ITLS

WORKING PAPER

ITLS-WP-18-16

How to better represent preferences in choice models: the contributions to preference heterogeneity attributable to the presence of process heterogeneity

By
Camila Balbontin, David A. Hensher and Andrew T. Collins

Institute of Transport and Logistics Studies, The University of Sydney, Australia

September 2018

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES
The Australian Key Centre in
Transport and Logistics Management

The University of Sydney
Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-18-16

TITLE: **How to better represent preferences in choice models: the contributions to preference heterogeneity attributable to the presence of process heterogeneity**

ABSTRACT: Discrete choice studies, with rare exception, assume that agents act as if sources of observed utility are captured through a linear in parameters and additive in attributes (LPAA) form, with some interactions. A growing number of transport (and other) choice studies have investigated one or more alternative processing rules adopted by agents in arriving at a choice, raising interest in how best to represent the utility expressions in a joint process and outcome choice model. Given the popular and appealing random parameter treatment of LPAA in mixed logit as a way of identifying non-systematic preference heterogeneity in a sample, this paper considers the possibility that we might be able to interact specific process heuristics with LPAA to uncover sources of systematic preference heterogeneity hidden in the standard LPAA form, and hence establish a link between the LPAA form and candidate process heuristics, offering a way to embellish and hence clarify the contributions to preference heterogeneity attributable to the presence of process heterogeneity. Specifically, we are interested in the extent to which there is a systematic relationship between the simple LPAA form and the more complex (albeit behaviourally realistic) process heuristics emerging in the transport literature which we call conditioning by random process heterogeneity (CRPH). In this paper, in addition to LPAA, we consider two process heuristics - Value Learning, and Relative Advantage Maximisation - with an overlay to account for risk attitudes, perceptual conditioning, and overt experience. The findings, using two data sets, suggest that empirically there exists a significant attribute-specific relationship between preference heterogeneity identified through specific process heuristics and through the LPAA assumption.

KEY WORDS: ***Travel behaviour; preference heterogeneity; multiple heuristics; risk attitudes; perceptual conditioning; value learning***

AUTHORS: **Balbontin, Hensher and Collins**

Acknowledgment: This research was funded by ARC Discovery Project (DP): DP140100909: Integrating Attribute Decision Heuristics into Travel Choice Models that accommodate Risk Attitude and Perceptual Conditioning. It was also partially funded by the CONICYT PFCHA/DOCTORADO BECAS CHILE/2017 - 72150522. The authors acknowledge the facilities, and the scientific and technical assistance of the Sydney Informatics Hub at the University of Sydney and, in particular, access to the high performance computing facility Artemis. We also thank Chandra Bhat for his comments on an earlier version.

CONTACT:	INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73) The Australian Key Centre in Transport and Logistics Management The University of Sydney NSW 2006 Australia Telephone: +612 9114 1824 E-mail: business.itlsinfo@sydney.edu.au Internet: http://sydney.edu.au/business/itls
DATE:	September 2018

1 Introduction

Discrete choice modelling has, in the main, focussed on a simplified functional form for the way in which the observed sources of utility are represented. It is typically assumed that agents act *as if* they are utility maximisers under a linear in the parameters and additive in the attributes (LPAA) paradigm. This simplified but appealing context independent form is unambiguously consistent with the Random Utility Maximisation (RUM) model – see Hess et al. (2018) for a comprehensive review of compliance (and approximate compliance) with RUM. It is also typically assumed that agents when making a choice act *as if* they are taking into account all the attributes and alternatives presented to them in a stated choice experiment or obtained from questions on revealed preferences, and value the attribute levels exactly as presented in a choice experiment or as reported in a series of revealed preference questions. A growing number of studies have questioned the behavioural richness of this paradigm, investigating the role of a range of process heuristics in choice modelling, including multiple heuristics (see Hensher et al. 2015 for a review). However, the existing literature remains limited even though there are commonly suggestions that the consideration of multiple heuristics, which is referred to in this study as process heterogeneity, significantly improves the goodness of fit of choice models, and provides a richer understanding of individuals' behaviour.

Given the popularity and ease of estimating a RUM model under LPAA with random parameters and obtaining estimates of willingness to pay, but also the interest in a number of alternative process rules that may be used by agents to assess the attributes defining choice options and arrive at a choice, we wonder whether there is an informative behavioural relationship between the two. Namely, can the appealing LPAA specification that accounts for preference heterogeneity through random parameters (hence a non-systematic allocation of agents over the preference distribution), which in a sense hides potentially important behavioural information, be a further representation of preference heterogeneity captured by specific process rules (or heuristics) that can reveal systematic sources of influence? This is effectively a created interaction between the utility identified under LPAA preference heterogeneity (i.e., non-systematic random parameters) and additional sources of utility identified through process heterogeneity which we call conditioning by random process heterogeneity (CRPH) (Balbontin et al., 2017a). A number of previous studies have also suggested (but not tested for) a possible interaction between preference heterogeneity embedded in process heuristics and that captured in the simplified LPAA form (Hess et al., 2012; Hensher et al., 2013a; Collins et al., 2013; Campbell et al. 2014).

