Australasian Transport Research Forum 2012 Proceedings
26 - 28 September 2012, Perth, Australia
Publication website: http://www.patrec.org/atrf.aspx

Stated Preference Survey Experiment Design for Transit-Oriented Development Modelling

Li Meng, Michael A P Taylor, Nicholas Holyoak

Transport Systems, School of Natural & Built Environments, University of South Australia

Menly004@mymail.unisa.edu.au

Abstract

Stated preference experiment design for discrete choice modelling has been recognised as an important yet difficult task with regard to the selection of choice alternatives, attributes and their levels. This paper illustrates techniques for stated preference experiment design and data application in discrete choice modelling for transit-oriented development (TOD) study in a rail corridor. Based on analyses of local census data, station observations and focus groups, this experiment design selects major transport modes as alternatives for building a railway station access mode choice model and local house types for building a residential location choice model. Factors that contributed to people's choices were selected as attributes for the models, and the resident's preferences on these attributes were defined as attribute levels. To optimise survey design, this experiment adopted Bayesian efficient design with estimated prior parameters derived from transport experts in the field. These parameters then applied in utilities for testing different types of computing algorithm selections and draws to obtain an optimised efficient design resulting in lower value of D-error and S-estimates. The designed survey was tested in a pilot study and a full scale survey is just starting. Latent Class models, Random Parameters model and Error Component model were derived from pilot data with the comparison of different draws and random parameter distributions. The initial results suggest that the waiting time for bus is a significant contributor for the station access mode choice, while house type, affordability and the distance from home to preferred school are important for residential choice. The distance from home to the railway station is vital for both choices. The empirical experiment showed the suggested design technique has a high potential for being able to provide policy indications for TOD planning.

Key words: discrete choice modelling, station access, residential choice, land development

1 Introduction

Discrete choice modelling has been promoted as a method to assist in analysing the multi-faceted factors of travel behaviour in regards to travel demand and travel mode choice, including the built-environment and residents characteristics. A list of relevant modelling studies can be found in McFadden (1972, 1974, 1978), McFadden and Train (2000), Walker and Li (2007), Olaru, Smith and Taplin (2011). These studies have demonstrated well established theoretical modelling techniques not only in fundamental Multinomial Logit model (MNL) but also in extended forms of Latent Class model (LCM), Nested Logit model (NL) model and Mixed Multinomial Logit (MMNL) model. However, in these studies, it is hard to find any explanation of how choice data surveys have been designed. Experiment design for discrete choice modelling data collection remains an under investigation and an interesting topic even though it is not new in the field.

Analysts collecting choice data are required to make assumptions on what sort of choice would be likely to be optimal. This 'assumption' had been discussed for centuries. Sir Donald R. Fisher's study, 'The design of experiments' (1935), suggested that inductive inference should be adopted while also advocating Tomas Bayes's prior parameter (1763) for statistic estimation. However, his work was considered as highly controversial at the time by many conventional statisticians. An alternative method for making assumptions is called orthogonal design (Louviere 1988; Tarokh, Jafarkhani & Calderbank 1999; Louviere, Hensher & Swait 2000; Hensher, Rose & Greene 2005; Lancsar, Louviere & Flynn 2006). It assumes the attribute levels are allocated within an orthogonal matrix which may be deficient if orthogonality is lost during the design. Interestingly, some other studies have claimed that using Bayesian parameters usually has an advantage in making this assumption (Fowkes & Wardman 1988; Huber & Zwerina 1996; Sándor & Wedel 2001; Kanninen 2002; Sándor & Wedel 2002). Their main argument is that the properties of orthogonal design are unrelated to real world scenarios, as if they were it would reduce the strength of the relationship among parameters while the Bayesian design distinguishes the unobserved heterogeneity between the attributes. These works reintroduce a Bayesian theory to discrete choice modelling application and the theory has been recognised by many researchers as an improved experiment design compared with orthogonal designs (Hensher & Rose 2007; Rose et al. 2008)

In recent practical modelling works (Jaeger & Rose 2008; Olaru, Smith & Taplin 2011), Bayesian design has been adopted but it is rare to find any study that provides detailed estimations of how to define prior parameters, particularly in regards to transport modelling. The research detailed in this paper is based on an empirical modelling study using a rail corridor as the case site investigate why people choose their mode of transport. This is intended to be used in helping promoting a local transit-oriented development (TOD). The study demonstrates how to define specific transport issues, select discrete choice model's alternatives, attributes and levels, and how to estimate Bayesian prior parameters for constructing optimised hypothetical choice survey questionnaires, from where the derived models could provide robust and statistically significant modelling results.

This paper firstly introduces the specifications of discrete choice modelling and explains the advanced model options. In section 3, Bayesian efficient experiment design and efficient measuring criteria are demonstrated. Section 4 demonstrates the fundamental processes and techniques in experiment design, which includes estimating prior parameters and explaining techniques that can be applied to gain lower efficiency criteria before defining an optimized survey questionnaire. Section 5 uses data collected from the designed questionnaire to build different forms of discrete choice models and explains the model results. The final section provides discussion on experiment design and future study.

2 Discrete choice model

Discrete choice models analyse individual behaviour related to variable attributes in hypothetical choices. The MNL model is a fundamental fomulation. It is defined as:

$$P_{mi} = \frac{\exp(U_{mi})}{\sum_m \exp(U_{mi})} \tag{1}$$

where p_{mi} is the probability that individual i will select alternative m from a set of alternative choices, where the value of each alternative to i is given by its utility function U_{mi}.

$$U_{mj} = V_{mj} + \varepsilon_{mj} \tag{2}$$

where V_{mj} is a function of the measured attributes, ε_{mj} is unobserved attributes.

A full derivation of the MNL model and description of the utility function may be found in references such as Ortúzar and Willumsen (2002). MNL models are confined by the Independence of Irrelevant Alternatives (IIA) axiom, which suggests that with any two alternatives, one chosen probability is unaffected by the existence of other alternatives in the choice set (Ortúzar & Willumsen 2002). This requries necessary and sufficient conditions in experimental designs for model formalization (Louviere 1988).

2.1 *Alternative discrete choice models*

The IIA enables the MNL model to be used to simplify econometric estimations and forecasting, but it cannot estimate accurate choices if the IIA assumption is violated. There are alternative model types, which relax the IIA assumption and demonstrate different statistical properties. This study applies the variety of MMNL models which relax the IIA assumption.

The Mixed Multinomial Logit (MMNL) model, or Mixed Logit (ML) model, provides the flexibility to accommodate general characteristics as well as differences across individuals presented in the variables (Bhat 2001). This model treats variance and covariance in the random component that it represents as an “unobservable” component in the utility.

$$U_{jnt} = \theta_n X_{jnt} + \varepsilon_{jnt} \tag{4}$$

where X_{jnt} is a vector of observable variables, θ_n is a vector of unknown coefficients that vary randomly according to the individual and ε_{jnt} is unobserved attributes. There are also different formulations of MMNL models. Three of which, being the Latent Class model, Random Parameter model and Error Component model will be described below.

Latent Class model

An Laten Class model (LCM) refers to a choice model formulation that considers the inclusion of ‘classes’ which are defined priori by the analysts depending on observable attributes and unobserved latent heterogeneity (Greene & Hensher 2003). The overall probabilities of classes are defined on the basis of estimating the class specific parameters for each respondent.

$$U_{nsj} = V_{nsj/c} + \varepsilon_{nsj/c} \tag{3}$$

where $V_{nsj/c}$ is a function of the measured attributes in a latent class c, and $\varepsilon_{nsj/c}$ is the unobserved attributes in a latent class c. Within each class, the probability assumption is treated the same as in MNL models.

Random Parameter model

MMNL models can be interpreted in several ways by specifying different utilities, e.g. the cross sectional model:

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj} \tag{5}$$

where ε_{nj} is a random term and an Independent and Identically Distributed (IID) extreme value. Probability then depends on covariance density f (β) based on β which is distributed normally as $\beta \sim N(\mu,\sigma^2)$ or another distribution (Train 2003). This specification estimates the heterogeneity existing both within and between individuals and we named it Random Parameter model (RPM).

Error Component model

When the MMNL model ignores random-coefficients, then the error components create correlations among alternatives in a different utility:

$$U_{nj} = \alpha_n' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj} \tag{6}$$

where, $\mu_n' z_{nj} + \varepsilon_{nj}$ are error components as a random portion of the utility which is dependent on $z_{nj,}$ for a standard MMNL, $z_{nj}=0$ (Train 2003). This interpretation is called an Error Component model (ECM). Error terms are added into the utility function, to estimate the heterogeneity between random parameters associated with alternatives or nested alternatives, by estimating different error variances associated with these alternatives.

2.2 ***Data requirements***

Discrete choice models require two types of data sets for parameter estimation, being revealed preference data and stated preference data. Revealed preference data comes from real markets but is based on the decision maker's perceptions of the real market and can be collected by asking individuals in each transit node about their current travel behaviour (Louviere, Hensher & Swait 2000). Revealed preference data questionnaires are designed with straight forward questions to collect socio-demographic information, such as age, gender, income, car ownership and also travel patterns.

