

University of Wollongong  
**Research Online**

---

Faculty of Engineering and Information  
Sciences - Papers: Part A

Faculty of Engineering and Information  
Sciences

---

1-1-2013

**Analysing the merit of latent variables over traditional objective attributes  
for traveller mode choice using RPL model**

Ahm Mehbub Anwar  
*University of Wollongong*, [amanwar@uow.edu.au](mailto:amanwar@uow.edu.au)


A Kiet Tieu  
*University of Wollongong*, [ktieu@uow.edu.au](mailto:ktieu@uow.edu.au)

Peter Gibson  
*University of Wollongong*, [peterg@uow.edu.au](mailto:peterg@uow.edu.au)

Khin Than Win  
*University of Wollongong*, [win@uow.edu.au](mailto:win@uow.edu.au)

Matthew Berryman  
*University of Wollongong*, [mberryma@uow.edu.au](mailto:mberryma@uow.edu.au)

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

 Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

---

**Recommended Citation**

Anwar, Ahm Mehbub; Tieu, A Kiet; Gibson, Peter; Win, Khin Than; and Berryman, Matthew, "Analysing the merit of latent variables over traditional objective attributes for traveller mode choice using RPL model" (2013). *Faculty of Engineering and Information Sciences - Papers: Part A*. 2547.  
<https://ro.uow.edu.au/eispapers/2547>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Analysing the merit of latent variables over traditional objective attributes for traveller mode choice using RPL model

## Abstract

In real life, the attributes that influence individual choice may be complex. The traditional objective attributes can be incorporated easily into choice models. However, there are also latent preference heterogeneities that are often overlooked by the traditional thinkers of transport planners. This paper deals with this issue firstly by testing the adequacy of objective attributes representing latent variables (LVs). It then quantifies the effect of LVs over objective attributes on traveller mode choice using the random parameter logit (RPL) model. Understanding these attributes is essential if transport agencies are to understand traveller behaviour when determining effective transport policies. This paper emphasises travellers' LVs along with objective attributes during the mode choice process as a method by which the utility of the traveller can be maximised. Thus the issue of utility function is raised and evaluated using a discrete choice experiment, i.e. RPL model. An empirical study was carried out in the context of traveller behaviour in the Sydney Statistical Division (SSD). We consider six LVs and thirteen objective attributes to analyse the importance/merits of LVs over objective attributes in traveller mode choice. The results show that indicators of LVs and traveller choice attributes are found to be significant, while objective attributes show a very minimal (0% to 10% on average) capacity to reflect LVs in traveller choice processes. LVs are found to be more influential than objective attributes on the mode choice made by travellers and our results also show that hybrid RPL is superior to traditional RPL models that ignore the effect of LVs. Our results support the contention that latent factors are important in traveller mode choice in ways that are relevant to transportation planners and policy-makers. Although possibly not directly susceptible to policy intervention, a better understanding of these relationships is useful for decision makers and transportation planners when designing and developing sustainable transportation policies or projects for the city dwellers.

## Disciplines

Engineering | Science and Technology Studies

## Publication Details

Anwar, A. H. M. Mehbub., Tieu, K., Gibson, P., Win, K. Than. & Berryman, M. J. (2013). Analysing the merit of latent variables over traditional objective attributes for traveller mode choice using RPL model. International Choice Modelling Conference (pp. 1-23). United Kingdom: ICMC.

## **Analysing the Merit of Latent Variables over Traditional Objective Attributes for Traveller Mode Choice Using RPL Model**

A.H.M. Mehbub Anwar<sup>1,\*</sup> Kiet Tieu<sup>1,\*</sup> Peter Gibson<sup>1,♦</sup>

Khin Than Win<sup>2,•</sup> Matthew J. Berryman<sup>3,▲</sup>

<sup>1</sup> Faculty of Engineering

<sup>2</sup> Faculty of Informatics

<sup>3</sup> SMART Infrastructure Facility

University of Wollongong  
Wollongong NSW 2522, Australia

### **Abstract**

In real life, the attributes that influence individual choice may be complex. The traditional objective attributes can be incorporated easily into choice models. However, there are also latent preference heterogeneities that are often overlooked by the traditional thinkers of transport planners. This paper deals with this issue firstly by testing the adequacy of objective attributes representing latent variables (LVs). It then quantifies the effect of LVs over objective attributes on traveller mode choice using the random parameter logit (RPL) model. Understanding these attributes is essential if transport agencies are to understand traveller behaviour when determining effective transport policies. This paper emphasises travellers' LVs along with objective attributes during the mode choice process as a method by which the utility of the traveller can be maximised. Thus the issue of utility function is raised and evaluated using a discrete choice experiment, i.e. RPL model. An empirical study was carried out in the context of traveller behaviour in the Sydney Statistical Division (SSD). We consider six LVs and thirteen objective attributes to analyse the importance/merits of LVs over objective attributes in traveller mode choice. The results show that indicators of LVs and traveller choice attributes are found to be significant, while objective attributes show a very minimal (0% to 10% on average) capacity to reflect LVs in traveller choice processes. LVs are found to be more influential than objective attributes on the mode choice made by travellers and our results also show that hybrid RPL is superior to traditional RPL models that ignore the effect of LVs. Our results support the contention that latent factors are important in traveller mode choice in ways that are relevant to transportation planners and policy-makers. Although possibly not directly susceptible to policy intervention, a better understanding of these relationships is useful for decision makers and transportation planners when designing and developing sustainable transportation policies or projects for the city dwellers.

**Keywords:** *Traditional objective attributes, latent variables, hybrid RPL model, mode choice, and traveller.*

---

\* Corresponding author, T: +61 2 42515 0727, Email: ahmma324@uowmail.edu.au

♦ T: +61 2 4221 3061, Email: ktieu@uow.edu.au

♦ T: +61 2 4221 5968, Email: peterg@uow.edu.au

• T: +61 2 4221 4142, F: +61 2 4221 4045, Email: win@uow.edu.au

▲ T: +61 2 4239 2330, F: +61 2 4298 1489; Email: mberryman@uow.edu.au

## 1 Introduction

Traveller preference heterogeneity is well understood by transportation planners. Some people value travel time savings more than others; some pay more attention to the environmental effects of transport options; some are more sensitive to social status; some prefer more convenient options than faster options and so on. Most transportation departments over the world set a goal to provide suitable options for travellers, which demands a good understanding of the heterogeneity in traveller preferences in order to serve the diverse needs of each individual. Usually, choice models in transportation research include traditional objective choice attributes (*hereafter referred to as objective attributes*), such as travel time, travel cost etc. as well as socio-economic characteristics (e.g. income, age etc.). Real life is complex with individual preference heterogeneity, for example, comfort, convenience, flexibility etc. also influences the choice process significantly (Anwar et al., in press). In traditional choice models, this heterogeneity is assumed, at least partially, to be controlled for by individual specific variables (Johansson et al., 2006).

However, the latent factors people consider in making their travel decisions are more salient than travel time and cost alone. Furthermore, people's travel preferences are much more complex than their socio-economic and trip characteristics (Anwar et al., 2011). There is strong evidence in extant research that recent developments including latent variables, latent classes, structural equation modelling (SEM) and integrated frameworks have advanced ways to examine a wider array of variables that might influence travel behaviour. This framework explicitly treats psychological factors, such as attitudes and perceptions, using psychometric indicators instead of objective attributes (Johansson et al., 2006; Ben-Akiva et al., 1994; Gopinath, 1995; Walker and Ben-Akiva, 2002; Ashok, 2002; Temme et al., 2008), but the extent to which objective attributes explain latent factors has not been evaluated to date.

A discrete choice analysis is the most popular method for investigating the nature of modal choice decision-making processes amongst many modes (Train, 2009). The economic theories of random utility are the fundamental concept of this analysis and it assumes that a traveller chooses the mode with the highest utility under a rational circumstance (Bhat, 1998; Bolduc, 1999; Train, 2009). Though discrete choice analysis was introduced to analyse transport related problems, it has been applied in various fields for the last two decades (Bolduc, 1999). These studies have focused on analysing the behaviour of the decision-making process, such as modal choice (Bhat, 2000; Bolduc, 1999; Cohen and Harris, 1998; Commins and Nolan, 2011; Dissanayake and Morikawa, 2005; Ewing et al., 2004; Habib, 2012; Train, 1980), choice of car type (Choo and Mokhtarian, 2004; McCarthy, 1996), tourists' mode choice (Can, 2013; Jialing et al., 2013; Fesenmaier, 1988; Nicolau and Mas, 2006; Train, 1998), traveller latent perspective (Daly et al., 2012; Fleischer et al., 2012), survey quality to perceptual and attitudinal questions (Hess and Stathopoulos, 2011), and heterogeneous decision rules (Hess et al., 2011).

Observed heterogeneity can be incorporated easily into the models by introducing individual socio-economic characteristics and integrating a level of service attributes as well as trip characteristics. In addition, there are also unobserved heterogeneities of individuals that are often overlooked by the traditional transportation modellers, because it is assumed that the latent aspects are sufficiently represented by the objective attributes. There is a clear need to test whether it is true or not. The commonly used choice attributes are defined here as objective attributes which are defined in this paper as follows:

- 1) *Level of services (LOS)*: travel time, travel cost and waiting time;
- 2) *Socio-economic characteristics (SEC)*: age, income, family size, gender, car ownership, number of children (age 0-14 years), and number of full time workers in household; and
- 3) *Trip characteristics (TC)*: trip rate per day, distance travelled per trip and trip purpose.