LPAA and two process heuristics will be the focus of this paper as well as three additional behavioural refinements (see below). The first process heuristic considers that competing alternatives influence the way an individual assesses an alternative (referred to as the local choice context dependent heuristics). Specifically, we consider Relative Advantage Maximisation (RAM) where an individual will ponder the advantages and disadvantages of an alternative relative to the competing alternatives as a function of the difference between the alternatives' attributes. The other process heuristic considers that, when individuals are faced with more than one choice sequentially, the previous scenarios have an effect on their current choices (referred to as choice set interdependent heuristics). We include this effect through a Value Learning (VL) heuristic which assumes that the best attribute levels of the alternatives that have been previously chosen will have an effect on how an individual assesses the alternatives in a current scenario.

Traditional choice studies also usually assume that respondents act *as if* they are neutral to risk (i.e., indifferent towards a risky alternative or a sure alternative of equal expected value), although respondents can be risk averse or risk takers. Choice studies also often consider that respondents perceive the levels of attributes in choice experiments in a way that suggests the absence of perceptual conditioning. However, heterogeneity is also present for perceptual conditioning in cases where there is variability in the outcomes of an attribute(s), which allows for differences between the stated probability of occurrence (in a choice experiment) and the perceived probability used when evaluating the prospect. Finally, the (accumulated) overt experience that individuals' have with each alternative might also influence their decisions.

In summary, the motivation of this study is to explore potential relationships between LPAA, process heuristics, random parameters, behavioural refinements and experience, as contributions to revealing preference heterogeneity within a sampled population. The concept of process homogeneity refers to the assumption that all respondents use one process strategy, which in this study refers to either VL or RAM; the LPAA is considered as the reference model. Alternatively, process heterogeneity refers to a situation when more than one process strategy might be being used across respondents in decision-making, together with LPAA. Preference heterogeneity will be included in this study in the form of random parameters, and preference homogeneity will be represented by fixed parameters. The proposed model form, that includes all of the features defining CRPH will be compared to (i) a model that considers the presence of either LPAA or one process strategy in the heterogeneous (i.e. random parameter) or homogeneous (i.e., fixed parameter) form and (ii) a model that allows for LPAA together with process heterogeneity (i.e., multiple process strategies) through fixed parameters under a latent class functional form (i.e., implicitly assuming preference homogeneity). This latter is referred to as the *probabilistic decision process* (PDP) method and it will be the only other approach – besides CRPH – that will consider multiple process strategies (i.e., process heterogeneity) so it will be very relevant for comparison between them and with the mixed and MNL logit models.

The paper is structured as follows. In Section 2 we present the process heuristics of interest and provide additional commentary on the interactive nature of such heuristics with the more commonly identified preference heterogeneity under LPAA. Section 3 details the model forms and how the various behavioural elements are integrated and the derived willingness to pay estimates. We then overview the two data sets in section 4, followed by the estimated models and their interpretation and implications in section 5. The concluding section summarises the main findings of the paper and directions for ongoing research.

2 Background

Mainstream discrete choice modelling approaches have evolved in a setting in which some very specific behavioural assumptions are made in specifying choice models and estimation methods. While they have served the literature well and are often the 'bread and butter' procedures in practical applications, they are not without question. The great majority of choice studies assume that decision makers act *as if* they are rational (Luce, 1959), take into account all the attributes included in a stated choice experiment or as listed in a revealed preference model, and value the levels exactly as are presented to them. The attributes are also assumed to be adequately represented as sources of (relative) utility under a linear in the parameters and additive in the attributes (LPAA) assumption, with allowable linear interactions. Criticisms of these assumptions as behaviourally appropriate, especially in psychology and marketing research, have led to the development of a growing number of process strategies (or rules or heuristics) as possible alternative ways of representing choice making.

The spectrum of process rules or heuristics that have been presented in the literature can be divided into three categories: (1) those that are independent of the alternatives presented in the choice set – context free heuristics; (2) those that depend on all the alternatives shown in a choice set – local choice context dependent; and (3) those that depend on the multiple choice tasks shown to an individual – choice set interdependent. The focus of this study is on studying the role of LPAA and two heuristics: Relative Advantage Maximisation (the second category); and Value Learning (the third category).