Stated preference data comes from choice experiments which require the generation of hypothetical choice scenarios. These scenarios need to be composed by the analyst to be as close to a realistic situation as possible. The data collected can be used to investigate people's perception of new transport facilities or future residential developments that may not yet exist. This information may be useful for policy makers in planning forecasts and in reforming urban structure and transport infrastructure, especially for TOD planning.

It is important to combine revealed preference data and stated preference data sets, because the combination of the two overcomes the limitations of each single set and identifies estimated parameters of an optimal design (Louviere, Hensher & Swait 2000; Hensher, Rose & Greene 2005).

3 Experiment design

A stated choice experiment is 'a way of manipulating attributes and their levels to permit rigorous testing of criteria in hypotheses of interest' (Louviere, Hensher & Swait 2000, p. 84). In the design, alternative attribute levels will be constructed into an asymptotic variance-covariance (AVC) matrix, with each column representing an attribute and each row representing a choice task (Rose & Bliemer 2005). A survey questionnaire with efficient allocated attribute levels enables the collection of high quality information that can be used for discrete choice modelling and minimises the burden and fatigue of respondents. This study will focus on Bayesian design methodology as it has been recognised as an improvement on the original orthogonal design (Hensher & Rose 2007).

3.1 ***Bayesian efficient design***

The uncertainty in obtaining prior information for a discrete choice model utility function is referred to as an expected loss in the Bayesian expected utility function. It can be presented by parameter θ representing a vector or matrix. A particular action will be denoted as a, while all possible actions will be denoted as *A*. The random variable outcome will be denoted X (a vector), $X=(X_1,X_2,\dots,X_n)$, X_i represents the independent observation from a common distribution. A particular realization of X is denoted *x*. The probability distribution of X depends upon the unknown θ. P_θ (A) that denotes the probability of the event A, assumed to be with a density f (x|θ), then

$$P_\theta\ (A) = \int_A\ \mathrm{f(x|\theta)\,dx} \tag{7}$$

The expectation over X of a real value function u(x), the expected utility function, will be

$$E_\theta[u(x)] = \int_\chi\ u(x) f(x|\theta)\, dx \tag{8}$$

then

$$E_\theta[u(x)] = \int_\chi\ u(x)\, d\, F^X(x|\theta) \tag{9}$$

The posterior distribution is combining prior information π (θ|x), which is the conditional distribution of θ given the sample observation x.

$$P_\theta\ (A) = \int_A\ dF(x|\theta) = \int_A\ \pi(\theta|x)d\theta \qquad (10)$$

A more detailed explanation was found in Berger (1985). In Bayesian efficient designs, prior distribution π (θ|x) presents the likely parameter values and optimizes the design over that distribution. The more reliable this prior information is, the more accurate parameter estimation will be.

3.2 *Efficiency criteria*

Efficient design assumes parameters for standard error and approximates the AVC matrix prior to conducting the complete survey. Consequently the applied mathematical derivation can only be called a best guess of the true value of the parameters.

The log likelihood function is described as:

$$L_{N(\beta|X,y)} = \sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} y_{jsn}\ logP_{jsn}(X|\beta) \qquad (11)$$

where

$$P_{jsn}(X|\beta) = \frac{expV_{jsn}(X|\beta)}{\sum_{i=1}^{J} expV_{jsn}(X|\beta)} \qquad (12)$$

$$V_{jsn}(X|\beta) = \sum_{k=1}^{K} \beta_k\, X_{jksn} \qquad (13)$$

The Fisher information matrix is obtained after two derivations:

$$I_N(\beta|X,y) = -E_y\left[\frac{\partial^2 L_N(\beta|X,y)}{\partial\beta\partial\beta'}\right] \qquad (14)$$

$$\frac{\partial^2 L(X|\beta)}{\partial\beta_{k1}\partial\beta_{k2}} = \sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{j=1}^{J} X_{jk_1sn}P_{jsn}(X|\beta)\left(X_{jk_2sn} - \sum_{i=1}^{J} X_{jk_2sn}P_{jsn}(X|\beta)\right) \qquad (15)$$

So the AVC matrix can be:

$$\Omega_N(\beta|X) = I_N^{-1}(\beta|X) \qquad (16)$$

$$\Omega_N(\beta|X) = \begin{bmatrix} \frac{se_1^2}{N} & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \frac{se_k^2}{N} \end{bmatrix} \qquad (17)$$

where, j represents alternative (j = 1, …, J), k represents attribute (k = 1, …, K), s represents choice situation (s = 1, …, S), n represents respondent (n = 1, …, N), design X consisting of attribute levels $X_{jksn,}$ choice observations y, where y_{jsn}=1 if respondent n chooses alternative j in choice situation s (and 0 otherwise), β is parameters to be estimated. After the second derivation, the AVC matrix is independent of observation y, then prior parameters can be estimated without responding data and greater efficiency is then also given by the AVC matrix. These equations were originally stated by McFadden (1974), and slightly modified by Bliemer and Rose (2009).

D-error

The standard errors for parameter constants have a large impact on the efficiency. The measure of efficiency can be the determinant by the AVC matrix:

$$\text{D-error}=\det(\Omega)^{1/K} \qquad (18)$$

In Bayesian efficient design, the prior parameter values are only approximately known, assuming the prior parameter is randomly distributed. Using D-efficient design, assuming $\beta \sim N(\mu, \sigma^2)$:

$$\mathrm{D}_b - error = \int_{\beta} \left(\Omega(\beta|X)\right)^{1/K} f(\beta|u, \sigma^2)\, d\beta \tag{19}$$

A detailed explanation can be referred to Bliemer, Rose and Hess (2008).

S-estimates

S-estimates generate a lower sample size by applying asymptotic t-ratios. If the t-ratio is larger than 1.96, a 95% certainty is obtained (Bliemer & Rose 2009).

Sample size for estimating parameter β_k:

$$N \geq \left[\frac{se_1(\beta_k)t^*}{\beta_k}\right]^2 \tag{20}$$

Where N is sample size. This provides for the minimum sample size and minimum observations. It can be explained for a given assumed prior as, at least N times for all parameters to be statistically significant with a t-ratio of at least 1.96. The highest asymptotic t-ratio may provide the optimal parameter values.

A-error

A-error is the trace of the AVC matrix, which is the summation of all diagonal elements of the matrix. This arithmetic mean is variant according to the design of matrix, e.g. level coding (Zwerina, Huber & Kuhfeld 1996).

4 Experimental design applied to a TOD case study

Applying experimental design to a TOD case study requires a consideration of TOD factors which forms the basis for choosing the model, alternatives, attributes and levels and designing utility functions. Prior parameters need to be elicited and algorithms need be carefully applied and adjusted by computing to achieve optimal efficiency. This study uses Ngene, an experiment design software developed for computational assistance by Choice Metrics Pty Ltd. (2009).

4.1 *Observations*

To assist in experimental design and as part of the wider research scope, this study has collected and analysed a number of related data bases. This has included the analysis of Australian Census data on travel to work modes, an observational survey conducted on railway station access modes and focus groups invited residents from the corridor. The analyses of these observed data provide important information for experiment design.

Adelaide’s Northern rail corridor has a good overall mix of land uses with local residents tending to use a car as their main transport mode. Even people who live close to the railway line use rail less than their car (Australian Bureau of Statistics 2006). A railway station observational survey conducted in 2010 at major transport interchanges, such as Mawson Lakes, covered all the station access points from 6am–7pm on one day each and recorded in 5 minute intervals the passenger transport demands. The survey results from Mawson Lakes show that ‘park and ride’ users caused a car park occupancy rate of 85 per cent for most of the day out of a total of 418 available car parks. A total of 1602 passengers used the train mode to depart the station, whist arriving at the station by bus, car, cycling or walking. Nine feeder bus routes bring in 740 passengers per day. Walk and cycle arrivals only account for 10 per cent of total train users while 17 per cent of users arrived by someone drop off as ‘kiss and ride’. This demonstrates that people use motorised modes of arrival more than walking or cycling.

In follow-up focus group sessions, questions were designed based on literature reviews about TOD features and observations. Discussions with local residents centred on questions such as ‘how often do you use the train?’, ‘why do you use or not use the train?’, ‘where and what kind of house do you live in?’, ‘how do you travel in your local area?’. The highlighted issues focus groups were combined with TOD literatures to form the basis for the discrete choice model alternatives and attributes levels. More information about this corridor observation can be found in Meng, Holyoak and Taylor (2011).

4.2 *The models, alternatives, attributes and levels*

Based on an analysis of the observational survey and focus groups, two models were designed. The station access mode choice (SAMC) model is developed for the purpose of assisting in increasing rail patronage in the short term and evaluates the passenger’s choice of mode to access the train. The residential location choice (RLC) model was designed to assist in evaluating policies to encourage people to move closer to public transport and services in the long term. Table 1 shows detailed alternatives and attributes belonging to the specific models. It was important to describe the attributes to be easy-to- understand and as simple as possible.