In general, both individual specific attributes, such as income, age and gender, and mode specific attributes, such as travel time and travel cost, are analysed as functions of travel mode choice models. However, in the last decade this level of analysis has been criticised and some researchers have recommended the need to integrate latent variables (LVs) into choice models (McFadden 1986; Ashok et al., 2002; Morikawa et al., 2002; Anwar et.al, 2011). Latent factors are the true and adequate representation of traveller behaviour that helps to acquire valuable insight in the decision-making process of the individual (Johansson et al., 2006). Other research indicates that more intangible constructs, e.g. values, nature of lifestyle, and personality traits, might also have an effect on travel mode choice (Choo and Mokhtarian 2004; Nordlund and Garwill 2003; Collins and Chambers 2005). Thus, it is proposed that traveller preferences (e.g. latent factors) affect the mode choice process ( $P_1$ ):

***Proposition 1 ( $P_1$ ): Traveller preference heterogeneity structure mode choice***

Socioeconomic characteristics, such as age, gender, and income, are the dominant features which orient people in their lifestyle in terms of their choice, preferences, and expectations. In choice models with latent variables, Ben-Akiva et al. (2002b) observed that preferences (e.g. flexibility) of individuals are affected by socioeconomic characteristics. Johansson et al. (2006) also concluded that demographic variables impacted on preferences of flexibility and comfort. We accordingly propose the following proposition ( $P_2$ ):

***Proposition 2 ( $P_2$ ): Socioeconomic characteristics shape travellers' preferences***

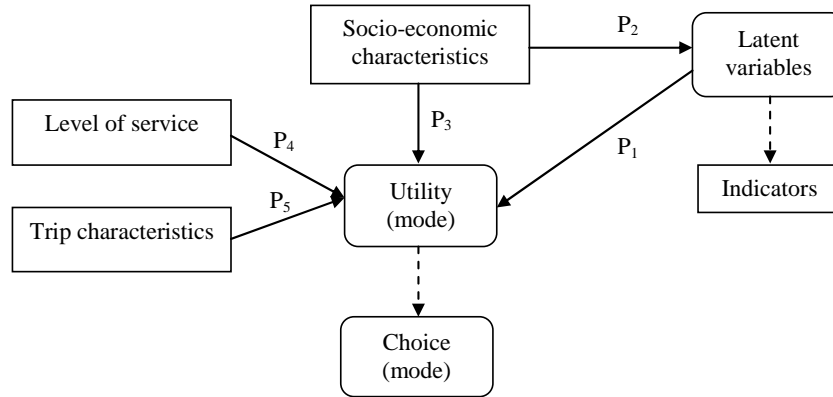
Inclusion of level of service attributes, such as travel time and travel cost, is very common in most of the empirical models on travel mode choice in addition to individual socioeconomic characteristics such as income, gender, and age (Johansson et al., 2006). The interaction between travel and purpose may also indicate the individual trip nature (Ory and Mokhtarian, 2009). Thus, we expect similar effects in our study and propose ( $P_3 - P_5$ ):

***Proposition 3 ( $P_3$ ): Socioeconomic characteristics (e.g., age, income, gender) affect mode choice***

***Proposition 4 ( $P_4$ ): Mode specific attributes (e.g., LOS attributes) affect mode choice***

***Proposition 5 ( $P_5$ ): Trip characteristics (e.g. trip purpose) affect mode choice.***

The analysis specifically investigates the influences of LVs in concert with objective attributes on traveller mode choice (Figure 1). Five propositions detailed above were derived from the literature review and are tested in an empirical analysis of traveller mode choice. Figure 1 describes the structure of hybrid choice model.



**Figure 1** Understanding the structure of hybrid discrete choice model with identified propositions (Anwar et al., in press)

Studies of transportation agencies related to traveller preferences still largely focus on people's responses to travel time and cost, and traveller's observed socio-economic characteristics. This does not properly reflect traveller latent behaviour. Thus, it is necessary to test whether objective attributes embody latent factors and if so, to what extent. In view of the above discussion, this paper examines the inadequacy of objective attributes reflecting latent factors that influence travel behaviour and also analyses the importance of LVs over objective variables in traveller mode choice.

The remainder of this paper is divided into six sections: Section 2 reports on the data sources and collection process and also describes the details of variables used in this research; Section 3 illustrates econometric methods that have been employed in this study; Section 4 describes the estimated results of relationships between LVs and objective attributes in order to quantify their ability to reflect latent factors. It also shows how the RPL model is used to explore the effects of choice variables on mode choice. Section 5 discusses the obtained results; conclusion and implications are included in Section 6; and finally, Section 7 contains some limitations accompanied by future research directions.

## 2 Data and Collection Process

The Sydney Household Travel Survey (HTS) is the largest and most comprehensive source of personal travel data, which is the key data source of this study, for the Sydney Greater Metropolitan Area. This area includes the Sydney and Illawarra Statistical Divisions and the Newcastle Sub-Statistical Division. The investigation in this paper is confined to travel by residents of the Sydney Statistical Division (SD) only. The HTS is the longest running household travel survey in Australia. It began in 1997 and has been operating continuously since then. The survey collects detailed trip information for each day of the year by face-to-face interview. This collection method ensured high data quality and maximised response rates too. Socio-demographic information about the residents of the selected household are also collected. Data collected from 82121 individuals were used in this analysis as a sample size. Each respondent was requested to maintain a simple travel diary to record the details of all trips undertaken for their nominated 24-hour period. An interviewer then interviewed each respondent to collect the details of each trip. For further details about the HTS, its scope, coverage and methodology, please see BTS (2012).

Six LVs and thirteen objective attributes have been evaluated to determine the impact on travellers' mode choice with the adequacy of objective attributes reflecting LVs. Latent variables are: (i) comfort, (ii) convenience, (iii) Safety, (iv) flexibility, (v) reliability, and (vi) satisfaction and twenty indicators described in Table 1 were set to explain them. Thirteen explanatory variables (objective attributes) are: personal annual income (in Australian dollar), age (in years), gender (1 if male, 0 otherwise), having children (0-14 years), car ownership per adult, family size, full time workers of household, travel time (in minutes), travel cost (in Australian dollar), waiting time (in minutes), trip rate (trip per person per day), trip purpose (1 if work, 0 otherwise) and distance travelled (in kilometre).

The following is the list of psychometric indicators (Table 1) that were considered in the modelling approach of this study for structuring the influence of LVs in traveller preference.

**Table 1** Description of latent variables

<i>Latent factors</i>	<i>Explained by (indicators)</i>	<i>Definitions</i>
Comfort	- Enjoy time to read/relax on vehicle	Importance with 1, otherwise 0
	- Stressfulness on vehicle	Importance with 1, otherwise 0
	- Service slower	Importance with 1, otherwise 0
Convenience	- Mode availability	Importance with 1, otherwise 0
	- Accessibility (does not go where required)	Importance with 1, otherwise 0
	- Timetable availability	Importance with 1, otherwise 0
Safety	- Safety response for mode used in 1 <sup>st</sup> trip	Importance with 1, otherwise 0
	- Safety response for mode used in 2 <sup>nd</sup> trip	Importance with 1, otherwise 0
	- Safety response for mode used in 3 <sup>rd</sup> trip	Importance with 1, otherwise 0
Flexibility	- Fixed start and finish times – each day can vary	Importance with 1, otherwise 0
	- Rotating shift	Importance with 1, otherwise 0
	- Roster shift	Importance with 1, otherwise 0
	- Variable hours	Importance with 1, otherwise 0
Reliability	- Frequency	Importance with 1, otherwise 0
	- Punctuality	Importance with 1, otherwise 0
	- Faster	Importance with 1, otherwise 0
Satisfaction	- Cleanliness	Importance with 1, otherwise 0
	- Travel time	Travel time in minutes
	- Travel cost	Travel cost in Australian dollar
	- Waiting time	Waiting time in minutes

Reliability of the indicators listed in Table 1 was tested using factor analytic models (exploratory and confirmatory factor model) with the model fit criteria, such as GFI, AGFI, NFI, CFI and RMSEA with lower and upper bound. The factor analytic model focuses solely on how, and the extent to which, the observed variables are linked to their underlying latent factors (Byrne, 2010). However, due to the limited space allocation for this paper, the findings of factor analytic models are not presented here. For further details about the findings of factor analytic models ( $\gamma$  vector matrix of equation 1), please see Anwar et al. (2011).

### 3 Econometric Methods<sup>1</sup>

There are two approaches available now for incorporating LVs into the choice models (i) *sequential (also known as two-step) approach*, where the LVs are needed to be constructed before being included into the discrete choice model as regular explanatory variables (Yanez et al., 2010; Johansson et al., 2006); and (ii) *simultaneous approach*, where both processes are done simultaneously (Ashok et al., 2002; Bolduc et al., 2008). The first approach i.e. a two-step approach is performed to estimate the results in this paper. Step 1 is the estimation of a MIMIC (multiple indicators and multiple causes) model; a type of regression model with a latent dependent variable(s). Step 2 is the estimation of a choice model with random parameters; information from the first step is incorporated in the second step.

Ben-Akiva et al. (2002a) argued that results obtained using second approach are more consistent and rational than other approach. Conversely, second approach is not popular due to its high complexity (Raveau et al., 2010). However, the authors were biased to implement the sequential approach in this study due to the following reasons:

- i) the decision underlying a travel pattern is decomposed into a series of interrelated choices and analysed one by one that is more relevant to sequential approach rather than simultaneous approach (Kitamura, 1984);
- ii) the estimated results using both sequential and simultaneous approaches were not statistically different (Raveau et al., 2010);
- iii) it is less cumbersome to estimate the model sequentially (Johansson et al. 2006);
- iv) the sequential approach can be easily linked to discrete choice analysis than simultaneous method to analyse the traveller behaviour over a specified period of time, and
- v) it is assumed that travel decision itself is sequential to a certain extent because of uncertainty involvement in traveller decision making process.