2.1 Extremeness Aversion Heuristics: Relative Advantage Maximisation

Simonson and Tversky (1992) propose two hypotheses of how context might influence respondent decisions. The first hypothesis states that the attractiveness of an alternative depends on whether the trade-offs within the choice set are favourable towards that alternative. The second hypothesis, which they refer to as *extremeness aversion*, states that an alternative is more attractive if it is an intermediate option within the choice set. Hence, the extreme options are less attractive to respondents. They define the extreme alternatives within a choice set as the ones that have high advantages and high disadvantages relative to each other, and have small advantages and small disadvantages relative to the intermediate alternative. The intermediate alternatives are the ones that have small advantages and small disadvantages relative to the extremes (there are two extreme alternatives). Simonson and Tversky (1992) developed the *extremeness aversion* heuristic within the framework of loss aversion, according to which individuals assign a higher weight to losses than they do to gains (Tversky and Kahneman 1991). The losses or gains of an alternative are measured using a reference point, which in this case is considered to be the other alternatives presented in the choice set. The extremeness aversion heuristic has been implemented in papers such as Chernev (2004); Sharpe et al. (2008); Leong and Hensher (2012a); Hensher et al. (2018), and several theoretical models accommodate its effects, such as random regret minimisation, RRM (Chorus et al., 2008; Chorus, 2010; Chorus, 2012; van Cranenburgh et al., 2015), and relative advantage maximisation (RAM).

The RAM model was introduced by Tversky and Simonson (1993) to consider how individuals compare the attribute levels across alternatives taking into account the compromise and polarisation effect, which they referred to as the ‘componential context model’. They suggest that each attribute level is an advantage or disadvantage relative to the other alternatives, and therefore, the utility function for each alternative is the sum of its advantages and disadvantages. Kivetz et al. (2004) first refer to this model formulation as the ‘relative advantage model’. Their starting point is the Tversky and Simonson (1993) model described above, defining the disadvantage of an alternative with respect to another alternative as an increasing convex function.

Leong and Hensher (2014, 2015) propose a new version of the RAM model, which has some desirable properties from the classical RRM functional forms, and has the symmetry between advantage and disadvantage as proposed by Tversky and Simonson (1993). This is the form used in this study and the utility functions will be presented in section 3.2.

As was initially proposed by Tversky and Simonson (1993), RAM required at least three alternatives. However, Leong and Hensher (2014) show that, oppositely to the RRM model, the RAM model can still take into account the relative advantage/regret effects in binary choice data. The Leong and Hensher (2015) results show that the differences in these models’ fit were quite small. However, they state that this decision process requires further investigation and has the real potential to improve the performance of choice models.

2.2 Reference Points: Value Learning

Several studies have proposed decision rules under which alternatives are evaluated relative to a reference point. However, the use and/or origin of the reference point changes across the proposed processing rules. For example, the reference point might be used as a rational comparison to evaluate the attribute levels of the alternative presented, as is the case of the 'reference revision' heuristic (DeShazo, 2002; Day and Pinto, 2010; McNair et al., 2011), and 'value learning'. It could also be considered as past experiences influencing the expected outcome, referred to as 'Case Based Decision Theory' (Gilboa and Schmeidler, 1995; Gilboa and Pazgal, 2001), or it might be considered as "real" market levels to evaluate the fairness of the alternatives, as the 'Cost Expectations Model' (Carson et al., 1994; Alberini et al., 2017).

The value learning heuristic defines a situation where the principle of preference stability is violated, and assumes that throughout the experiment, individuals discover their preferences. This heuristic underlies a theory that individuals have weak preferences and can be influenced by the alternatives shown to them. This was originally proposed by Plott (1996) and was later analysed in a discrete choice experiment context by Bateman et al. (2006), Bateman et al. (2008), McNair et al. (2011), and Hensher and Collins (2011). Although they have similarities, there is not one common way of incorporating this heuristic. For instance, Bateman et al. (2006) state that repetition was the key for learning, and so the starting point bias, i.e., the comparison to the initial level, is reduced as the respondent is presented with more decisions or choice sets. McNair et al. (2011) test whether choices are affected when a respondent faces four choice tasks instead of one, and their focus is in the cost attribute. Hensher and Collins (2011) incorporate this heuristic by including a dummy variable that is equal to 1 if the alternative was chosen in the previous choice set, and 0 otherwise.

2.3 Multiple Heuristics or Process Heterogeneity

The evidence reviewed thus far suggests that the traditionally used LPAA paradigm is not always adequate, and decision process strategies are sometimes better at describing decision-making. However, the majority of the literature on alternative heuristics has focussed on explaining decision-making using only one decision process strategy. Some studies have studied the possibility that individuals might be using different heuristics in decision making by including multiple heuristics in the modelling (e.g., Hensher et al., 2018), even though the literature on this topic is rather limited. One of the most common approaches used in literature to incorporate process heterogeneity is the probabilistic decision process approach (PDP), which will be explained in more detail in the methodology section.

