{13} + \mathbf{c}_1'\widehat{\boldsymbol{\Upsilon}}\mathbf{c}_3)$.

For identification we need to restrict the parameter space (see e.g. Hausman and Wise, 1978). We thus let $\boldsymbol{\Xi}$ be the identity matrix i.e. $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$ and $\sigma_{jk} = 0$ for all $j \neq k$.

WORKING PAPERS*
Editor: Nils Gottfries

2003:26 Niclas Berggren and Henrik Jordahl, Does Free Trade Really Reduce Growth? Further Testing Using the Economic Freedom Index. 19 pp.

2003:27 Eleni Savvidou, The Relationship Between Skilled Labor and Technical Change. 44 pp.

2003:28 Per Pettersson-Lidbom and Matz Dahlberg, An Empirical Approach for Evaluating Soft Budget Contraints. 31 pp.

2003:29 Nils Gottfries, Booms and Busts in EMU. 34 pp.

2004:1 Iida Häkkinen, Working while enrolled in a university: Does it pay? 37 pp.

2004:2 Matz Dahlberg, Eva Mörk and Hanna Ågren, Do Politicians' Preferences Correspond to those of the Voters? An Investigation of Political Representation. 34 pp.

2004:3 Lars Lindvall, Does Public Spending on Youths Affect Crime Rates? 40 pp.

2004:4 Thomas Aronsson and Sören Blomquist, Redistribution and Provision of Public Goods in an Economic Federation. 23 pp.

2004:5 Matias Eklöf and Daniel Hallberg, Private Alternatives and Early Retirement Programs. 30 pp.

2004:6 Bertil Holmlund, Sickness Absence and Search Unemployment. 38 pp.

2004:7 Magnus Lundin, Nils Gottfries and Tomas Lindström, Price and Investment Dynamics: An Empirical Analysis of Plant Level Data. 41 pp.

2004:8 Maria Vredin Johansson, Allocation and Ex Ante Cost Efficiency of a Swedish Subsidy for Environmental Sustainability: The Local Investment Program. 26 pp.

2004:9 Sören Blomquist and Vidar Christiansen, Taxation and Heterogeneous Preferences. 29 pp.

2004:10 Magnus Gustavsson, Changes in Educational Wage Premiums in Sweden: 1992-2001. 36 pp.

2004:11 Magnus Gustavsson, Trends in the Transitory Variance of Earnings: Evidence from Sweden 1960-1990 and a Comparison with the United States. 63 pp.

2004:12 Annika Alexius, Far Out on the Yield Curve. 41 pp.

* A list of papers in this series from earlier years will be sent on request by the department.

2004:13 Pär Österholm, Estimating the Relationship between Age Structure and GDP in the OECD Using Panel Cointegration Methods. 32 pp.

2004:14 Per-Anders Edin and Magnus Gustavsson, Time Out of Work and Skill Depreciation. 29 pp.

2004:15 Sören Blomquist and Luca Micheletto, Redistribution, In-Kind Transfers and Matching Grants when the Federal Government Lacks Information on Local Costs. 34 pp.

2004:16 Iida Häkkinen, Do University Entrance Exams Predict Academic Achievement? 38 pp.

2004:17 Mikael Carlsson, Investment and Uncertainty: A Theory-Based Empirical Approach. 27 pp.

2004:18 N. Anders Klevmarken, Towards an Applicable True Cost-of-Living Index that Incorporates Housing. 8 pp.

2004:19 Matz Dahlberg and Karin Edmark, Is there a "Race-to-the-Bottom" in the Setting of Welfare Benefit Levels? Evidence from a Policy Intervention. 34 pp.

2004:20 Pär Holmberg, Unique Supply Function Equilibrium with Capacity Constraints. 31 pp.

2005:1 Mikael Bengtsson, Niclas Berggren and Henrik Jordahl, Trust and Growth in the 1990s – A Robustness Analysis. 30 pp.

2005:2 Niclas Berggren and Henrik Jordahl, Free to Trust? Economic Freedom and Social Capital. 31 pp.

2005:3 Matz Dahlberg and Eva Mörk, Public Employment and the Double Role of Bureaucrats. 26 pp.

2005:4 Matz Dahlberg and Douglas Lundin, Antidepressants and the Suicide Rate: Is There Really a Connection? 31 pp.

2005:5 Maria Vredin Johansson, Tobias Heldt and Per Johansson, Latent Variables in a Travel Mode Choice Model: Attitudinal and Behavioural Indicator Variables. 31 pp.

See also working papers published by the Office of Labour Market Policy Evaluation http://www.ifau.se/

ISSN 0284-2904