#### 3.1 Modelling with Latent Variables

Integration of LVs in choice model is becoming popular for last three decades (Koppelman and Hauser, 1978; Koppelman and Pas, 1980) because inclusion of LVs in choice model provides better understanding of individual's decision making process. Some studies have shown its advantages before (Spear, 1976; Mokhtarian and Salomon, 1997; Kuppam et al., 1999) due to its significantly improvement of explanatory power of traditional models, but until inadequacy determination of objective attributes representing LVs and a comparison between traditional RPL (in which LVs are not integrated) and hybrid RPL (in which LVs are integrated) models towards traveller mode choice have not been done so far and it is absolutely a new challenge that has been presented in this paper.

A MIMIC model, that defines LVs appropriately, is estimated first, where the LVs ( $\eta_{ijl}$ ) are explained by characteristics ( $s_{ijr}$ ) from the users (individuals), alternatives (mode alternative) and trip nature through structural equation (Eq. 1); as the analysts cannot collect data on LVs directly, indicators ( $v_{ijp}$ ) are assigned to explain them through measurement equation (Eq. 2):

$$\eta_{ijl} = \sum_r \alpha_{jlr} * s_{ijr} + v_{ijl} \quad (1)$$

$$y_{ijp} = \sum_l \gamma_{jlp} * \eta_{ijl} + \zeta_{ijp} \quad (2)$$

<sup>1</sup> The authors employ the similar econometric methods that have been used in Anwar et al., (in press).



where,  $i$  to an individual,  $j$  refers to an alternative,  $l$  to a LV,  $r$  to an explanatory variables belong to LOS, SEC and TC and  $p$  to an indicator;  $\alpha_{jlr}$  and  $\gamma_{jlp}$  are parameters to be estimated, while  $v_{ijl}$  and  $\zeta_{ijp}$  are error terms with mean zero and standard deviation to be estimated. The above specifications of MIMIC model are not restricted on the estimation of parameters and the results of model depend on the selected variables.

Eq. (1): Structural

$$\begin{pmatrix} \eta_{\text{Comfort}} \\ \eta_{\text{Convenience}} \\ \eta_{\text{Flexibility}} \\ \eta_{\text{Safety}} \\ \eta_{\text{Reliability}} \\ \eta_{\text{Satisfaction}} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \\ \alpha_{51} & \alpha_{52} & \alpha_{53} \\ \alpha_{61} & \alpha_{62} & \alpha_{63} \end{pmatrix} \begin{pmatrix} \text{Socio-economic} \\ \text{Level of service} \\ \text{Trip characteristics} \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix}$$

Eq. (2): Measurement

$$\begin{pmatrix} 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{pmatrix} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \gamma_{1,4} & \gamma_{1,5} & \gamma_{1,6} \\ \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} & \gamma_{2,4} & \gamma_{2,5} & \gamma_{2,6} \\ \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} & \gamma_{3,4} & \gamma_{3,5} & \gamma_{3,6} \\ \gamma_{4,1} & \gamma_{4,2} & \gamma_{4,3} & \gamma_{4,4} & \gamma_{4,5} & \gamma_{4,6} \\ \gamma_{5,1} & \gamma_{5,2} & \gamma_{5,3} & \gamma_{5,4} & \gamma_{5,5} & \gamma_{5,6} \\ \gamma_{6,1} & \gamma_{6,2} & \gamma_{6,3} & \gamma_{6,4} & \gamma_{6,5} & \gamma_{6,6} \\ \gamma_{7,1} & \gamma_{7,2} & \gamma_{7,3} & \gamma_{7,4} & \gamma_{7,5} & \gamma_{7,6} \\ \gamma_{8,1} & \gamma_{8,2} & \gamma_{8,3} & \gamma_{8,4} & \gamma_{8,5} & \gamma_{8,6} \\ \gamma_{9,1} & \gamma_{9,2} & \gamma_{9,3} & \gamma_{9,4} & \gamma_{9,5} & \gamma_{9,6} \\ \gamma_{10,1} & \gamma_{10,2} & \gamma_{10,3} & \gamma_{10,4} & \gamma_{10,5} & \gamma_{10,6} \\ \gamma_{11,1} & \gamma_{11,2} & \gamma_{11,3} & \gamma_{11,4} & \gamma_{11,5} & \gamma_{11,6} \\ \gamma_{12,1} & \gamma_{12,2} & \gamma_{12,3} & \gamma_{12,4} & \gamma_{12,5} & \gamma_{12,6} \\ \gamma_{13,1} & \gamma_{13,2} & \gamma_{13,3} & \gamma_{13,4} & \gamma_{13,5} & \gamma_{13,6} \\ \gamma_{14,1} & \gamma_{14,2} & \gamma_{14,3} & \gamma_{14,4} & \gamma_{14,5} & \gamma_{14,6} \\ \gamma_{15,1} & \gamma_{15,2} & \gamma_{15,3} & \gamma_{15,4} & \gamma_{15,5} & \gamma_{15,6} \\ \gamma_{16,1} & \gamma_{16,2} & \gamma_{16,3} & \gamma_{16,4} & \gamma_{16,5} & \gamma_{16,6} \\ \gamma_{17,1} & \gamma_{17,2} & \gamma_{17,3} & \gamma_{17,4} & \gamma_{17,5} & \gamma_{17,6} \\ \gamma_{18,1} & \gamma_{18,2} & \gamma_{18,3} & \gamma_{18,4} & \gamma_{18,5} & \gamma_{18,6} \\ \gamma_{19,1} & \gamma_{19,2} & \gamma_{19,3} & \gamma_{19,4} & \gamma_{19,5} & \gamma_{19,6} \\ \gamma_{20,1} & \gamma_{20,2} & \gamma_{20,3} & \gamma_{20,4} & \gamma_{20,5} & \gamma_{20,6} \end{pmatrix} \begin{pmatrix} \eta_{\text{Comfort}} \\ \eta_{\text{Convenience}} \\ \eta_{\text{Flexibility}} \\ \eta_{\text{Safety}} \\ \eta_{\text{Reliability}} \\ \eta_{\text{Satisfaction}} \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \\ \zeta_6 \\ \zeta_7 \\ \zeta_8 \\ \zeta_9 \\ \zeta_{10} \\ \zeta_{11} \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{14} \\ \zeta_{15} \\ \zeta_{16} \\ \zeta_{17} \\ \zeta_{18} \\ \zeta_{19} \\ \zeta_{20} \end{pmatrix}$$

### 3.1.1 Specification of Latent Variable Model

The factor analysis was employed to investigate the structural relationships in MIMIC model that guides the specification for computation of LVs (Figure 2 illustrates the results of this process), which results in the following set of equations.

$$\text{Comfort}_{ij} = \alpha_{inc-com,j} * \text{Income}_i + \alpha_{age-com,j} * \text{Age}_i + \alpha_{gen-com,j} * \text{Gender}_i + \alpha_{car-com,j} * \text{Car ownership}_i + \alpha_{f1w-com,j} * \text{Full time workers}_i + \alpha_{dt-com,j} * \text{Distance travelled} + \alpha_{chi-com,j} * \text{Having children} + v_{com,ij}$$

$$\text{Convenience}_{ij} = \alpha_{age-conv,j} * \text{Age}_i + \alpha_{gen-conv,j} * \text{Gender}_i + \alpha_{car-conv,j} * \text{Car ownership}_i + v_{conv,ij}$$

$$\text{Safety}_{ij} = \alpha_{inc-saf,j} * \text{Income}_i + \alpha_{age-saf,j} * \text{Age}_i + \alpha_{gen-saf,j} * \text{Gender}_i + \alpha_{fs-saf,j} * \text{Family size}_i + \alpha_{tr-saf,j} * \text{Trip rate}_i + v_{saf,ij}$$

$$\text{Flexibility}_{ij} = \alpha_{gen-fle,j} * \text{Gender}_i + \alpha_{chi-fle,j} * \text{Having children}_i + \alpha_{car-fle,j} * \text{Car ownership}_i + \alpha_{tr-fle,j} * \text{Trip rate}_i + v_{fle,ij}$$

$$\text{Reliability}_{ij} = \alpha_{tti-rel,j} * \text{Travel time}_i + \alpha_{wti-rel,j} * \text{Waiting time}_i + \alpha_{fti-rel,j} * \text{Full time workers}_i + \alpha_{tp-rel,j} * \text{Trip purpose}_i + v_{rel,ij}$$

$$\text{Satisfaction}_{ij} = \alpha_{tti-sat,j} * \text{Travel time}_i + \alpha_{tco-sat,j} * \text{Travel cost}_i + \alpha_{wti-sat,j} * \text{Waiting time}_i + \alpha_{dt-sat,j} * \text{Distance travelled}_i + v_{sat,ij}$$