The current literature on multiple heuristics, also referred to as process heterogeneity when more than one heuristic is included in the model, suggests significant improvements in the statistical performance of choice models (Hensher and Collins, 2011; McNair et al., 2012; Leong and Hensher, 2012b; Balbontin et al., 2017; Balbontin et al. 2017a); however this topic is relatively new and evidence is still accumulating. The consideration of multiple heuristics has helped researchers to further understand individual behaviour by differentiating the influence of several heuristics. The majority of these studies have shown significant influences on the WTP estimates (in their mean and standard deviation). However, some studies did not find significant differences compared to a Mixed Multinomial Logit (MML) model under LPAA. Nevertheless, the choice set specific preferences suggested by alternative heuristics produce different behavioural insights which lead to a richer interpretation of the trade-offs which is equally as relevant in decision-making.

2.4 Interaction between Process and Preference Heterogeneity

The relationship between process and preference heterogeneity has been mentioned numerous times in choice studies; however it has rarely been studied in detail. Collins (2012) demonstrated the biasing impact that attribute non-attendance (ANA) behaviour can have on the mean and the variance of random parameters through simulated data, suggesting a highly dependent relationship between the ANA process strategy and random parameters. Hensher et al. (2013a) included two process rules in a mixed latent class model structure. Their results showed a small improvement in statistical performance when adding the random parameters into the latent class structure. In their conclusions, they mention a possible 'confoundment' between attribute processing - as a mechanism to reveal process heterogeneity - and random parameters. Campbell et al. (2014) studied one process strategy and the LPAA paradigm using a latent class approach where they acknowledge a possible confounding of preference and process heterogeneity, but do not test it any further.

Hess et al., (2012) propose latent mixed logit models, where the first class considers the traditional RUM heuristic and the second class is a process strategy (i.e., they test different heuristics in the second class, one at a time). Their results show that when allowing for random parameters in the RUM class, there is a reduction in the share of the other class. Their results also show a decrease in the degree of random heterogeneity in the RUM class compared to a simple multinomial mixed logit model by itself as a one class model. They suggest that what might be retrieved as taste heterogeneity in traditional models may be heterogeneity in decision rules, leading to a questioning of whether there is confoundment, which they suggested should be investigated further in future research.

Balbontin et al. (2017a) propose a method to incorporate an interaction between the traditional LPAA estimates of an attributes' mean and standard deviation parameters – and the properties of the VL heuristic, which we refer to as Conditioning by Random Process Heterogeneity (CRPH). As mentioned above, the current paper extends the CRPH method to include more than one process strategy (while extending the definition of VL which will be seen in the next section) and to incorporate, risk attitudes and perceptual conditioning, and overt experience. Moreover, this study defines the method necessary to calculate the willingness to pay estimates with their confidence intervals in such models.

This section has provided a short review of what has been conducted in the field of process heuristics and other behavioural refinements. The growing number of studies have shown a significant influence on the model estimates and the interpretation derived from them. However, there are still many research gaps related to process heterogeneity when more than one heuristic is included (i.e. the idea of multiple heuristics), especially considering their relationship with random parameters (under LPAA in particular), experience, risk attitudes and perceptual conditioning, which is the focus of this paper.

3 Proposed Method

As the literature on process heuristics grows in interest within a discrete choice setting, and especially where preference heterogeneity is increasingly accommodated by a random parameter specification, the question arises as to whether there is a systematic relationship between random parameters as a representation of preference heterogeneity under LPAA and one or more process heuristics. Specifically, is there a relationship between preference heterogeneity and process heterogeneity such that process heterogeneity, as represented by specific multiple heuristics, conditions the distribution of preferences in a sampled population in such a way that it adds a systematic (in contrast to random) explanation of preference heterogeneity? Effectively, is a treatment under a process heuristic aligned with LPAA as an interaction effect and thus a contribution to a richer specification of preference heterogeneity? And how does the interaction effect work together with behavioural refinements and overt experience?

The traditionally used LPAA model is set out within a theory of random utility (McFadden, 1974) which assumes that each individual behaves rationally and it is assumed that they act *as if* they choose the alternative that maximises utility. The utility of each alternative is defined as a function of its attributes, where each attribute is weighted by an estimated coefficient as a measure of the marginal (dis)utility associated with a specific attribute. We can describe this utility for each individual *q* with a utility function U_{iq} assigned to each alternative A_i, as described in equation (1). This utility function has a systematic component which is measurable, V_{iq}, that is a function of the alternatives' attributes measurable by the modeller. It also has a random component, $\varepsilon_{\mathrm{iq}}$, which allows for possible preference heterogeneity and for potential errors induced by the modeller (i.e., in measuring or observing) including excluded relevant attributes. It is important to understand that the modeller is likely only able to see or account for a subset of the attributes considered by sampled individuals when reaching a decision, and hence is not able to include all the attributes that could influence the respondent choice.