Models

A MMNL model structure is powerful for identifying heterogeneity across individual preferences (Jaeger & Rose 2008). Both the SAMC and RLC models were expected to run the RPM and ECM models to test the efficiency of the model and to identify heterogeneity. An LCM is preferable for estimating unobserved preferences with latent heterogeneity in preferences and to separate the population into classes to focus on TOD planning (Greene & Hensher 2003; Olaru, Smith & Taplin 2011). Latent classes are constituted by different levels of socio-demographic factors, such as age, gender, income, and family size, which will be developed to analyse different groups of people’s preferences on mode choices and housing choices. In SAMC model, some environmental attributes influence the mode choice but are not a distinguishing feature of the mode itself. These attributes form a ‘travel occasion’ which reduces their impact on model design compared with mode-related major attributes. These include station safety and security, time of day, weather conditions, train frequency, and accompanied or unaccompanied travel. Each of the 12 sub-model condition scenarios are applied as a condition for each of the 12 main model scenarios. For a similar example of this design application type, see (Jaeger & Rose 2008).

Alternatives

Some research suggests that the number of alternatives forms a U–shape relationship with the variance of the error term. Three to four alternatives possess the highest scale parameter and designs (DeShazo & Fermo 2002; Caussade et al. 2005). For the SAMC model, the four labelled alternatives, Car, Bus, Walk and Bike, which contain different attributes, are major modes for TOD station access. For the RLC model, the 3 unlabelled alternatives, A, B and C, which all have the same attributes are presented as different house types, being separate house, semi-detached/townhouse and apartment/flat. House type is an attribute of all the alternatives. A ‘no choice’ alternative is included in the RLC model for when a respondent has no perceived attractiveness of the available options in comparison with other alternatives (Dhar & Simonson 2003).

Table 1: SAMC model and RLC model alternatives, attributes, levels and coding

Attributes	Attribute Index	Level No.	Attribute Levels	Level Code
SAMC model				
[Dist]Travel Distance to station	A	1, 2, 3, 4, 5, 6	3km, 2.5km, 2km, 1.5km,1.0km, 0.5km	2, 4, 6, 8, 10, 12
[Parka]Parking availability	B	1, 2, 3, 4	$4.00, $2.00, Free parking, drop off	-3, -1, 1, 3
[WtimeB]Wait Time for Bus	C	1, 2, 3, 4	20mins, 15mins, 10mins, 5mins	-3, -1, 1, 3
[Wway]Quality walk route	D	1, 2	Poor, Good	-3, 3
[Bway]Quality bike route	E	1, 2	Poor, Good	-3, 3
SAMC model -Sub Conditional model				
[Ssafe]Station design/Security	A	1, 2	Not safe, Safe	-6, 6
[Weather]Weather Condition	B	1, 2, 3	Wet, Hot, Fine	1, 2, 3
[Trainf]Train frequency	C	1, 2, 3, 4	20mins, 15mins, 10mins, 5mins	2, 4, 6, 8
[Soc]Social Interaction With Others	D	1, 2	Not with friend, With friend	-1, 1
[TimeD]Safety/ Time of Day	F	1, 2	Nighttime, Daytime	-4, 4
RLC model				
[HouseT]House Type	A	1, 2, 3	Separate House, Semi-Detached/Townhouse, Apartment/Flat	2, 4, 6
[Haffor]House Cost/Affordability	B	1, 2, 3, 4	40%, 30%, 20%, 10%	4, 8, 12, 16
[DistTS]Travel Distance to Rail Station	C	1, 2, 3, 4, 5, 6	3km, 2.5km, 2km, 1.5km, 1.0 km 0.5km	2, 4, 6, 8, 10, 12
[DistBS]Distance to Nearest Bus Stop	D	1, 2, 3, 4, 5, 6	0.6km, 0.5km, 0.4km, 0.3km, 0.2km, 0.1km	2, 4, 6, 8, 10, 12
[WorkA]Employment opportunity distance from house	E	1, 2, 3	2.4km, 1.6km, 0.8km	2, 4, 6
[School]Facilities and Service - Preferable School	F	1, 2, 3, 4, 5, 6	1.8km, 1.5km, 1.2km, 0.9km, 0.6km, 0.3km	2, 4, 6, 8, 10, 12
[Shop]Facilities and Service - Shops	G	1, 2, 3, 4, 5, 6	1.8km, 1.5km, 1.2km, 0.9km, 0.6km, 0.3km	2, 4, 6, 8, 10, 12
[Park]Facilities and Service - Parks and Outdoor Areas	H	1, 2, 3, 4, 5, 6	1.8km, 1.5km, 1.2km, 0.9km, 0.6km, 0.3km	2, 4, 6, 8, 10, 12

Attributes

Allowing too many attributes could increase the error variance due to inconsistent choices (DeShazo & Fermo 2002; Caussade et al. 2005). The distance from home to the railway station is a shared alternative for both models. Other attributes included in the SAMC model include station parking fee, bus waiting time, and the quality of walk and bike route as alternative specific attributes. Other alternatives included in the RLC model are house affordability, the distance to public transport nodes, employment, shops, schools and parks.

Levels

Levels of attributes take into account respondents' weights for each attribute for determining preferred alternatives in the process of estimating parameters. Huber and Zwerina (1996) claimed that a level is only meaningful when compared to others in a choice set, although, the predicted average attribute levels may influence the D-error (Rose et al. 2008). A wide range of attribute levels is preferred, but too wide a range will result in a higher error term (Caussade et al. 2005). Columns 3 and 4 in Table 1 demonstrate the levels of attributes for both models, which are inferred from the perception of respondents of focus groups and literatures for related factors. Most of these levels were defined by taking into account the local conditions of the study site. One example is the waiting time for a bus is generally 15 minutes while some routes have a 10 minute waiting period in peak hour and 5 minutes might be a possible scenario for the future. Another example is the distance to the nearest bus stop which has been set at 100m to 600m, because the standard distance between bus stops in Adelaide is planned to be no longer than 600m.

4.3 *Utility function, prior parameters and algorithm*

A computer programming technique was applied for the estimation of the optimised efficient design, deciding whether to use random or Bayesian parameters, estimating prior parameters and choosing algorithms which all influence the optimal design result.

Utility functions

The utility function of the alternative is constituted by attributes and parameters. The number of attributes included and the type of parameter, such as a generic, alternative constant or alternative specific parameter directly defines the number of choice scenarios. As number of choice scenarios have a significant influence upon error variances and we should choose small enough to enable respondents to complete the survey without feeling over-burdened or fatigued (Caussade et al. 2005). Rose and Bliemer (2005) suggested that total number of choice probabilities should be equal to or greater than the number of parameters to be estimated.

The SAMC model has labelled alternatives. Therefore the parameter is either alternative specific or a generic parameter if an attribute assigns the same weight to each mode (Rose & Bliemer 2005). There are 8 alternative specific parameters, 3 constant parameters and 1 error component, totalling 12 choice sets. The RLC model design with unlabelled attributes will only include 10 generic parameter estimates (Bliemer & Rose 2009). There are 10 choice sets possible to estimate 10 parameters. However, to balance the attribute levels of the models (Bliemer & Rose 2009), this number was increased to 12. To reduce the chance of losing data due to respondents not answering all questions, the choice sets of both models were designed to separate into 2 blocks.

Prior parameters

Estimating prior parameter values by providing a prior distribution on parameter values were applied in studies, such as Box and Lucas (1959) and Chaloner and Verdinelli (1995). Researchers (e.g. Rose et al. 2008), claimed that such estimations involve uncertainty and therefore confronts challenges. Studies have tried to acquire experts who 'assess the probability in an actual decision situation' based on their experience and intuition then directly sketching a prior density (Murphy & Winkler 1970; Berger 1985). This method has been discussed as the 'paper-and-pencil' elicitation method in studies of Van Lenthe (1993) Sándor & Wedel (2001) and Rose et al. (2008).

This study first invited three assessors, who have extensive research experience in transport and land use, to estimate the probabilities of specific levels for one particular attribute, as if this choice is provided by all possible combinations of the attributes and levels. The distributions of the parameters were then derived as a normal distribution (e.g. Kessels et al.

2009), see Figure 1 and Figure 2. For example, 'Dist car' means distance to railway station for car alternative parameters that follows a prior density β~N (0.17, 0.07) in the SAMC model. For negative attributes, we define the attributes level as a minus, which enables the prior parameter to be positive (Kanninen 2002). For improved certainty, the assumed prior parameter could be tested by a pilot study, with small samples that may provide reasonable priors (Huber & Zwerina 1996).