$$y_{y1,ij} = \gamma_{y1,j} * \text{Comfort}_{ij} + \zeta_{y1,ij}$$

$$y_{y2,ij} = \gamma_{y2,j} * \text{Comfort}_{ij} + \zeta_{y2,ij}$$

$$y_{y3,ij} = \gamma_{y3,j} * \text{Comfort}_{ij} + \zeta_{y3,ij}$$

$$y_{y4,ij} = \gamma_{y4,j} * \text{Convenience}_{ij} + \zeta_{y4,ij}$$

$$y_{y5,ij} = \gamma_{y5,j} * \text{Convenience}_{ij} + \zeta_{y5,ij}$$

$$y_{y6,ij} = \gamma_{y6,j} * \text{Convenience}_{ij} + \zeta_{y6,ij}$$

$$y_{y7,ij} = \gamma_{y7,j} * \text{Safety}_{ij} + \zeta_{y7,ij}$$

$$y_{y8,ij} = \gamma_{y8,j} * \text{Safety}_{ij} + \zeta_{y8,ij}$$

$$y_{y9,ij} = \gamma_{y9,j} * \text{Safety}_{ij} + \zeta_{y9,ij}$$

$$y_{y10,ij} = \gamma_{y10,j} * \text{Flexibility}_{ij} + \zeta_{y10,ij}$$

$$y_{y11,ij} = \gamma_{y11,j} * \text{Flexibility}_{ij} + \zeta_{y11,ij}$$

$$y_{y12,ij} = \gamma_{y12,j} * \text{Flexibility}_{ij} + \zeta_{y12,ij}$$

$$y_{y13,ij} = \gamma_{y13,j} * \text{Flexibility}_{ij} + \zeta_{y13,ij}$$

$$y_{y14,ij} = \gamma_{y14,j} * \text{Reliability}_{ij} + \zeta_{y14,ij}$$

$$y_{y15,ij} = \gamma_{y15,j} * \text{Reliability}_{ij} + \zeta_{y15,ij}$$

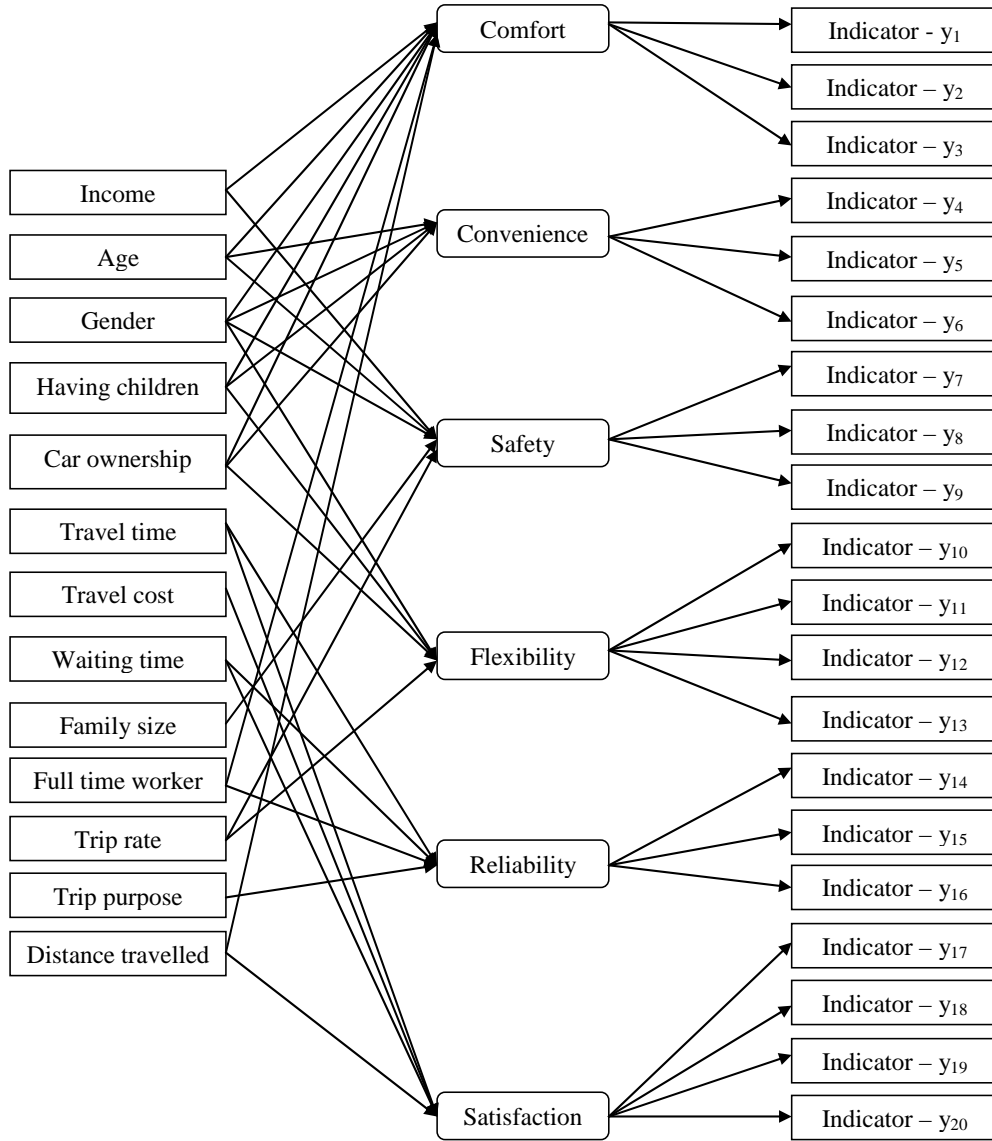
$$y_{y16,ij} = \gamma_{y16,j} * \text{Reliability}_{ij} + \zeta_{y16,ij}$$

$$y_{y17,ij} = \gamma_{y17,j} * \text{Satisfaction}_{ij} + \zeta_{y17,ij}$$

$$y_{y18,ij} = \gamma_{y18,j} * \text{Satisfaction}_{ij} + \zeta_{y18,ij}$$

$$y_{y19,ij} = \gamma_{y19,j} * \text{Satisfaction}_{ij} + \zeta_{y19,ij}$$

$$y_{y20,ij} = \gamma_{y20,j} * \text{Satisfaction}_{ij} + \zeta_{y20,ij}$$



**Figure 2** Process of structural and measurement relationship

### 3.2 Hybrid Discrete Choice Modelling

By maximising the utility ( $U_{ij}$ ), individuals take a decision based on the assumption of random utility theory. It is also assumed that an analyst can only determine a representative portion (systematic component) of utility ( $V_{ij}$ ) function, therefore, an error term ( $\varepsilon_{ij}$ ) to each alternative (Ortuzar and Willumsen, 2001) is required to be included in the function as stochastic component. Mathematically the utility function becomes as below:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad (3)$$

where  $V_{ij}$  is a function of objective attributes  $X_{ijk}$ , i.e. travel time and cost, socio-economic and trip characteristics of the individual, etc. and  $k$  stands for all objective variables together).

Eq. (4) is derived by including latent variables in the utility function, where  $\theta_{jk}$  and  $\beta_{jl}$  are parameters to be estimated:

$$V_{ij} = \sum_k \theta_{jk} * X_{ijk} + \sum_l \beta_{jl} * \eta_{ijl} \quad (4)$$

Only the alternative  $j$  is chosen, if the utility of alternative, ' $j$ ', is greater than or equal to the utility of all other alternatives<sup>2</sup>, ' $t$ ', in the choice set,  $C$ . This can be expressed mathematically with binary variables  $d_{ij}$ :

$$d_{ij} = \begin{cases} 1 & \text{if } U_{ij} \geq U_{it}, \forall t \in C \\ 0 & \text{other case} \end{cases} \quad (5)$$

As sequential approach is used in this study, discrete choice model is estimated with MIMIC model's structure (Eq.1) and measurement (Eq.2) equations (Ben-Akiva et al., 2002b).

### 3.2.1 Specification of Random Parameter Logit<sup>3</sup> (RPL) Model

RPL model has been chosen to analyse the data due to its some advantages. The RPL model is capable to measure random taste variation and to allow unrestricted substitution pattern and correlation among unobserved factors that help to address the limitations of initially innovated logit models, e.g. multinomial (MNL) and nested logit (NL) models. An analyst collects data from the sample population and it is not possible to observe the intangible factors related to the respondents. Therefore, it is common to have the existence of intangible heterogeneity in the sample population and this unobserved heterogeneity is accommodated by the random parameters in RPL model. The estimated constants in MNL and NL models may handle this heterogeneity through data segmentation, but the intangible heterogeneity is more general and representative adequately as it is expressed by using random parameters in RPL model (Hensher and Greene, 2003). The standard deviations of random parameters depict the degree of unobserved heterogeneity and heterogeneity around the mean describes the interaction between random parameters and specified attribute.

Utility is a mathematical representation to an individual. Generally, utility is derived from the attributes of its set of alternatives; e.g., total set of transport mode usage in a given period. The utility maximization rule states that an individual will select the alternative from his/her set of available alternatives that maximizes his or her utility (Koppelman and Bhat, 2006). Thus, an individual selects an alternative once he/she perceives highest utility from that alternative. Further, the value of utility is determined as a function of alternative-specific and individual-specific attributes. According to Eq. (3), the utility that individual  $i$  receives from alternative  $j$  is denoted by  $U_{ij}$ , which is the sum of systematic component  $V_{ij}$  and a stochastic component  $\varepsilon_{ij}$  and in linear relationship.

Within a logit context the condition is imposed that  $\varepsilon_{ij}$  is independent and identically distributed (IID) extreme value type 1 (Gumbel Distribution) and independence of irrelevant alternatives (IIA) property is also existed in initially innovated logit model such as MNL and NL models. These limitations (IID and IIA) should be taken into account in some way. One

<sup>2</sup> All  $t$  includes alternative  $j$

<sup>3</sup> Random parameters logit is also known as "mixed logit (ML)," "mixed multinomial logit (MMNL)," "Kernel logit", "hybrid logit", "random coefficients logit," and "error components logit".

way is to do that the stochastic component can be divided into two additive parts that are uncorrelated. One part is correlated and heteroskedastic among alternatives and, and another part is IID over alternatives and individuals.