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{1}$$

The systematic part described above in its simplest form can be written as follows (Hensher et al., 2015; Ortúzar and Willumsen, 2011) in the LPAA form:

$$V_{iq} = \theta_{i1} \cdot x_{i1q} + \theta_{i2} \cdot x_{i2q} + \theta_{i3} \cdot x_{i3q} + \ldots + \theta_{in} \cdot x_{inq} \tag{2}$$

θ_{in} represents the parameter estimate (or marginal (dis) utility) associated with attribute *n*, which is assumed to be fixed for all individuals but can vary between alternatives i, and x_{inq} represents the level of attribute *n* of alternative *i* for the individual *q*. This is referred to as the multinomial logit model (MNL)

A more complex model that is widely used in transportation studies, is the mixed logit model (MML). The current form was proposed by two research groups: Ben-Akiva and Bolduc (1996) and McFadden and Train (2000) (see also Hensher and Greene 2003). The ML form consists of any model for which choice probabilities can be expressed as an open form (Ortúzar & Willumsen 2011):

$$E(P_{iq}) = \int P_{iq}(\theta) \cdot f(\theta) \; d\theta \tag{3}$$

$P_{iq}(\theta)$ represents the standard MNL probabilities evaluated at a set of parameters θ , and $f(\theta)$ is their density function (also known as a 'mixing distribution'). If the density function is degenerate at fixed parameters θ, the choice probability is equivalent to the MNL form shown above. If, however, the density function is discrete, the MML becomes a *latent class model*. This latter is the model form used by the PDP approach, which will be explained in section 3.4.

A specification of the MML model considers *random parameter estimates,* which is the base of the CRPH method that will be outlined below. In this case, the parameters are allowed to vary over individuals *q* (but not over choice sets *t*) with density function $f(\theta)$ in order to capture preference heterogeneity. Not all the parameters θ have to be considered random; some can be specified as preference homogeneous. The utility function of alternative *i* for individual *q* in choice situation *t* (*t*=1,…,*T*), where attribute *n* is defined as random, is defined as follows:

$$U_{iqt} = \theta_{inq} \cdot x_{inqt} + \ldots + \varepsilon_{iqt} \tag{4}$$

where the parameter θ_{inq} can be decomposed in its mean, θ_{in}^{m}, which is common for all individuals *q*, and standard deviation across the sample, σ_{in}; and a distribution v specified by the modeller:

$$U_{iqt} = \theta_{in}^{m} \cdot x_{inqt} + \sigma_{in} \cdot v \cdot x_{inqt} + \ldots + \varepsilon_{iqt} = \left(\theta_{in}^{m} + \sigma_{in} \cdot v\right) \cdot x_{inqt} + \ldots + \varepsilon_{iqt} \tag{5}$$

The main objective of this paper is to estimate and compare models that consider preference revelation under: (a) LPAA or one process rule (i.e., process homogeneity treatment) that assumes preference homogeneity (fixed parameters), as MNL models; (b) LPAA or one process rule (i.e., process homogeneity treatment) assuming preference heterogeneity (random parameters), as MML models; (c) LPAA and multiple process rules assuming preference homogeneity (fixed parameters), which is the PDP (latent class) model; and (d) LPAA and multiple process rules assuming preference heterogeneity, which is the CRPH model. All the models analysed in this study include behavioural refinements and experience[1]. Table 1 and Table 2 provide a description of the LPAA and process homogeneity (models 'a' and 'b') and process heterogeneity models (models 'c' and 'd'), respectively, together with the acronyms that will be used to refer to them in the section 5. The two models that include LPAA and multiple heuristics are the PDP and CRPH, and each one proposes a different way in which these multiple heuristics are being used (which will be explained in the sections below). Therefore, the comparison between our proposed CRPH model and the PDP model will be of particular interest as the PDP is the most common method in literature that has been used to include multiple heuristics.

Table 1: Description of standalone LPAA and four process homogeneity (i.e., a single heuristic) models

	LPAA	**RAM**	**VL**	**Parameters***	**Behavioural refinements**	**Experience**
LPAA_MNL	Yes	No	No	Fixed	Yes	Yes
LPAA_MML	Yes	No	No	Random	Yes	Yes
VL_MNL	No	No	Yes	Fixed	Yes	Yes
VL_MML	No	No	Yes	Random	Yes	Yes
RAM_MNL	No	Yes	No	Fixed	Yes	Yes
RAM_MML	No	Yes	No	Random	Yes	Yes

*If the parameters are fixed, then the model accommodates preference homogeneity, and if they are random then it accommodates preference heterogeneity.