Figure 2: Estimated prior parameter distribution for SAMC model attributes

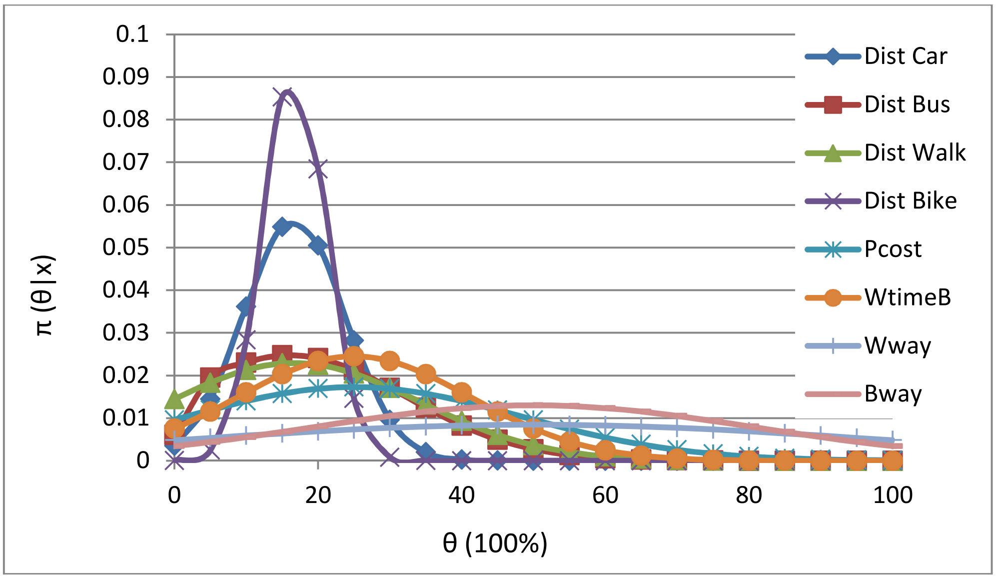
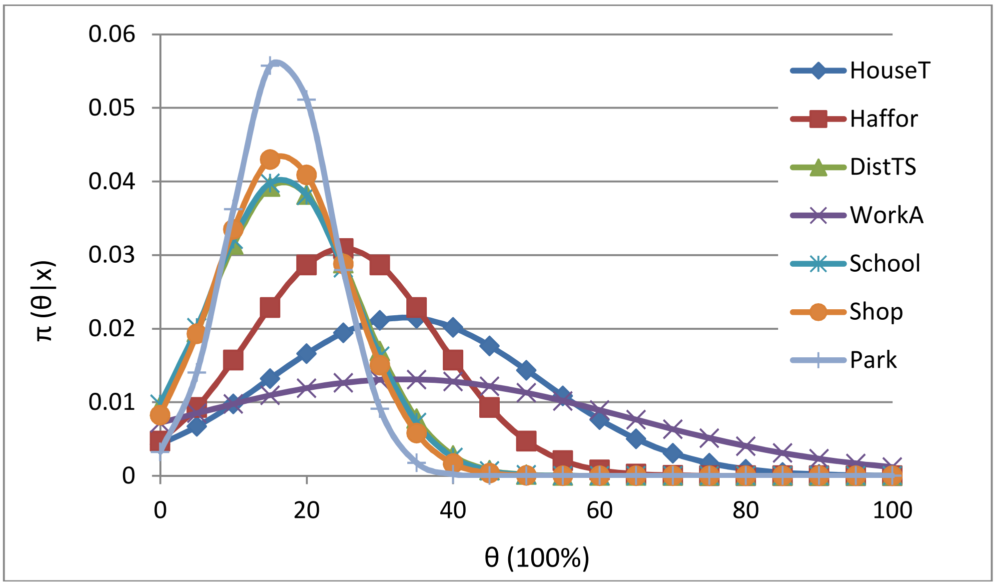


Figure 3: Estimated prior parameter distribution for RLC model attributes



Efficiency and algorithm

Several algorithms can be used for design generation, such as relabeling, swapping, cycling or a Modified Federov algorithm (Huber & Zwerina 1996; Sándor & Wedel 2001). The first three are column based algorithms that reassign, shift and rotate the levels of attributes in choice sets for smaller errors. The modified Federov algorithm searches for the lowest efficient error in all possible combinations of the choice situation and is based on rows. Algorithms will improve the result of the AVC matrix, by providing lower D-error and A-error

(Sándor & Wedel 2002). They suggested that the cycling algorithm performance suits Bayesian design more than relabeling or swapping, but it requests all attribute have the same set of levels. This study does meet the condition of same levels of each attribute, we elected swapping generation (as relabeling experienced a difficulty in the simulation which need further investigate).

For computing the design, the choice can be Quasi-random Monte Carlo simulation (MC) or Gaussian quadrature (Bliemer, Rose & Hess 2008). Quasi-random MC is computed by Halton sequences which divide 0-1 spaces into prime segments or Sobol sequences which provide a more multi-dimensional coverage in a higher dimension than Halton. The Gaussian quadrature method uses cubature methods for orthogonal polynomials, up to 10 abscissas (Bliemer, Rose & Hess 2008).

This study compares the three draw types in random draws and Bayesian draws. Initially, all parameter estimations were assigned with a Bayesian draw parameter, however, this resulted in additional choice sets. The simple solution is to apply the parameters with a lower estimated t-ratio to a random draw parameter (Bliemer, Rose & Hess 2008). The experiment tried different algorithms by using both Bayesian and random parameter (see Figure 4). In the experiments for the RLC model, we have found Gaussian draws with 2 abscissas for random parameters and with different abscissa (e.g., G 2 G 1322223), for Bayesian draws, providing D_b-error=0.038, A_b-error= 0.125, outperformed other draws in D_b-error and A_b-error which meets with the suggestion that Gaussian draws outperform other draws in a previous study (Bliemer, Rose & Hess 2008). Figure 5 shows the S-estimate for the RLC model, G 2 G 2 provides a lower S-estimate of 21 but a higher D_b error with 0.04.

Figure 4: RLC model A-Error and D-Error result with different draws

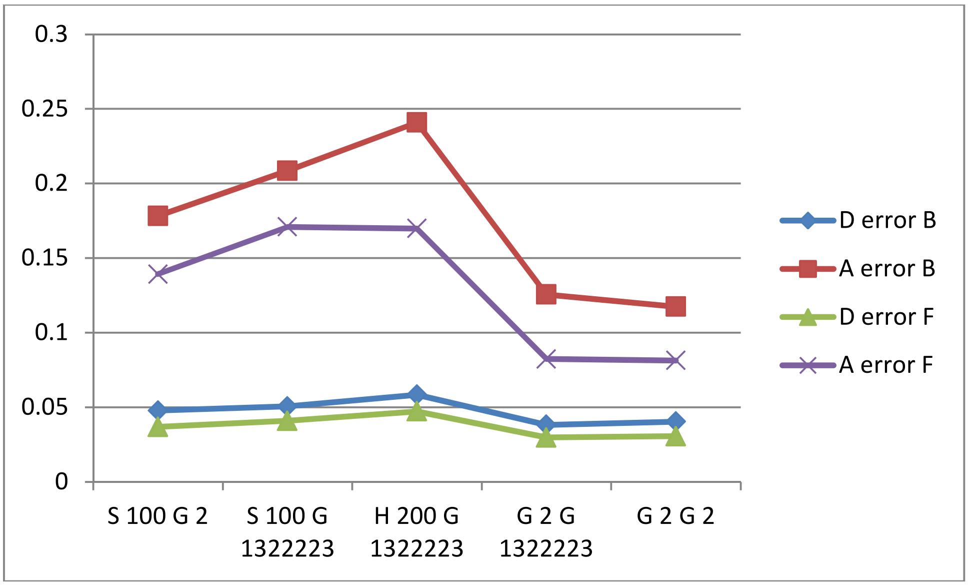
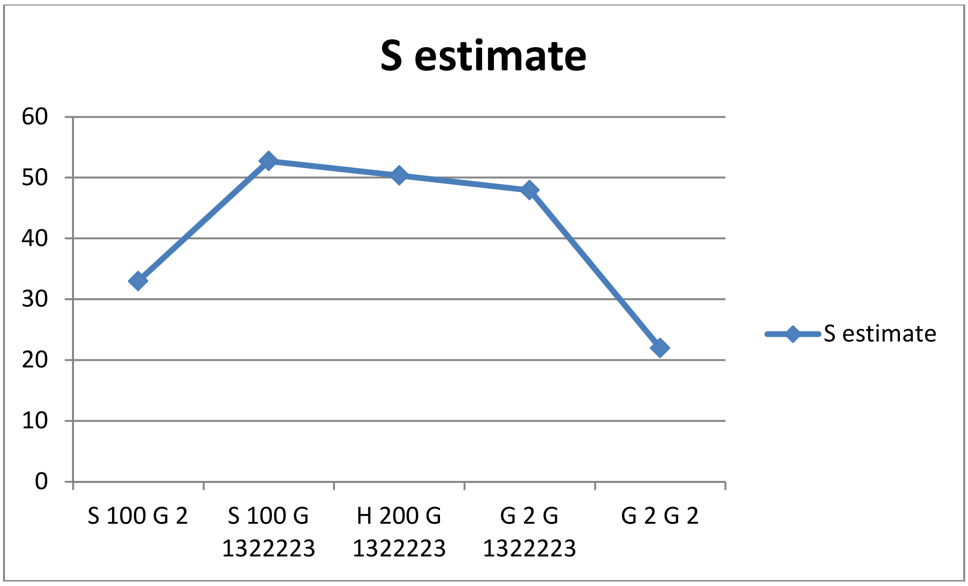


Figure 5: RLC model s-estimate result with different draws



In performing tests of draws for the SAMC model, a high S-estimate of 595.8 was computed. This suggested further investigation to overcome the high number.