The systematic component of utility  $V_{ij}$  can be rewritten as  $x_{ij}\beta_j$ , where  $x_{ij}$  is a vector of explanatory variables that are observed by the analyst from any source related to individuals and alternatives.  $\beta_j$  is a vector of parameters to be estimated. The stochastic component of utility  $\varepsilon_{ij}$  can also be rewritten as  $z_{ij}\eta_i + e_{ij}$ , where  $z_{ij}$  is a vector of characteristics that can vary over individuals, alternatives, or both (there may have some or all common elements in both  $z_{ij}$  and  $x_{ij}$ ), and  $e_{ij}$  is a random term with zero mean that is IID over individuals and alternatives and is normalised to set the scale of utility. The random variable  $\eta_i$  is a vector of random terms with zero mean that varies over individuals according to the distribution  $f(\eta|\Omega)$ , where  $\Omega$  are the fixed parameters of the distribution  $f$ . Accordingly, the utility  $U_{ij}$  that individual  $i$  gets from alternative  $j$  can be written as  $[x_{ij}\beta_j + (z_{ij}\eta_i + e_{ij})]$ . In matrix form, it can be written as:

$$U = X\beta + (Z\eta + e) \quad (6)$$

If IIA exists, then  $\eta = 0$  for all  $i$  and so utility  $U$  depends on only the systematic and IID stochastic portion of utility. Initially innovated logit models assume that IIA does not estimate  $Z\eta$ ; thus  $\eta$  is assumed as zero. Because of that, unobserved taste variations have not been addressed in initially innovated logit models. Hence, by incorporating the effect of  $Z\eta$  in utility function, discrete choice models can be able to accommodate those impacts and thus avoid the IIA assumption. These models estimate  $\Omega$  (the parameters of the distribution of  $\eta$ ) as well as  $\beta$ .

To derive a RPL model from Eq. (6),  $e$  is assumed as IID extreme value, while  $\eta$  follows a general distribution,  $f(\eta|\Omega)$ . If  $\eta = 0$ , it is MNL which has the IIA property. Estimation of the RPL generally involves estimating  $\beta$  and  $\Omega$ . The choice probabilities depend on  $\beta$  and  $\eta$  and the probability to select alternative  $j$  for individual  $i$  with conditional on  $\eta$  is similar as MNL below:

$$P(j|\eta) = L_j(\eta) = (e^{X_j\beta_j + Z_j\eta}) / (\sum_k e^{X_k\beta_k + Z_k\eta}) \quad (7)$$

As  $\eta$  is not given, by integrating over all values of  $\eta$  weighted by the density of  $\eta$  the unconditional choice probability for each individual can be obtained as below.

$$P(j) = \int_{\eta} [(e^{X_j\beta_j + Z_j\eta}) / (\sum_k e^{X_k\beta_k + Z_k\eta})] f(\eta|\Omega) \partial\eta \quad (8)$$

$$i.e. \quad P(j) = \int_{\eta} L_j(\eta) f(\eta|\Omega) \partial\eta \quad (9)$$

Models of this form are called *random parameter logit (RPL)*. The probabilities do not exhibit the IIA property, and the specification of  $f$  describes different substitution patterns. The RPL model handles it in two ways. *One* way is known as random parameter specification that specifies each  $\beta_i$  with both a mean and a standard deviation. The error component is another way to deal with the unobserved taste variation as a separate error component in the random parameter that is by estimated with standard deviation as an additional error component which is an identical outcome.

## 4. Empirical Results: Merits of LVs in Traveller Choice Process

### 4.1 Overall Explanation of Latent Variables by Objective Attributes

Table 2 shows the overall capacity of objective attributes to represent the latent variables. The capacity has been determined by *R-square* values (coefficient of determination) of regression and does not exceed 10%, 12% and 10% of SEC, LOS and TC respectively. That is, LVs are explained at on average 9.1%, 11.9% and 9.3 by SEC, LOS and TC respectively, which is very minimal. This is an indication of inadequacy of objective attributes to describe LVs.

**Table 2** R-square values determining the contribution of objective attributes to latent variables

Latent variables	SEC	LOS	TC
	<i>Beta (t- value)</i>	<i>Beta (t- value)</i>	<i>Beta (t- value)</i>
Comfort	0.096 (12.44)	0.144 (2.45)	0.077 (4.45)
Convenience	0.122 (2.99)	0.135 (4.10)	0.141 (3.17)
Flexibility	0.080 (1.98)	0.080 (1.65)	0.099 (4.24)
Safety	0.104 (11.5)	0.118 (4.15)	0.102 (7.25)
Reliability	0.058 (1.99)	0.117 (2.44)	0.087 (6.08)
Satisfaction	0.085 (10.00)	0.120 (4.01)	0.051 (4.17)

Significant at 90% level of confidence if  $1.960 > t \geq 1.645$ ;

Significant at 95% level of confidence if  $2.576 > t \geq 1.960$ ;

Significant at 99% level of confidence if  $2.810 > t \geq 2.576$ ;

Significant at 99.5% level of confidence if  $3.290 > t \geq 2.810$ ;

Significant at 99.9% level of confidence if  $t \geq 3.290$ .

Specifically, about 12% variation of convenience is explained by unit changes of SEC; about 14% variation of comfort is elucidated by LOS and about 14% variation of convenience is described by unit changes of TC. The influences of SEC on other LVs are less than LOS and TC. LOS and TC do not affect the variation of latent factors substantially. However, on average, the objective attributes account for only 10.1% variation of LVs.

### 4.2 Effects of Traveller Choice Attributes on the Choice of Mode

This section discusses the results of a series of RPL models: 3 traditional RPL (TRPL) and one hybrid RPL (HRPL) models that illustrate the effects of choice attributes on mode choice. The TRPL model deals with objective attributes only and in HRPL, LVs are included with objective attributes. In order to reduce the space, only the results of  $\alpha$  vector matrix in the structural equation of MIMIC model are presented here (Table 3). The estimated coefficients were valid according to model fit criteria, such as GFI, AGI, NFI, CFA and RMSEA with lower and upper bound that were calculated by computer software AMOS v.19. For the detail explanations of the results of structural equation of MIMIC model, please see Anwar et al. (2011). The results obtained from MIMIC model have been used to quantify latent variables that are incorporated in RPL models as explanatory variables. The coefficients of attributes to mode choice are interpreted using traditional (in which LVs are not integrated) and hybrid (in which LVs are integrated) RPL models (Table 4). The models were estimated in LIMDEP (Nlogit 4), econometric software, using maximum likelihood estimation procedures.

**Table 3** MIMIC model results:  $\alpha$  vector matrix of structural equations (t-values in the parenthesis)

	<i>Travel time</i>	<i>Travel cost</i>	<i>Waiting time</i>	<i>Age</i>	<i>Income</i>	<i>Family size</i>	<i>Gender</i>	<i>Car ownership</i>	<i>No. child</i>	<i>Full time</i>	<i>Trip rate</i>	<i>Distance travelled</i>	<i>Trip purpose</i>
Comfort	-0.055 (-2.10)	-0.202 (-5.77)	-0.175 (-2.00)	-0.014 (-11.1)	0.145 (2.72)	-0.008 (-3.15)	0.054 (3.35)	0.221 (5.00)	0.221 (4.21)	0.008 (2.03)	0.058 (4.68)	0.111 (4.84)	0.063 (1.75)
Convenience	-0.127 (-9.51)	-0.058 (-2.00)	-0.222 (-4.35)	-0.132 (-2.45)	0.189 (2.33)	-0.006 (-3.45)	0.189 (2.85)	0.132 (5.63)	0.136 (2.89)	0.071 (3.44)	0.137 (3.43)	0.115 (2.05)	0.171 (2.00)
Flexibility	-0.171 (-7.52)	-0.004 (-1.99)	-0.067 (2.99)	-0.184 (-4.12)	0.082 (-3.50)	0.021 (5.10)	-0.106 (-3.13)	-0.011 (-2.50)	-0.121 (-6.37)	-0.037 (-3.63)	0.012 (2.00)	0.160 (8.00)	0.126 (10.5)
Safety	-0.166 (-6.23)	-0.100 (-3.04)	-0.089 (-1.97)	-0.258 (-3.45)	-0.136 (-4.49)	0.011 (6.0)	-0.08 (-6.85)	-0.087 (-6.78)	-0.121 (-6.37)	-0.037 (-3.44)	0.012 (2.00)	0.168 (6.41)	0.126 (5.73)
Reliability	-0.444 (-5.24)	-0.022 (1.87)	-0.107 (-3.33)	-0.142 (-4.44)	0.026 (2.17)	-0.009 (-2.10)	0.074 (3.85)	0.122 (3.21)	0.013 (4.25)	0.025 (3.13)	0.019 (3.17)	0.212 (3.45)	0.031 (2.58)
Satisfaction	-0.129 (-1.98)	-0.155 (-6.66)	-0.077 (-2.80)	-0.143 (-11.11)	0.028 (4.52)	-0.086 (-4.44)	-0.086 (-3.45)	0.102 (6.19)	0.109 (15.25)	0.045 (5.63)	0.107 (17.83)	0.022 (7.33)	0.025 (2.08)
Model fit criteria													
GFI							0.927						
AGFI							0.902						
NFI							0.964						
CFI							0.911						
RMSEA							0.043						
Lower bound							0.030 (90% CI of RMSEA)						
upper bound							0.051 (90% CI of RMSEA)						

Significant at 90% level of confidence if  $1.960 > t \geq 1.645$ ;  
Significant at 95% level of confidence if  $2.576 > t \geq 1.960$ ;  
Significant at 99% level of confidence if  $2.810 > t \geq 2.576$ ;  
Significant at 99.5% level of confidence if  $3.290 > t \geq 2.810$ ;  
Significant at 99.9% level of confidence if  $t \geq 3.290$ .  
(Source: Anwar et al., 2011)

Table 4 presents the results obtained from RPL models. A series of four RPL models were developed using objective attributes and LVs. The attributes were incorporated in the models with the sequence of LOS → LOS+SEC → LOS+SEC+TC → LOS+SEC+TC+LVs. The first model (*TRPL*<sup>1</sup>) deals with the simplest specification considering only LOS attributes and SEC and TC have been included additionally in the second (*TRPL*<sup>2</sup>) and third (*TRPL*<sup>3</sup>) model respectively. The fourth model, called HRPL, evaluates the effects of LVs in estimations.