Table 2: Description of the process heterogeneity (i.e., multiple heuristics) models: including LPAA, RAM and VL

	Parameters*	**Behavioural refinements**	**Experience**	**Description**
PDP	Fixed	Yes	Yes	LPAA and each heuristic is represented by a different class
CRPH	Random	Yes	Yes	Interactions between the **mean** and **standard deviation** normally defined under LPAA, and the process strategies VL and RAM

*If the parameters are fixed, then the model accommodates preference homogeneity, and if they are random then it accommodates preference heterogeneity.

We now present the various components of an integrated model that captures the VL and RAM process heuristics, behavioural refinements and experience, together with testable interactions with the LPAA components. We include a brief explanation of the PDP approach, followed by a detailed explanation of the CRPH method that includes LPAA and the two three process strategies (VL and RAM). We also present the formulae used to obtain willingness to pay estimates including the form of the confidence intervals. The method is set out in the context of a stated choice experiment, given the data will be using, although all components can also be studied within a revealed preference setting.

[1] The behavioural refinements and experience included in each model are those found to be statistically significantly different from zero. The reader is referred to Balbontin (2018) for a detailed comparison of the models with and without behavioural refinements and experience.

3.1 Value Learning

Value Learning assumes that throughout a choice experiment individuals' preferences change. The reason for their preferences to change can be due to the attribute levels presented to them, often referred to as "rules of the market" (i.e., institutional learning), or simply because they are gaining knowledge and discovering their preferences. Both can be influencing preferences, and the way in which they change is defined in each study. We propose that when valuing the alternatives, individuals compare each of the alternatives' attribute levels to a reference level (De Borger and Fosgerau, 2008). For example, there is one reference level for the travel time (whether it is car or public transport) and individuals will compare the travel times of the alternatives to that reference level. Our value learning proposal is that when an individual faces a new decision, an attribute's reference level is updated only if the attribute level of the chosen alternative is better (i.e., preferred) than the current reference level. In a stated preference experiment, the starting reference level corresponds to the attribute levels of the mode they use in real life.

One of the simplest model formulations of Value Learning is to consider directly the difference between the attribute levels and reference level, without any type of transformation, as used in Balbontin et al. (2017b). In this case, the observed part of the utility function for alternative *i* can be written as follows:

$$U_{iq} = \theta_{i1} \cdot \left(x_{i1q} - ref_{1q} \right) + \theta_{i2} \cdot \left(x_{i2q} - ref_{2q} \right) + \ldots + \theta_{in} \cdot \left(x_{inq} - ref_{nq} \right) + \varepsilon_{iq} \tag{6}$$

θ_{in} are the estimates representing the difference between the level of attribute *n* and alternative *i* and the reference level for that same attribute *n*; x_{inq} represents the level of attribute *n* of alternative *i* and individual *q* ; and ref_{nq} represents the reference level for attribute *n* and individual *q*. This is a very simple model formulation and has some restrictions. For example, it collapses to a simple MNL model when the choice context is unlabelled, or when the same attributes are present in all the alternatives and their parameters are considered generic in a labelled experiment. In these cases, the reference levels will be the same for all the alternatives, and since the MNL models are estimated based on the differences, the reference levels will be nulled. Moreover, as has been widely mentioned in the literature, the valuation of gains and losses represented by $\left(x_{inq} - ref_{nq} \right)$ may not be linear. Therefore, in this study we incorporate a concavity factor φ to transform the differences between the attribute levels and the reference levels[2] as follows:

$$U_{iq} = \theta_{i1} \cdot \left(x_{i1q} - ref_{1q} \right)^{\varphi} + \theta_{i2} \cdot \left(x_{i2q} - ref_{2q} \right)^{\varphi} + \ldots + \theta_{in} \cdot \left(x_{inq} - ref_{nq} \right)^{\varphi} + \varepsilon_{iq} \tag{7}$$

The difference between the attribute levels and the reference levels can be positive or negative[3]. As φ has to be a continuous parameter (with decimal points), to avoid any estimation problems when the difference is negative, the transformation explained above will be considered as follows:

$$VL(\mathrm{x}_{inq}) = \begin{cases} \left(x_{inq} - ref_{nq} \right)^{\varphi} & \text{if } \left(x_{inq} - ref_{nq} \right) \geq 0 \\ -\left[-\left(x_{inq} - ref_{nq} \right) \right]^{\varphi} & \text{if } \left(x_{inq} - ref_{nq} \right) < 0 \end{cases} \tag{8}$$