Efficiency and attribute, attribute levels and prior parameter

In re-considering the design of the SAMC model, the influence of the selected attributes and levels needed to be analysed to consider whether they fit the objectives of the model. One area that could have created uncertainty could be the attribute of car parking provided in the train station. In the original design, the attribute 'parking cost' had a 'None' level, which includes existing levels of 'drop off' and 'free parking'. There is a significant difference between those two options which strongly affect station area land use and station access mode choice. Capturing the 'drop off' and 'free parking' values separately might be more valuable for making policy suggestions on station land use. Therefore the 'parking cost' attribute was changed to 'car parking availability', and the levels were changed from 'None, $2, $4, $6' to 'Drop off bay, free parking, $2/day parking, $4/day parking'. Accordingly, the prior parameter was changed. The designed result gained a lower value of D-error, A-error, B-estimates, and particularly a lower S-estimate which dropped from 596 to 42 (see Table 2). Consequently the change resulted in higher design efficiency.

Table 2: SAMC model attribute and level adjustment and efficency result

	Before change				**After change**			
Attribute name	Car parking cost				Car parking availability			
Attribute levels	None, $2, $4, $6				Drop off bay, free parking, $2/day parking, $4/day parking			
Prior parameter	n (0.25, 0.23)				n (0.25, 0.09)			
Efficient criteria	D error	A error	B estimate	S estimate	D error	A error	B estimate	S estimate
Fixed	0.121	1.706	35.244	68.928	0.120	1.518	32.252	60.699
Bayesian mean	0.130	1.781	0.294	595.864	0.126	1.571	0.284	41.865

For all mentioned techniques practiced, the experiment design created choice scenarios which are optimised based on the efficient criterion of D-error and S-estimates and other statistical properties, e.g. A-error. The stated choice survey questionnaire was constituted of 12 scenarios of each of the SAMC model and RLC model, see samples of them in Figures 6 and 7.

Figure 6: SAMC choice scenario example

	Car	Bus	Walk	Bike
Travel distance to the train station	1.5 km	1 km	2 km	3 km
Parking availability at the train station	Drop off bay			
Waiting time for a bus to train station		20 mins		
Quality of walking route			Poor	
Quality of bicycle route				Good
Which of the mode alternatives would you choose for your journey to the train station?	☐	☐	☐	☐

Figure 7: LCM choice scenario example

	House "A"	House "B"	House "C"	None
House type	Apartment/Flat	Apartment/Flat	Semi-Detached/ Townhouse	
House affordability (percent of weekly income spent on loan repayments or rent)	10%	10%	30%	
Travel distance to the train station	0.5 km	3 km	1 km	
Distance to nearest bus stop	0.4 km	0.5 km	0.3 km	
Travel distance to nearest potential employment opportunity from house	1.6 km	1.6 km	1.6 km	
Travel distance to a preferred school	1.5 km	0.9 km	1.5 km	
Travel distance to shops/supermarket	1.8 km	0.3 km	1.2 km	
Travel distance to parks/recreation areas	1.5 km	0.9 km	0.9 km	
Which of the house alternatives do you prefer?	House "A" ☐	House "B" ☐	House "C" ☐	None ☐
If you answered "none" previously and had to choose between House A, B or C, which of these alternatives do you prefer?	House "A" ☐	House "B" ☐	House "C" ☐	

Revealed preference data was collected by asking 24 questions about the respondents' socio-demographic information, travel activities, mode choice, car ownership, family structure, residential type and service availability. The full survey form including both revealed preference questions and stated preference questions were distributed small number of respondents for a pilot test. Comments were collected for obtaining a broad aspect of view and improving the designed survey questions to provide a quality database for robust modelling.

5 Discrete choice modelling results

The pilot study was conducted by surveying staff and PhD students at the University of South Australia. Over 100 survey forms were handed out to mail boxes or in person. 50 answered forms were collected, of which three of them were missing one or two answers in the choice scenarios and as such were deleted from the data set. The remaining 47 samples were sufficient to meet the minimum sample size requirement for S-estimates of 42 for the SAMC model and 21 for the LCM model. By analysing the characteristic of respondents, e.g. age, gender, income, we found that the variables of daily activities (DAAC) and the number of people living in the respondent's dwelling (NPID) distinguish the sample into groups. Figure 1 shows the density of different daily activities, where 1 represents respondents whose daily activity is full time work, 5 represents full time study, while 11 represents respondents who do full time work and part time study. Figure 2 shows the number of people living in the dwelling. Two people living in one dwelling is the most common category. In this pilot modelling study, we focused on DAAC and NPID as distinctive characteristics, in particular respondents who are doing full time study and have 2 people living in their dwelling.

Figure 8: Daily activity density

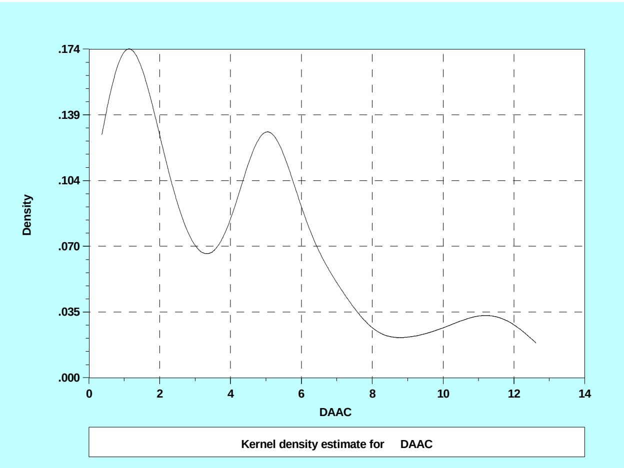
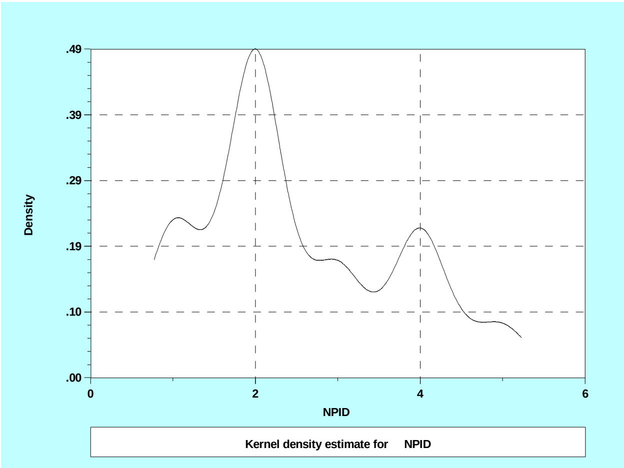


Figure 9: Number of person in dwelling density



5.1 *SAMC model*

An SAMC model was designed to study residents' preference for their mode of accessing the train station, providing 12 choice scenarios sets. A MNL, LCM, RPM and ECM models have been developed to estimate each attribute in different model structures

Latent Class model

Since the likelihood ratio test is not appropriate for the LCM (Greene & Hensher 2003), the AIC criterion is the tool to use to improve an LCM model. The results in table 3 show that the LCM gained a smaller AIC value of 2.260 compared to the MNL model with 2.470. Two latent classes were selected as the best fit ahead of using 3 or 4 classes. We found that overall people consider their bus stop distance from home, waiting time for a bus, the time of day to go to railway station, walking distance to the station and weather when deciding on their choice of transport mode. One class consisting of 73 per cent of respondents drove their car to the train station due to a bus stop being too far from home or too long a waiting time for both the bus and the train. We named this the car access class. The remaining 27 per cent of respondents shift their mode to access the train station between car, bus, walking or cycling depending on various factors including the time of day, the weather, the frequency of trains and convenience of each mode, e.g. car parking availability or walking route quality. We named this class the multi-mode access class. Within the LCM classes, we further create a class probability model with the utility function consisting of a constant (-0.103) and variables DAAC (0.002) and NPID (0.459). The results showed that the full time students surveyed with two people living in their house have 76 per cent belonging to car access class which is 3 per cent higher than the average in the sample.

Random Parameter model

The RPM provides rich information on behaviour preference analysis. It is however difficult to decide which attributes of alternatives should have random parameters and what random parameter should be used. We first tested statistically significant variables in the MNL model, using 15 Halton draws to test using a normal distribution. Waiting time for a bus and walking distance showed statistically significant t-ratios. Next, a combination of a Normal distribution, a lognormal distribution and a triangle distribution were used to compare these two variables using 100 Halton draws. The results showed a Normal distribution for waiting time for a bus and a triangle distribution for walking distance provided an improved model fit. We then used a further 1000 Halton draws to obtain an estimation that is shown in Table 5.

The results for waiting time for a bus showed an estimated mean of 0.399 and an estimated standard deviation of 0.840, with 68 percent of the distribution above zero and 32 percent below. This implies that shorter waiting time for a bus is a positive inducement for attracting about two-third of train users, while the other one-third of train users may choose to take the bus to the station regardless of waiting time. The results for walking distance showed over four-fifths of train users may choose to walk to the station if the distance is preferable, while less than one-fifth might have other reasons to choose not to walk to the train station.

A RPM is able to estimate the interaction effects of each random parameter with other attributes to determine whether heterogeneity may exist in the data. In this RPM model, daily activity and the number of persons in a dwelling were tested for the preference of heterogeneity around the mean of the random parameter. Table 3 shows that the interaction between walking distance and the number of persons in a dwelling is statistically significant with a t-ratio -2.302. Respondents who have more people living in their home (e.g. children) might lack time or find it too difficult to be able to choose walk to the train station.