**Table 4** Results of random parameter logit models (t-values within the parenthesis)

Attributes	TRPL <sup>1</sup>	TRPL <sup>2</sup>	TRPL <sup>3</sup>	HRPL
Random parameter in utility functions				
Travel cost (mean)	-3.14 (-2.11)	-3.19 (-2.56)	-3.20 (-5.55)	-2.11 (-2.62)
Travel cost (st.dev.)	1.07 (1.99)	1.02 (2.45)	1.05 (3.45)	1.06 (4.21)
Waiting time (mean)	-1.72 (-2.12)	-1.85 (-3.11)	-1.93 (-3.15)	-1.75 (-3.14)
Waiting time (st.dev.)	0.08 (3.11)	0.03 (3.41)	0.004 (2.48)	0.004 (2.99)
Age (mean)		-0.22 (-1.89)	-0.11 (-1.11)	-0.09 (-2.84)
Age (st.dev.)		0.48 (1.66)	0.22 (2.01)	0.58 (2.63)
Car ownership (mean)		1.84 (3.52)	1.91 (5.21)	1.89 (4.00)
Car ownership (st.dev.)		0.03 (3.51)	0.02 (4.21)	0.04 (4.44)
Having children (mean)		-1.78 (-6.44)	-1.80 (-5.41)	-1.77 (-5.02)
Having child (st.dev.)		0.11 (3.65)	0.26 (3.11)	0.12 (2.87)
Trip purpose (mean)			0.07 (3.44)	0.06 (2.15)
Trip purpose (st.dev.)			0.003 (2.33)	0.001 (3.63)
Comfort (mean)				3.32 (7.89)
Comfort (st.dev.)				0.12 (5.66)
Convenience (mean)				3.18 (4.66)
Convenience (st.dev.)				0.22 (5.66)
Safety (mean)				5.18 (11.11)
Safety (st.dev.)				0.45 (9.84)
Flexibility (mean)				0.73 (1.00)
Flexibility (st.dev.)				0.30 (2.16)
Reliability (mean)				5.17 (11.10)
Reliability (st.dev.)				0.01 (9.15)
Satisfaction (mean)				1.23 (2.66)
Satisfaction (st.dev.)				0.09 (2.99)
Nonrandom parameter in utility functions				
Age	-0.08 (-0.99)			
Having children under 5 yrs	-0.97 (-3.62)			
Car ownership	1.27 (3.91)			
Trip purpose	0.97 (2.89)	0.97 (2.91)		
Travel time	-1.17 (-7.85)	-1.17 (-8.77)	-1.19 (-6.42)	-1.11 (-3.63)
Gender	0.29 (1.89)	0.32 (2.13)	0.39 (2.15)	0.21 (2.69)
Income	1.32 (1.85)	1.69 (1.11)	1.98 (1.91)	1.50 (0.89)
Family size	-0.94 (-0.45)	0.94 (1.01)	0.93 (0.99)	0.94 (1.00)
Full time workers of HH	0.97 (0.32)	0.97 (1.45)	0.97 (0.85)	0.97 (1.01)
Trip rate	0.91 (1.11)	0.91 (1.00)	0.91 (1.74)	0.91 (1.86)
Distance travelled	-0.19 (-1.89)	-0.17 (-1.11)	-0.78 (-1.01)	-0.24 (-1.12)
Mode constant				
Car as a passenger (base)	0	0	0	0
Car as a driver	-2.22 (-2.45)	-2.23 (-2.54)	-2.22 (-3.10)	-2.41 (-9.00)
Train	-1.00 (-1.99)	-1.17 (-1.98)	-2.18 (-3.41)	-2.39 (-7.15)
Bus	-0.11 (-0.52)	-0.12 (-1.23)	-0.14 (-1.22)	-0.10 (-1.53)
Heterogeneity around the mean				
Travel cost :Income	-0.11 (-4.21)	-0.10 (-2.98)	-0.12 (-3.62)	-0.01 (-3.99)
Waiting time :Income	-0.54 (-3.56)	-0.54 (-2.56)	-0.54 (-2.96)	-0.03 (-3.85)
Age: Income		-0.11 (-1.89)	-0.08 (-1.98)	-0.12 (-2.14)
Car ownership: Income		0.02 (3.12)	0.01 (3.01)	0.65 (5.14)
Having child: income		-0.02 (-1.99)	-0.09 (-2.66)	-0.17 (-3.01)
Purpose: Income			0.01 (4.01)	0.05 (3.01)



Comfort: Income				0.09 (3.10)
Convenience: Income				0.10 (2.89)
Safety: Income				0.45 (11.52)
Flexibility: Income				0.05 (2.45)
Reliability: Income				0.31 (10.20)
Satisfaction: Income				0.08 (5.10)
Model statistics				
Log likelihood function	-812.41	-768.31	-715.28	-613.37
McFadden Pseudo R-squared	0.21	0.25	0.27	0.36
Akaike Information Criterion (AIC)	0.019	0.018	0.017	0.014
Modal choice probability				
Car as a driver	0.713	0.721	0.731	0.785
Car as a passenger	0.080	0.075	0.055	0.010
Train	0.159	0.160	0.181	0.190
Bus	0.048	0.044	0.033	0.015

Legend: TRPL<sup>1</sup> = Effect of LOS; TRPL<sup>2</sup> = Effect of LOS and SEC; TRPL<sup>3</sup> = Effect of LOS, SEC and TC; HRPL = Effect of LV along with LOS, SEC and TC.

Significant at 90% level of confidence if  $1.960 > t \geq 1.645$ ;

Significant at 95% level of confidence if  $2.576 > t \geq 1.960$ ;

Significant at 99% level of confidence if  $2.810 > t \geq 2.576$ ;

Significant at 99.5% level of confidence if  $3.290 > t \geq 2.810$ ;

Significant at 99.9% level of confidence if  $t \geq 3.290$ .

HRPL model has the similar specification of model TRPL<sup>3</sup>, but the effect of LVs is allowed to vary among individuals; and also models introduce an interaction between the mean estimate of the random parameter and a covariate, which is equivalent to revealing the presence or absence of heterogeneity around the mean parameter estimate. In the TRPL models (TRPL<sup>1</sup>, TRPL<sup>2</sup> and TRPL<sup>3</sup>), all variables except age, income, family size, full time workers of household, trip rate per day per person and distance travelled are significantly associated with the choice of travel mode. The HRPL model provides a better representation of the nature of preferences, as it accounts for variation in travellers' preference heterogeneity (i.e. LVs) across socioeconomic and other characteristics.

In particular, the parameters of LVs are statistically significant in the HRPL model. Moreover, the high significance of the LVs standard deviation parameters in the HRPL model implies that the effects of LVs over the choice process effectively and importantly vary between individuals. Results from the TRPL<sup>1</sup>, TRPL<sup>2</sup> and TRPL<sup>3</sup> models are similar to published research on travel mode choice. Interestingly, it is observed in the estimated parameters that the significance level of TRPL<sup>2</sup> is stronger than TRPL<sup>1</sup>, and TRPL<sup>3</sup> is stronger than RPL<sup>2</sup>. These findings indicate the good explanatory power of the models while socioeconomic and trip characteristics variables are included with the level of service variables. It is also noticed that the signs of the estimated parameters are coherent. With respect to mode-related and individual-specific attributes, the car ownership per adult in a household exhibit strong effects on travel mode choice (Bresson et al., 2004) and the results are also supported it. As expected, owning a car per adult in a household increases the propensity to use a car for daily trips. In contrast, none of the socioeconomic and trip variables of age, personal income, family size, full-time workers in household, trip rate and distance travelled significantly impacted on mode choice. At least for household income this result was unexpected since previous research has identified income as a robust explanatory variable for mode choice (McFadden, 1974; Train, 1980; Kitamura, 1989). In this case, this issue is explored in more depth and eventually understood that owning a car might capture much of the income effect and thus car ownership was found as a strong explanatory variable with strong significance level at 0.001.

The estimated coefficients suggest that the most important attribute is *travel cost*, followed by *waiting time*, *car ownership*, *having child*, and *travel time* according to TRPL models. The estimated coefficients of *waiting time*, *travel time* and *travel cost* variables have the expected negative signs since the utility of a mode decreases as waiting time increases and/or travel time increases and/or the mode becomes more expensive. The expected negative signs of these three variables, in turn, imply that this reduces the choice probability of the corresponding mode. The variable *having children* has negative sign that indicates the sensitivity over the choice. The positive sign of coefficient of car ownership indicates that respondents were more likely to choose (and prefer) car to make trip.