[2] A different concavity factor can be estimated for gains and for losses. This was tested, but the results did not show statistically significant differences between the parameters, so the results consider a common concavity factor in gains and losses.
[3] We tested estimating a separate concavity factor for negative and positive differences, but the results showed that the difference was not statistically significant.

$$U_{iq} = \theta_{i1} \cdot VL(x_{i1q}) + \theta_{i2} \cdot VL(x_{i2q}) + \ldots + \theta_{in} \cdot VL(x_{inq}) + \varepsilon_{iq} \quad (9)$$

Figure 1 presents the Value Learning transformation for different values of φ, where $\varphi > 0$. As can be seen, φ acts upon the difference between the attribute levels, having a different influence over attributes that have larger differences between their levels than others that have smaller differences. The purpose of φ is to test whether the influence of VL in the utility function is non-linear, which will be considered common across attributes despite them having different units of measure. The θ_{in} parameters will vary according to the units of the attribute levels to represent the effect of the Value Learning transformation in the utility function.

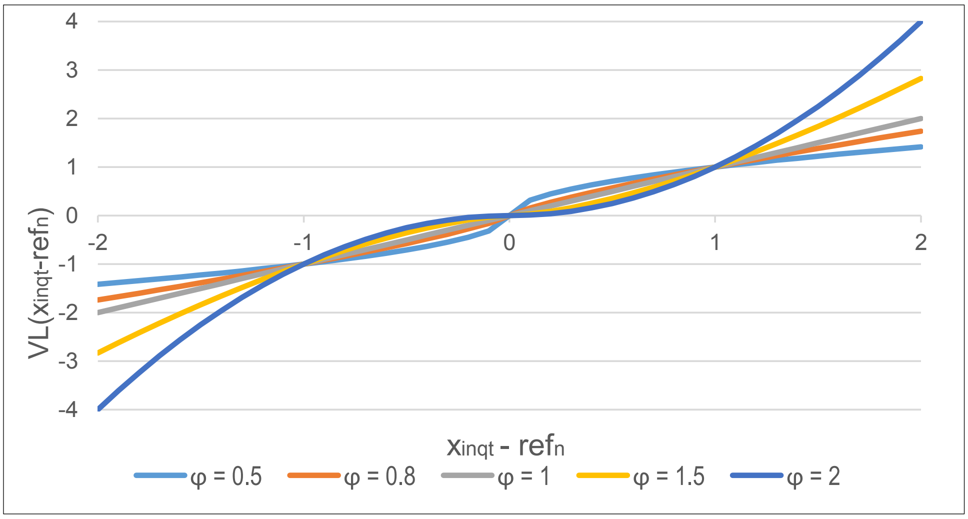


Figure 1: Value Learning transformation of attribute levels

3.2 Relative Advantage Maximisation (RAM)

The relative advantage maximisation (RAM) heuristic included in this study was proposed by Leong and Hensher (2014) and Leong and Hensher (2015), as mentioned in the background section. The authors define the disadvantage of alternative *i* relative *j*, $D(i, j)$, using the RRM formulation as follows:

$$D(i, j) = \sum_m \ln(1 + \exp(\theta_{jm} \cdot x_{jm} - \theta_{im} \cdot x_{im})) \quad (10)$$

where θ_{im} and θ_{jm} represent the estimates for attribute *m* in alternatives *i* and *j*, respectively; x_{im} and x_{jm} represent the level of attribute *m* of alternatives *i* and *j*, respectively.

With the symmetry condition between advantages and disadvantages, the advantage of alternative *j* over *i*, $A(j, i)$, can be expressed as follows:

$$A(j, i) = D(i, j) = \sum_m \ln(1 + \exp(\theta_{jm} \cdot x_{jm} - \theta_{im} \cdot x_{im})) \quad (11)$$

The relative advantage of *i* over *j*, denoted as $R(i, j)$ can be written as:

$$R(i, j) = \frac{A(i, j)}{A(i, j) + D(i, j)} \quad (12)$$

where the utility function of alternative *i* will be equivalent to the utility function under an LPAA heuristic plus the relative advantage of alternative *i* over all other alternatives *j* as follows:

$$U_{iq} = \sum_{m} \theta_{im} \cdot x_{imq} + \sum_{j \in S} R(i,j) + \varepsilon_{iq} \tag{13}$$

As can be noted, $A(\mathrm{j,i})$ and $D(i,\mathrm{j})$ are both positive; so the relative advantage of one alternative over another will always be a positive number lower than 1. That is why it is directly inserted in the utility function.