Error Component model

The ECM nests alternatives of Walk and Bike, and provides additional information for the preference heterogeneity associated with them which we might not be able to account for by random parameterisation (Hensher, Rose & Greene 2005). Table 3 shows a statistically significant t-ratio of 4.216 for these two alternatives.

The modelling specifications of the SAMC model also tested the relationship of elasticity between the distance from home to the train station and mode choices. The results obtained from different models, each demonstrated a similar effect that if the distance to the station is changed, there is a significant increase in respondents choosing to walk, but not as much for the car, bus and bike modes. Table 4 shows the elasticity for the distance from the house to the railway station which shows that for a 1 per cent change in distance, the possibility of choosing to access the station by walking changes far greater than for other modes. The MNL and LCM models showed a higher elasticity at around 1.8 than the RPM and ECM offer.

Table 4: Effect of changes in distance from home to the train station in tested models

Change in distance	MNL	LCM	RPM	ECM
Change in choice of car	-0.104	-0.117	-0.086	-0.051
Change in choice of bus	-0.081	-0.072	-0.068	-0.043
Change in choice of walk	1.854	1.829	1.541	1.432
Change in choice of bike	-0.013	-0.145	-0.009	-0.071

Figures 10 and 11 provide the histograms for the waiting time for a bus and walking distance to the train station. The graphs were estimated on the sample population as a whole rather than on the condition of any individual choice. The frequency of waiting time for a bus is roughly asymmetrically distributed either side of the mean, which might indicate that respondents consider this variable in a similar way to each other. The walking distance histogram is skewed to the right, which could possibly mean that if the walking distance is longer than the average acceptable distance, there might be people who still choose to walk. These unconditional parameter estimates can help predict results for the extended population outside of the sample, if the sample is large enough (Hensher, Rose & Greene 2005).

Figure 10: Waiting time for bus histogram

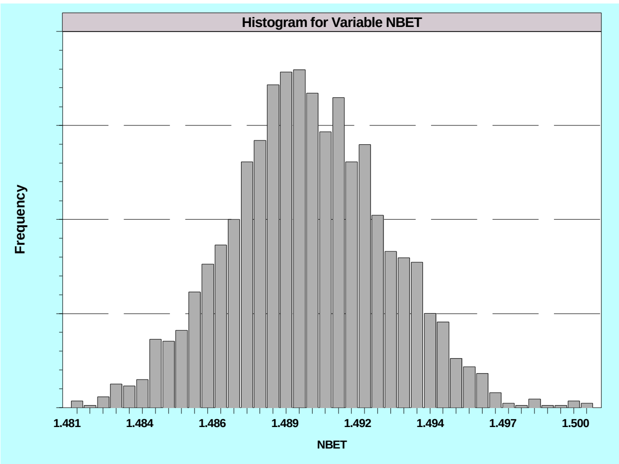


Figure 11: Walk distance histogram

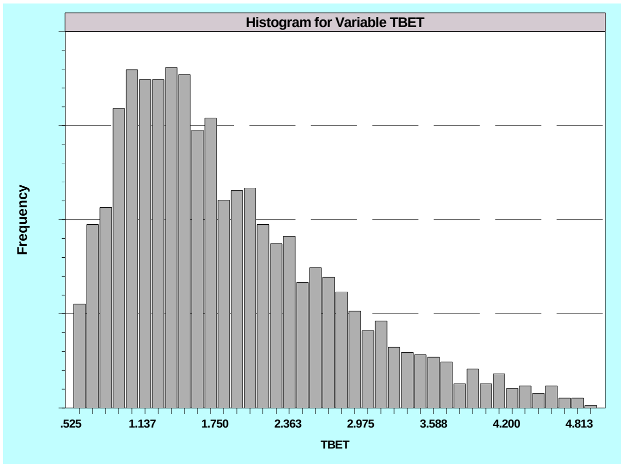


Table 3: The results of station access mode choice model of MNL, LCM , RPM and ECM models

	MNL			**LCM**-Class1			**LCM**-Class2			**RPM**			**ECM**		
Variable	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.
										Random parameters in utility functions					
Waiting time for Bus	0.254	5.063	***	0.184	3.318	**	0.961	5.383	***	0.399	3.252	**	0.424	2.805	**
Walk distance	0.316	8.558	***	0.293	6.058	***	0.681	7.649	***	0.466	6.620	***	0.585	5.454	***
										Nonrandom parameters in utility functions					
Bike distance	0.014	0.329		0.023	0.178		0.254	3.553	**	0.012	0.301		0.071	0.757	
Bike route quality	0.021	0.365		0.355	1.749		0.539	3.379	**	0.027	0.473		0.060	0.513	
Bus	-0.418	-0.764		5.643	2.490	*	-5.379	-4.113	***	-0.913	-1.588		-0.346	-0.363	
Bus station distance	0.046	1.375		0.047	1.138		0.047	0.619		0.052	1.506		0.049	1.008	
Car parking type	-0.128	-1.250		-0.543	-0.923		-0.466	-3.486	**	-0.131	-1.255		-0.182	-0.893	
Car	0.453	0.730		6.046	2.669	**	2.307	1.363		0.424	0.660		1.012	0.947	
Car drive distance	0.044	1.471		0.043	1.241		0.224	1.911		0.047	1.534		0.042	0.951	
Daily activity	0.067	1.610		-0.491	-2.931	**	0.483	5.309	***	0.071	1.659		0.167	1.804	
Driving license	0.071	1.608		0.101	1.514		0.096	1.129		0.078	1.718		0.129	2.419	*
Number of people	0.084	1.120		0.095	1.106		0.419	1.850		-0.056	-0.658		-0.039	-0.530	
Station parking availability	0.047	0.938		-0.020	-0.339		1.486	4.530	***	0.050	0.959		0.072	1.141	
Register vehicle	0.331	2.599	**	0.301	1.994	*	0.751	2.830	**	0.353	2.594	**	0.377	3.181	**
Social interaction with others	-0.238	-2.084	*	-0.256	-1.952		0.066	0.211		-0.243	-2.013	*	-0.231	-1.538	
Station design/security	-0.019	-0.723		-0.091	-1.004		-0.119	-2.495	*	-0.023	-0.862		-0.043	-0.918	
Time of day	-0.105	-3.030	**	-0.228	-1.544		-0.387	-5.727	***	-0.106	-3.059	**	-0.119	-2.490	*
Train frequency	-0.081	-1.482		-0.070	-1.125		-0.417	-2.308	*	-0.094	-1.652		-0.111	-1.616	
Walk	-2.797	-4.775	***	3.123	1.390		-7.379	-6.278	***	-3.416	-5.499	***	-3.792	-3.936	**
Walk way quality	0.107	2.536	*	0.028	0.515		0.556	5.821	***	0.138	2.754	**	0.154	2.765	**
Weather	0.384	2.546	*	0.208	1.109		4.344	4.929	***	0.418	2.649	**	0.461	2.011	*
				Class assignment											
Constant				-0.103	-0.178										
Daily activity				0.002	0.031										
Number of people in dwelling				0.459	2.085	*									

Ns Waiting time for bus					0.002	0.014		0.003	0.001	
Ts Walk distance					0.376	4.599	***	0.440	3.823	**
					Heterogeneity in mean, Parameter: Variable					
Waiting time for bus: daily activity					0.003	0.243		0.004	0.240	
Waiting time for bus: number of person in dwelling					-0.056	-1.514		-0.061	-1.368	
Walk distance: daily activity					-0.001	-0.114		-0.013	-1.149	
Walk distance: number of person in dwelling					-0.061	-3.013	**	-0.058	-2.302	*
SigmaE01 on Walk and Bike								2.282	4.216	**
Log likelihood function	-675.5		-592.2			-661.6			-625.4	
Info. Criterion: AIC	2.470		2.260			2.4419			2.317	
Finite Sample: AIC	2.473		2.274			2.4469			2.322	
Info. Criterion: BIC	2.631		2.606			2.6494			2.532	
Info. Criterion: HQIC	2.533		2.395			2.5229			2.401	
Restricted log likelihood			-781.9			-781.9			-781.8	
McFadden Pseudo R-squared			0.243			0.154			0.200	
Chi squared	-675.5	[18]	379.3	[45]		240.54	[27]		312.9	[21]
Prob [ChiSqd > value]	0.000		0.000			0.000			0.000	
At start values -675.4896			0.123			0.021			0.074	

Notes:

- * significant p value =0
- ** significant p value<0.01
- *** significant p value <0.05
- Parentheses indicates degree of freedom for Chi squared estimate

5.2 *RLC modelling*

The residential location model in this study was based on the answers to 12 hypothetic house location scenarios. MNL, LCM, RMP and MMNL models were derived from data obtained.

Latent Class model

The results (see Table 5) showed that the LCM has a better Critical AIC value of 1.894 than the 2.096 scored by the MNL model making it more statistically significant. We also determined that three-classes of LCM model provided a better estimation than the outputs of 2, 4 or 5 classes for this data. House affordability is an important attributor to house choice to all three classes in the model. The first class which consisted of 31 per cent of the whole sample, was made up of respondents who consider the convenience of their house location for their daily activities when choosing their desired house as well as the house type. This class can be named the activity class. The second class which represented 27 per cent of the whole sample, consisted of respondents who consider whether their house location is close to train or bus stops, and whether they or their family have easy access to shops, schools and parks. This can be named the transit class. The final class consisting of the remaining 43 per cent, strongly consider house type and whether their potential work place or a preferred school is close to their house when choosing their home. We named this class the working class. Using the same procedure as conducted previously for SAMC model, daily activity and the number of people in a dwelling parameter have been estimated for the RLC model. The parameters are -0.344 (constant), -0.083 (DAAC) and 0.107 (NPID). Using this data we can calculate that the full time students surveyed with 2 people living in their home are represented by 28 per cent in the activity class, 37 per cent in the transit class and 49 per cent in the working class.