Trip purpose (1 if work, 0 otherwise) and gender (1 if male, 0 otherwise) variables are assigned as dummy variables in this analysis. Signs are positive for both of them. Positive sign of estimated parameter for trip purpose indicates that travellers prefer to drive a car or take train rather than bus to commute to work. Positive sign of gender variables can be interpreted as females preferring the car as a passenger than male to make trips. However, overall the results of TRPL models incorporating LOS, SEC and TC variables are largely consistent with published research on travel mode choice. Having a child also influences preferences for comfortable, safe and reliable mode of transport. When latent variables are incorporated in the model, the significance level of objective variables decrease and this implies that travellers are more motivated by their latent preferences during the mode choice process. The number of cars owned influences the people to decide mode choice and inspires them to be more inclined to car usage.

By incorporating the latent variables in the HRPL model, the results have become more significant. Among the LVs, *safety* and *reliability* of transport mode are the most important variables that are evaluated by the travellers. The coefficients of these two variables are high which indicates its dominant influence over the mode choice process. The importance of a *convenient* and *comfortable* mode of transport is also adequately observed in terms of coefficient and level of significance. These are all as expected, confirming the theoretical validity. Only the *flexibility* of office hours is not significant statistically in this case. However, the introduction of HRPL allows us not only to improve model fit, but also to achieve better estimates of the parameters.

All models seem to fit the data reasonably well in terms of their predictive power and the model log likelihood function. The low pseudo  $R^2$  is to be expected, however, since most discrete choice models in the literature have a poor fit because of the inherent randomness in individual decision making (Cramer, 1991). However, increments in pseudo  $R^2$  from TRPL to HRPL model are the evident as additional parameters are included in the models. On the other words, integrating the relevant variables into the models has increased the explanatory power of the choice models. Thus, it indicates that objective attributes are not well enough to explain most of the variation in traveller choice behaviour. There are some other attributes that have influences on traveller mode choice. Therefore, it can be concluded that LVs have significant effect to increase explanatory power of choice models. In addition, HRPL model has the smallest AIC value that indicates its superiority over the TRPL models.

## 5 Discussions

Due to the integration of LVs, the HRPL choice model have been shown to be better than TRPL models for providing valuable insights into motivational processes in relation to mode choice. The results confirm previous research that among the objective attributes, travel time,

waiting time, travel cost, and car ownership are significant predictors of mode choice. Additionally results of the HRPL model show how preferences for comfort, convenience, safety, flexibility, reliability, and satisfaction impact mode choice. Interestingly, the inclusion of LVs changed the magnitude of coefficients of the objective variables substantially and in that sense delivered true additional insight. For example, the significance level of the income variable sharply declined once LVs were included in the hybrid RPL model. This can be interpreted as LVs being considered a preferred attribute than personal income for SSD people. However, it could be explained by socioeconomic variables affecting preferences and thereby also choice. Although LVs cannot be easily forecasted, the relation of these constructs to objective attributes may aid in forecasting such variables (Johansson et al., 2006), e.g. in an ageing society the salience of the safety value is increased and thereby also the relevance of security for mode choice becomes important.

It is found that the desire for *safety* and *reliability* are the important determinant of commuters' mode choice in SSD. Further understanding includes the desire for comfort and convenience positively impacting on commuter mode choice. Considering LVs, it is observed that the likelihood of train use has increased though car use as a driver is dominant. Thus train companies might consider how they can provide better services. In contrast, as the probability of bus usage is declining, bus companies need to improve the services to attract passengers.

The aim of this paper was to examine the preferences that travellers attach to various attributes of mode choice. Latent preferences significantly impact mode choice thus corroborating our *first proposition* ( $P_1$ ). This was the case for all travellers respective of their socioeconomic characteristics. Designing and implementing modal service with incorporating travellers' desire is valued highly by the travellers and it is an important factor in choosing transport modes.

The *second proposition* ( $P_2$ ) is also supported by the estimated results. LVs are constituted by the socioeconomic characteristics of people. For example, those of a high income and those concerned about children may focus on a comfortable journey, while most professional people prefer a reliable journey. Older and female travellers may pay more attention on safe journey rather than travel time or cost. The *third proposition* ( $P_3$ ) is that socioeconomic characteristic affect mode choice too. In choosing a mode of transport, gender matters. Travellers owning car and having a child are more likely to use a car. In contrast, elderly people prefer to use public transport (bus or train) rather than private car due to decaying health conditions. Our model demonstrates the influence of socioeconomic characteristics on mode choice.

The *fourth proposition* ( $P_4$ ) explains the relevance of LOS (mode-specific attributes) to mode choice. LOS has been described by travel time, travel cost and waiting time that are found as significant variables to influence travellers' decision. It is also observed that LOS variables are mostly influential in TRPL models but once LVs are included in the HRPL model, the influences of LOS variables decrease slightly although still they are strongly significant. It means, the  $P_4$  is adequately supported by the model estimations. The *fifth proposition* ( $P_5$ ) elucidates that travellers consider trip characteristics in their mode choice decision making process. Though influence of TC variables is low, the level of significance is high. It implies that the influence of TC variables cannot be ignored.

The results were plausible and consistent with the literature and predictions. This gives some evidence of the technique's convergent and theoretical validity. To sum up, the results support the contention that travellers' preference heterogeneity is an important determinant in the process of mode choice. The general theoretical conclusion of this study is that future projects can be successful by including LVs of travellers.

## **6 Conclusions and Implications**

This paper evaluates the effects of objective attributes towards LVs in traveller choice behaviour. In extant research, choice process is dominated by the objective mode attributes but in real life the scenario is different. Psychological factors, which have previously been treated as a *black box*, are dominating the choice process considerably, in addition to objective variables (Anwar et al., 2011). Contemporary researchers assume that objective variables explain traveller preference sufficiently. However, this study observes that LOS, SEC and TC make a very limited contribution to LVs. The important relationship between LVs and objective attributes are measured and it is also found that the overall capacity of objective attributes is not sufficient (10.1%) to explain the LVs. Therefore, other factors indicate the distinctness of LVs in traveller preferences that is required to be measured separately from LOS, SEC and TC.

This research makes some methodical and theoretical contributions. Concerning our methodical contribution, we have extended existing research in two major ways: (i) determining the inadequacy of objective choice attributes representing traveller LVs; and (ii) analysing the importance/merits of LVs over objective attributes by comparing between TRPL and HRPL models towards mode choice. The HRPL model clearly outperforms a TRPL model on several counts and provides valuable insights into the motivational process that determine mode choice. A further contribution of this paper is that it suggests and demonstrates a convenient alternative for estimating HRPL model with a structural equation modelling (SEM). From a substantial point of view, HRPL model can be considered as one of the most interesting advances in discrete choice modelling in the last decade.

With respect to the theoretical contribution, we set out to develop a more comprehensive model of choice that also maps the impact of such abstract motivational constructs as values on travellers' real choices. The general structure of our HRPL model consists of a discrete choice part where LVs enter as explanatory variables in addition to the observed attributes of the different choice options as well as attributes of the decision maker. The latent variable part of the model allows for relations between the LVs and objective variables, as well as the contribution of LVs to traveller mode choice. Additionally, socio-economics are included as explanatory variables both in the discrete choice and the latent variable model in order to control for observed heterogeneity and to aid in forecasting the latent variables. In our empirical example, the HRPL model where personal characteristics determine latent preferences which in turn impact on actual behaviour, was proposed and validated.

In addition, we can also conclude as per our findings that the behavioural findings and the modelling techniques have direct policy and planning implications in SSD future transportation planning. These include:

- i) LVs are important in explaining travel behaviour. Ignoring them in the planning process could result in serious errors in model estimation and application. A

systematic effort of latent factors is required in transportation planning to achieve the set objectives fixed by the transport agencies and planners.

- ii) Incorporating these latent factors not only improves the explanatory power of the transportation models but also provides more realistic descriptions of travellers' decision making.

## 7 Limitations and Future Works

Among the limitations of this study are the following: (i) the field of study is SSD. It would be better if the results were reinforced by applications on other geographical areas in order to be able to generalise the conclusions; ii) the lack of possible choice attributes (e.g. habit, psychological distance among individuals) that could be used to evaluate the traveller mode choice behaviour were not included in this study; (iii) simultaneous estimation process has not been used in this study. Accordingly, future research may include a joint estimation of the models, which requires the use of a simulation based approach. Furthermore, it is interesting to extend the current framework to incorporate other known behavioural processes such as habit formation, cognitive learning (of causal knowledge), spatial search, variety seeking, choice-set formation and dynamic updating/adaptation of mental representations. Although substantive work already exists in all or most of these areas separately, the extended model that we presented may offer a starting point for an integrated approach, as it combines latent and non-latent components of choice behaviour.

### Acknowledgment

The authors are grateful to the personnel of Bureau of Transport Statistics (BTS) affiliated with Transport for NSW, Australia for providing access to the largest and most comprehensive household travel survey data of SSD. The University of Wollongong, especially SMART Infrastructure Facility, also deserves special thanks for financial support to carry out this research. Finally, the authors express wholehearted gratitude to the reviewers for their valuable comments that contributed to enrich the paper significantly.

### References

- Anwar, A.H.M.M., Tieu, K., Gibson, P., Win, K.T. and Berryman J.M. in press. 'Analysing the heterogeneity of traveller mode choice preference using a random parameter logit model from the perspective of principal-agent theory', *International Journal of Logistics Systems and management*.
- Anwar, A.H.M.M., Tieu, K., Gibson, P., Berryman, M., Win, K.T. 2011, Structuring the Influence of Latent Variables in Traveller Preference Heterogeneity', in *Transportdynamics: proceedings of the 16<sup>th</sup> International Conference of Hong Kong Society for Transportation Studies*, ed. W.Y. Szeto, S.C. Wong and N.N. Sze, Hong Kong, pp.141-148.
- Ashok, K., William, R.D., Yuan, S. 2002, 'Extending discrete choice models to incorporate attitudinal and other latent variables', *Journal of Marketing Research*, vol. 39, no. 1, pp. 31-46.
- Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. and Rao, V. 1994, 'Combining revealed and stated preferences data', *Marketing Letters*, vol. 5, no. 4, pp. 335-350.

Ben-Akiva, M., Mcfadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D.S., Daly, A., Palma, A.D., Gopinath, D., Karlstrom, A. and Munizaga, M. A. 2002a, 'Hybrid choice models: progress and challenges', *Marketing Letters*, vol. 13, no. 3, pp. 163–175.

Ben-Akiva, M., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. 2002b, 'Integration of choice and latent variable models', in *Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*, ed. HS Mahmassani, Pergamon, Amsterdam, pp. 431-470.

Bhat, C.R. 1998, 'Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling', *Transportation Research Part A: Policy and Practice*, vol. 32, no. 7, pp. 495–507.

Bhat, C.R. 2000, 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation Science*, vol. 34, no. 2, pp. 228–238.

Bolduc, D., 1999, 'A practical technique to estimate multinomial probit models in transportation', *Transportation Research Part B: Methodological*, vol. 33, no. 1, pp. 63–79.

Bolduc, D., Boucher, N. and Alvarez-Daziano, R. 2008, 'Hybrid choice modelling of new technologies for car choice in Canada', *Transportation Research Record*, no. 2082, pp. 63–71.

Bresson G., Dargay J., Madre J.L. and Pirotte A. 2004, 'Economic and structural determinants of the demand for public transport: an analysis on a panel of French urban areas using shrinkage estimators', *Transportation Research Part A: Policy and Practice*, vol. 38, no. 4, pp. 269–285.

Bureau of Transport Statistics (BTS) 2012, *2010/11 Household Travel Survey Summary Report, 2012 Release*, New South Wales Department of Transport, Sydney.

Byrne, B.M. 2010, *Structural equation modeling with AMOS: Basic concepts, applications, and programming, 2<sup>nd</sup> edition*, Routledge, New York.

Can, V.V. 2013, 'Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model', *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 149-159 (in press).

Choo, S. and Mokhtarian, P.L. 2004, 'What type of vehicle do people drive? The role of attitude and life style in influencing vehicle type choice?', *Transportation Research Part A: Policy and Practice*, vol. 38, no. 3, pp. 201–222.

Cohen, A. and Harris, N. 1998, 'Mode choice for VFR journeys', *Journal of Transport Geography*, vol. 6, pp. 43–51.

Collins, C.M. and Chambers, S.M. 2005, 'Psychological and situational influences on commuter-transport-mode-choice', *Environment and Behavior*, vol. 37, no. 5, pp. 640–661.

Commins, N. and Nolan, A. 2011, 'The determinants of mode of transport to work in the Greater Dublin Area', *Transport Policy*, vol. 18, no. 1, pp. 259–268.

Cramer, J.S. 1991, *The logit model: an introduction for economists*. Edward Arnold, New York.

Daly, A., Hess, S., Patruni, B., Potoglou, D. and Rohr, C. 2012, 'Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour', *Transportation*, vol. 39, no. 2, pp. 267–297.

Dissanayake, D. and Morikawa, T. 2005, 'Household travel behaviour in developing countries: Nested logit model of vehicle ownership, mode choice, and trip chaining', *Transportation Research Record* 1805, pp. 45–52.

Ewing, R., Schroeder, W. and Greene, W. 2004, 'School location and student travel analysis of factors affecting mode choice', *Transportation Research Record: Journal of the Transportation Research Board*, no. 1895, pp. 55–63.

Fesenmaier, D.R. 1988, 'Integrating activity patterns into destination choice models', *Journal of Leisure Research*, vol. 20, no. 3, pp. 175–191.

Fleischer, A., Tchetchik, A. and Toledo, T. 2012, 'The impact of fear of flying on travellers' flight choice : choice model with latent variables', *Journal of Travel Research*, vol. 51, no. 5, pp. 653–663

Gopinath, A.D. 1995, *Modeling heterogeneity in discrete choice processes: application travel demand*. PhD thesis, Massachusetts Institute of Technology, USA.

Habib, K.M.N. 2012, 'Modeling commuting mode choice jointly with work start time and work duration', *Transportation Research Part A: Policy and Practice*, vol. 46, no. 1, pp. 33–47.

Hensher D.A. and Greene W.H. 2003, 'Mixed logit models: State of practice', *Transportation*, vol. 30, pp. 133 – 176.

Hess, S. and Stathopoulos, A. 2011, Linking response quality to survey engagement: a combined random scale and latent variable approach', *Proceedings of the 2<sup>nd</sup> International Choice Modelling Conference*, Leeds, UK.

Hess, S., Stathopoulos, A. and Daly, A. 2011, Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies', *Proceedings of the 2<sup>nd</sup> International Choice Modelling Conference*, Leeds, UK.

Jialing, H., Jun, L.I., Xinjun, L. and Honggang, X. 2013, 'Modal Choice of Recreational Tourists under Regional Transportation Integration: A Case Study of Pearl River Delta', *Applied Mechanics and Materials*, Vols. 253-255, pp. 287-292.

Johansson M.V, Heldt, T. and Johansson, P. 2006, 'The effects of attitudes and personality traits on mode choice', *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp. 507-525.

Kitamura, R. 1984. Incorporating Trip Chaining into Analysis of Destination Choice, *Transportation Research part B: Methodological*, Vol. 18B, No. 1, pp. 67-81

Kitamura, R., 1989, 'A causal analysis of car ownership and transit use', *Transportation*, vol. 16, no. 2, pp. 155-73.

Koppelman, F.S. and Bhat, C. 2006, *A self-instructing course in mode choice modelling: Multinomial and nested logit models*, U.S. Department for Transportation, Federal Transit Administration.

Koppelman, F.S. and Hauser, J.R. 1978', 'Destination choice behaviour for non-grocery-shopping trips', *Transportation Research Record*, no. 673, pp. 157-165.

Koppelman, F.S. and Pas, E.I. 1980, 'Travel-choice behaviour: models of perceptions, feelings, preference, and choice', *Transportation Research Record*, no. 765, pp. 26-33.

Kuppam, A.R., Pendyala, R.M. and Rahman, S. 1999, 'Analysis of the role of traveller attitudes and perceptions in explaining mode-choice behaviour', *Transportation Research Record*, no. 1676, pp. 68-76.

McCarthy, P. 1996, 'Market price and income elasticities of new vehicle demands', *The Review of Economics and Statistics*, vol. 78, no. 3, pp. 543-547.

McFadden D. 1986, 'The choice theory approach to marketing research', *Marketing Science*, vol. 5, no. 4, pp. 275 - 297.

McFadden M. 1974, 'The measurement of urban travel demand', *Journal of Public Economics*, vol. 3, pp. 303-328.

Mokhtarian, P.L. and Salomon, I. 1997, 'Modeling the desire to telecommute: the importance of attitudinal factors in behavioural models', *Transportation Research Part A: Policy and Practice*, vol. 31, no. 1, pp. 35-50.

Morikawa T., Ben-Akiva M., and McFadden D. L. 2002, 'Discrete choice models incorporating revealed preferences and psychometric data', *Econometric Models in Marketing*, vol. 16, pp. 29-55.

Nicolau, J.L. and Mas, F.J. 2006, 'The influence of distance and prices on the choice of tourist destinations: the moderating role of motivations', *Tourism Management*, vol. 27, no. 5, pp. 982-996.

Nordlund, A.M. and Garwill, J. 2003, 'Effects of values, problem awareness, and personal norm on willingness to reduce personal car use', *Journal of Environmental Psychology*, vol. 23, no. 4, pp. 339-347.

Ortuzar, J. de D. and Willumsen, L.G. 2001, *Modelling Transport*, 3<sup>rd</sup> edition, John Wiley and Sons, Chichester.



Ory, D. and Mokhtarian, P.L. 2009, 'Modelling the structural relationships among short distance travel amounts, perceptions, affections and desires', *Transportation Research Part A: Policy and Practice*, vol. 43, no. 1, pp. 26-43.

Raveau, S., Alvarez-Daziano, R., Yanez, M.F., Bolduc, D. and de Dios Ortuzar, J. 2010, 'Sequential and simultaneous estimation of hybrid discrete choice models: some new findings', *Transportation Research Record*, no. 2156, pp.131-139.

Spear, B.D. 1976, 'Generalized attribute variable for models of mode choice behaviour', *Transportation Research Record*, no. 592, pp. 6–11.

Temme, D., Paulssen, M. and Dannewald, T. 2008, 'Incorporating latent variables into discrete choice models – a simultaneous estimation approach using SEM software', *BuR Business Research Journal*, vol. 1, no. 2, pp. 220-237.

Train, K. 1998, 'Recreation demand models with taste differences over people', *Land Economics*, vol. 74, no. 2, pp. 230–239.

Train, K. 2009, *Discrete Choice Method with Simulation*, 2<sup>nd</sup> edition, Cambridge University Press, New York.

Train, K.E. 1980, 'A structured logit model of auto ownership and mode choice', *Review of Economic Studies*, vol. 28, no. 2, pp. 357–370.

Walker J. and Ben-Akiva, M. 2002, 'Generalized random utility model', *Mathematical Social Sciences*, vol. 43, no. 3, pp. 303 – 343.