Random Parameters model

The RPM provided a better model fit (see Table 5) with a log likelihood of -577.1 compared with -561.1 for the MNL model. The same process as was done for the RPM of SAMC model was followed to determine the most statistically significant parameters for the RLC model. The results showed that a normal distribution for housing affordability and a lognormal distribution for distance to train station produced the best fit. With 1000 Halton draws, they offered more statistically significant results with a t-ratio of 4.131 and -4.129. The estimation of house type gives a mean of 0.242 and estimated standard deviation of 0.402, such that 73 percent of the distribution is above zero and 27 percent below. This implies that the price of a house being affordable is a positive inducement for nearly three quarters of respondents, while just over a quarter of respondents do not consider house affordability to be as important. The results for the distance from house to a train station shows that less than one-third of respondents would consider the distance to train station as an important factor in making their house choice, while more than two-thirds would not consider whether their house will be located close to a train station.

Figure 12 shows housing affordability with a normal distribution is asymmetrically distributed on both sides of the mean, which indicates that residents share a common judgement on the price of their house. Figure 13 shows the distance to train station histogram with a lognormal distribution. It has a very short right-hand tail, which shows that a very small proportion of unreasonable values contribute to the distance to train station distribution.

Figure 12: Housing affordability histogram

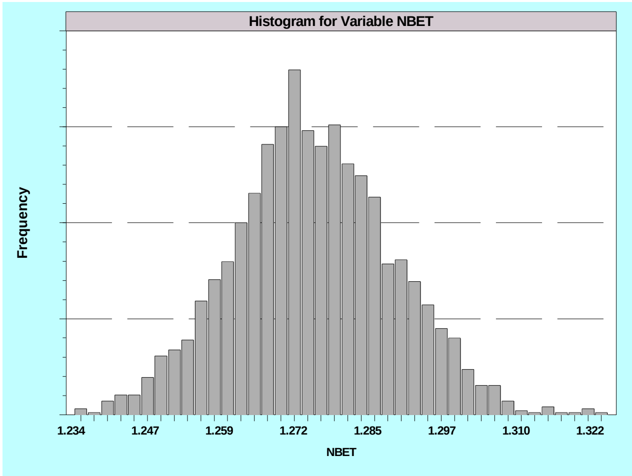
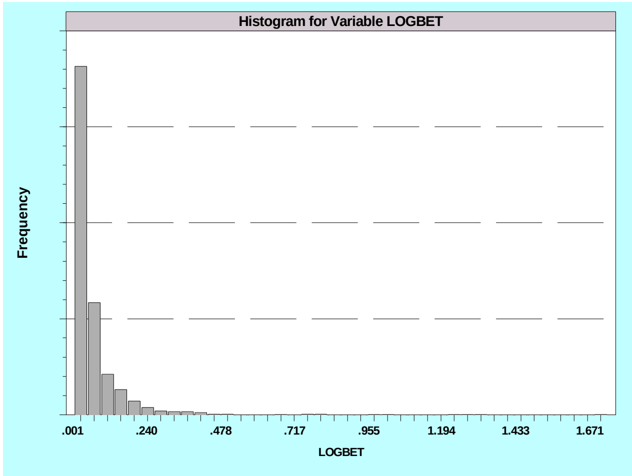


Figure 13: Distance to train station histogram



6 Discussions and conclusions

From the Stated and Revealed Preference data collected in the pilot study, different models were derived to gain an understanding of station access mode choice and residential location choice. Although the data only comes from a small sample of university staff and PhD students, the modelling results demonstrate the possibility of obtaining meaningful explanations from designed hypothetic choices. The distance from a residential house to train station was a significant contributor to both station access mode choice and residential location choice. The majority of participants preferred car access to the station which is consistent with the local car dominant habit. The 27 per cent of participants in the multi-mode class who shift their choice of modes might be the group of residents that government development policy should focus on first to meet their requirements for walking, cycling or public transport. The transit class which includes 27 per cent of the participants would like to live close to public transport regardless of the house type provided. This suggests that there may be a possibility to create a TOD using semi-detached houses or apartments/units around a train station with easy access to shops, schools and parks that could satisfy the needs of this class. In the case study area, the University of South Australia's Mawson Lakes Campus is in the core of Mawson Lake region. Further study on the travel behaviour of university staff and students will be one of most interesting outputs for local region development. We expect more valuable results deriving from real survey data. These results proved that the designed hypothetical scenarios are able to capture useful information, the Bayesian prior parameters are properly estimated and the models derived from the scenarios could indicate variable significance and weight.

From this experiment design, it can be concluded that establishing a research objective and understanding the study site in a realistic situation is the most important step to start for the design. With regard to transport and land use development, once a clear research question is defined, local data such as census data, train station observations and focus groups provide valuable fundamental information to help to define a minimised number of presentable choice alternatives, attributes and levels. A less biased assessment of attribute levels and the shares of each level are the core steps to estimating prior parameters. Experiment design efficiency relies on attributes and levels and prior parameters from where to achieve optimised efficient criteria, D-error, S-estimates and A-error. Efficiency criteria can also be sensitive to the selection of computing algorithms, in which Gaussian draws outperform other draws in efficiency. The A-error can also be reduced by changing the level coding and error components standard deviation.

Table 5: The results of residential location choice model of MNL, LCM , and RPM models

	MNL			**LCM**-Class1			**LCM**-Class2			**LCM**-Class3			**MMNL**-RPM		
Variable	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.	Coef	t-ratio	Sig.
													Random parameters		
House affordability	0.207	10.609	***	0.605	3.710	**	2.214	3.560	**	0.075	3.446	**	0.242	4.131	***
Distance to train station	0.079	3.956	**	-0.019	-0.214		0.966	3.787	**	0.004	0.172		-3.448	-4.129	***
													Nonrandom parameters		
Choice A	0.185	1.366		-2.051	-1.964	*	-3.518	-2.294	*	0.656	4.352	***	0.262	1.718	
Choice B	0.255	1.853		-0.363	-0.398		-2.165	-1.696		0.460	2.913	**	0.315	1.830	
Daily activity	0.033	0.644		-0.700	-3.045	**	-0.246	-0.891		0.031	0.327		-0.027	-0.211	
Distance to bus station	0.004	0.199		0.267	1.768		0.320	2.012	*	0.011	0.579		0.016	0.695	
House type	-0.357	-7.478	***	-2.803	-4.859	***	0.604	1.835		-0.229	-5.672	***	-0.359	-7.527	***
Income	0.000	-0.002		0.094	0.547		0.188	0.560		0.019	0.216		0.015	0.179	
Choice None	1.532	1.754		-8.373	-1.775		53.149	3.195	**	-1.784	-1.612		1.691	1.481	
Number of people in dwelling	-0.088	-0.679		-0.602	-1.703		-2.247	-2.279	*	0.237	1.001		-0.152	-0.561	
Park distance	0.044	2.298	*	-0.328	-1.895		1.219	3.474	**	-0.056	-2.424	*	0.059	2.929	**
School	0.047	2.256	*	0.027	0.237		1.518	3.173	**	-0.089	-3.757	**	0.059	2.687	**
Shops	0.052	2.723	**	-0.013	-0.126		1.637	3.073	**	-0.015	-0.770		0.064	2.912	**
Work place	0.085	2.019	*	-0.208	-0.658		0.319	1.871		0.106	2.663	**	0.127	2.668	**
				Class assignment											
Constant				-0.344	-0.503										
Daily activity				-0.083	-1.074										
Number of people in dwelling				0.107	0.507										
													Standard deviations of parameter distributions		
Ns House affordability													0.107	4.137	***
Ls Distance to train station													1.140	2.673	**
													Heterogeneity in mean		

House affordability: Daily activity					0.005	0.706	
House affordability: Number of person in dwelling					-0.006	-0.330	
Distance to train station: Daily activity					0.039	0.525	
Distance to train station: Number of person in dwelling					0.148	0.768	
Log likelihood function	-577.1		-486.2			-561.1	
Info. Criterion: AIC	2.096		1.894			2.061	
Finite Sample: AIC	2.097		1.911			2.063	
Info. Criterion: BIC	2.204		2.263			2.214	
Info. Criterion: HQIC	2.138		2.038			2.121	
Restricted log likelihood			-781.9			-781.9	
McFadden Pseudo R-squared			0.378			0.282	
Chi squared	285.5	[11]	591.3	[48]		441.5	[44]
Prob [ChiSqd > value]	0.000		0.000			0.000	
At start values -577.0860			0.157			0.028	*

Notes:

- * significant p value =0
- ** significant p value<0.01
- *** significant p value <0.05
- Parentheses indicates degree of freedom for Chi squared estimate

An additional output of the pilot study was that more than 30 respondents out of the 50 answered survey forms provided additional comments. The majority of these comments were made to improve the survey questions clearer and easier to understand. There were two comments that suggested the choice scenarios were too long and made them feel fatigued. However many other respondents commented that 'when I finished the first scenario, the later ones were pretty easy to answer'. This agrees with the conclusion of a fatigue effects study by Hess, Hensher and Daly (2012), that there is no conclusive evidence relating to respondent fatigue, but more that respondents undergo a learning process.

This design results in collecting valuable answers and building an explainable model output, but there are some limitations. A minimum attribute level overlap design still results in an approximate level balance. It might be possible for future experimental designs to compare the efficiency of designs between an unbalanced attribute level design with a lower number of choice sets and a balanced attribute level design with a higher number of choice sets. In relation to transport and land development studies, a panel model might be useful to observe the generation of a sample of respondents over several time periods (Train 2003; Baltagi 2009), to investigate the development of a TOD over time.

Further study will invite residents in the Adelaide northern rail corridor to answer the survey form both in paper form and online. A similar process will be followed to what was used in the pilot study, but with a wider range of variables and deeper analysis and statistical testing. The behavioural patterns observed could be extended from the sample to the whole population to provide policy suggestions for local transport and land planning. Some of the results such as the variable's histograms could be valuable for a same location study in further transport research.

Acknowledgement and disclaimer

This study is part of an Australian Research Council (ARC) Linkage Project supported by a committee of industry partners including the Land Management Corporation, the Town of Gawler, City of Playford, City of Salisbury, Delfin Lend Lease and the Department of Planning, Transport and Infrastructure (DPTI). The authors would like to express appreciations to the project steering committee and to project team members Professor Steven Hamnett, Drs Andrew Allan, Sekhar Somenahalli and Nikolaos Vogiatzis, and Messers Branko Stazic and Ivan Iankov.

The paper and its conclusions represent the views of the authors, and do not necessarily represent the views of DPTI, LMC and the South Australian Government, or those of the Cities of Playford and Salisbury, the Town of Gawler and Delfin Lend Lease. DPTI, its staff or the South Australian Government along with all of the other industry partners cannot be held responsible for any material contained within the paper, and are not liable for any loss or damage suffered by anyone who uses or relies on the information contained in the paper.

References

Australian Bureau of Statistics (2006) *Method of travel to work by sex*, viewed 10th August, 2010, <http://www.abs.gov.au/>.

Baltagi, BH (2009) *A Companion to Econometric Analysis of Panel Data*, John Wiley & Sons, Bayes, T (1763) 'An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS', *Philosophical Transactions,* vol. 53, p. 370.

Berger, JO (1985) *Statistical decision theory and Bayesian analysis*, 2nd edn, Springer-Verlag, New York.

Bhat, CR (2001) 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model', *Transportation Research Part B,* vol. 35, no. 7, pp. 677-693.

Bliemer, MCJ, Rose, JM & Hess, S (2008) 'Approximation of bayesian efficiency in experimental choice designs', *Journal of Choice Modelling,* vol. 1, no. 1, pp. 98-127.

Bliemer, MCJ & Rose, JM (2009) 'Efficiency and sample size requirements for stated choice experiments', *TRB 88th Annual Meeting*.

Box, GEP & Lucas, H (1959) 'Design of experiments in non-linear situations', *Biometrika,* vol. 46, no. 1/2, pp. 77-90.

Caussade, S, Ortúzar, JdD, Rizzi, LI & Hensher, DA (2005) 'Assessing the influence of design dimensions on stated choice experiment estimates', *Transportation Research Part B,* vol. 39, no. 7, pp. 621-640.

Chaloner, K & Verdinelli, I (1995) 'Bayesian Experimental Design: A review', *Statistical Science,* vol. 10, no. 3, pp. 273-304.

Choice Metrics Pty Ltd (2009) *Ngene 1.0 User Manual & Reference Guide*

DeShazo, JR & Fermo, G (2002) 'Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency', *Journal of Environmental Economics and Management,* vol. 44, no. 1, pp. 123-143.

Dhar, R & Simonson, I (2003) 'The effect of forced choice on choice', *Journal of Marketing Research*, pp. 146-160.

Fisher, RA (ed.) (1935) *The design of experiments*, Macmillan Pub Co,

Fowkes, T & Wardman, M (1988) 'The Design of Stated Preference Travel Choice Experiments: With Special Reference to Interpersonal Taste Variations', *Journal of Transport Economics and Policy,* vol. 22, no. 1, pp. 27-44.

Greene, WH & Hensher, DA (2003) 'A latent class model for discrete choice analysis: contrasts with mixed logit', *Transportation Research Part B,* vol. 37, no. 8, pp. 681-698.

Hensher, DA, Rose, JM & Greene, WH (2005) *Applied Choice Analysis: a Primer*, Cambridge University Press, Cambridge, UK.

Hensher, DA & Rose, JM (2007) 'Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study', *Transportation Research Part A,* vol. 41, no. 5, pp. 428-443.

Hess, S, Hensher, DA & Daly, A (2012) 'Not bored yet – Revisiting respondent fatigue in stated choice experiments', *Transportation Research Part A,* vol. 46, no. 3, pp. 626-644.

Huber, J & Zwerina, K (1996) 'The importance of utility balance in efficient choice designs', *Journal of Marketing Research,* vol. 33, no. 3, pp. 307-317.

Jaeger, SR & Rose, JM (2008) 'Stated choice experimentation, contextual influences and food choice: A case study', *Food Quality and Preference,* vol. 19, no. 6, pp. 539-564.

Kanninen, BJ (2002) 'Optimal design for multinomial choice experiments', *Journal of Marketing Research,* vol. 39, no. 2, pp. 214-227.

Kessels, R, Jones, B, Goos, P & Vandebroek, M (2009) 'An efficient algorithm for constructing Bayesian optimal choice designs', *Journal of Business and Economic Statistics,* vol. 27, no. 2, pp. 279-291.

Lancsar, E, Louviere, J & Flynn, T (2006) 'Several methods to investigate relative attribute impact in stated preference experiments', *Social Science & Medicine,* vol. 64, no. 8, pp. 1738-1753.

Louviere, JJ (1988) 'Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity', *Journal of Transport Economics and Policy,* vol. 22, no. 1, pp. 93-119.

Louviere, JJ, Hensher, DA & Swait, JD (2000) *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, UK.

McFadden, D (1972) 'Conditional logit analysis of qualitative choice behaviour', in Zarembka, P (ed), *Frontiers in Econometrics*, Academic Press, New York, pp. 105-142.

McFadden, D (1974) 'The measurement of urban travel demand', *Journal of Public Economics,* vol. 3, no. 4, pp. 303-328.

McFadden, D (1978) 'Modelling the choice of residential location', *Cowles Foundation Discussion Papers,* vol. 477.

McFadden, D & Train, KE (2000) 'Mixed MNL models for discrete response', *Journal of Applied Econometrics,* vol. 15, no. 5, pp. 447-470.

Meng, L, Holyoak, N & Taylor, MAP (2011) 'Increasing the Patronage of Adelaide's Northern Rail Corridor', *Papers of the 34th Australasian Transport Research Forum*, Adelaide, <http://patrec.org/atrf.aspx>.

Murphy, AH & Winkler, RL (1970) 'Scoring rules in probability assessment and evaluation', *Acta Psychologica,* vol. 34, pp. 273-286.

Olaru, D, Smith, B & Taplin, JHE (2011) 'Residential location and transit-oriented development in a new rail corridor', *Transportation Research Part A,* vol. 45, no. 3, pp. 219-237.

Ortúzar, JdD & Willumsen, LG (2002) *Modelling transport*, 3rd edn, Wiley, New York.

Rose, JM & Bliemer, MCJ (2005) *Constructing efficient choice experiments*, Institute of Transport and Logistics Studies, The University of Sydney.

Rose, JM, Bliemer, MCJ, Hensher, DA & Collins, AT (2008) 'Designing efficient stated choice experiments in the presence of reference alternatives', *Transportation Research Part B,* vol. 42, no. 4, pp. 395-406.

Sándor, Z & Wedel, M (2001) 'Designing conjoint choice experiments using managers' prior beliefs', *Journal of Marketing Research,* vol. 38, no. 4, pp. 430-444.

Sándor, Z & Wedel, M (2002) 'Profile construction in experimental choice designs for mixed logit models', *Marketing Science,* vol. 21, no. 4, pp. 455-475.

Tarokh, V, Jafarkhani, H & Calderbank, AR (1999) 'Space-time block codes from orthogonal designs', *IEEE Transactions on Information Theory,* vol. 45, no. 5, pp. 1456-1467.

Train, KE (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK.

Van Lenthe, J (1993) 'ELI: an interactive elicitation technique for subjective probability distributions', *Organizational Behavior and Human Decision Processes,* vol. 55, no. 3, pp. 379-413.

Walker, JL & Li, J (2007) 'Latent lifestyle preferences and household location decisions', *Journal of Geographical Systems,* vol. 9, no. 1, pp. 77-101.

Zwerina, K, Huber, J & Kuhfeld, WF (1996) *A general method for constructing efficient choice designs*, Durham, NC: Fuqua School of Business, Duke University.