3.3 Behavioural Refinements and Experience

3.3.1 Behavioural Refinements

The models in this study consider risk attitudes and perceptual conditioning. Eeckhoudt et al. (2005) define risk aversion as "the rate at which marginal utility decreases when wealth is increased by 1%". In general, we consider that people are risk averse if their utility function is concave and risk takers if their utility function is convex (Kahneman and Tversky, 1979). In this paper, risk attitudes are specified in the form of Constant Relative Risk Aversion (CRRA) as follows:

$$u\left(x_{inqt}\right) = \frac{x_{inqt}^{\;1-\alpha}}{1-\alpha} \tag{14}$$

α represents the risk attitude towards attribute x_{inqt} [4]: if α = 0 then there is risk neutrality, if α < 0 then there is risk taking attitude, and if α > 0 then there is risk aversion.

The definition of perceptual conditioning refers to outcome conditioning by an attribute influenced by perception (Kahneman and Tversky, 1979). The weighting function $w(p)$ proposed by Kahneman and Tversky (1979) is a specific case of perceptual conditioning, where the outcome is conditioned by its probability of occurrence, and this probability is subject to perception (through the weighting function). The Tversky and Kahneman functional form has proven to be the most adequate in several discrete choice studies (Camerer and Ho, 1994; Hensher et al., 2011), and it will be the one used in this study as follows:

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \tag{15}$$

This function is an inverse S-shape where γ represents the degree of curvature of the weighting function and p the probability of occurrence. An estimated parameter γ with a value between 0 and 1 suggests that individuals will over-weight low belief probabilities and under-weight medium to high belief probabilities.

In this study, risk attitudes and perceptual conditioning will be included in the LPAA specification in two cases:

(1) Attributes that are presented with levels of variation: will take into account risk attitudes and perceptual conditioning.
(2) Attributes that do not vary across occurrences: will take into account risk attitudes only.

For explanation purposes, we consider a utility function that has three attributes: a cost attribute, x_{i1q} ; an attribute presented with *L* levels of variation, x_{i2q} ; and a third attribute with no levels of variation and which does not represent cost, x_{i3q} . The utility function can be written as:

[4] The special case where α=1, collapses to 1/0 which is undefined (it goes to infinity); however the occurrence of such a case is very unlikely.

$$U_{iq} = \theta_{i1} \cdot x_{i1q} + \sum_{l \in L} w(p_{i2q,l}) \cdot u(x_{i2q,l}) + \theta_{i3} \cdot x_{i3q} + \varepsilon_i \tag{16}$$

where $x_{i2q,l}$ is the level *l* of variation for attribute x_{i2q} and $p_{i2q,l}$ is the associated probability of occurrence; $w(p_{i2q,l})$ is the weighting function which will be referred as perceptual conditioning; and $u(x_{i2q,l})$ the value function which is referred to as risk attitudes.

The utility function initially under LPAA when considering behavioural refinements can be written as:

$$U_{iq} = \theta_{i1} \cdot x_{i1q} + \theta_{i2} \cdot \sum_{l \in L} \left[\frac{p_{i2q,l}{}^{\gamma}}{(p_{i2q,l}{}^{\gamma} + (1-p_{i2q,l})^{\gamma})^{1/\gamma}} \cdot \frac{x_{i2q,l}{}^{1-\alpha_2}}{1-\alpha_2} \right] + \theta_{i3} \cdot \left[\frac{x_{i3q}{}^{1-\alpha_3}}{1-\alpha_3} \right] + \varepsilon_{iq} \tag{17}$$

where γ represents the degree of curvature of the weighting function, and α represents the risk attitudes towards the second and third attribute.

3.3.2 Experience

The overt experience that individuals' have with specific alternatives has been shown to influence their decisions (Hensher, 1975; Goodwin, 1977; Ben-Akiva and Morikawa, 1990; Cantillo et al., 2007). We consider experience as the alternative chosen in their most recent revealed preference decision. In a transportation context, this refers to the most recent mode used by the respondent. It will be included by conditioning the utility function (Hensher and Ho, 2016; Balbontin et al. 2017a), as follows:

$$U_{iq}{}^{experience} = \left(1 + \theta_{exp,i} \cdot x_{experience,iq}\right) \cdot U_{iq} \tag{18}$$

where $U_{iq}{}^{experience}$ is the transformed utility function; $x_{experience}$ is defined as a dummy variable equal to 1 if alternative *i* was chosen by respondent *q* in their most recent decision and 0 otherwise; and $\theta_{exp,i}$ is the associated parameter. The utility function of mode *i* can be negative in one scenario and positive in another scenario, depending on its characteristics. Thereby, the modeller needs to define the parameter $\theta_{exp,i}$ with a consistent interpretation across scenarios, as follows:

$$\theta_{exp,i} = \begin{cases} \theta^0_{exp,i} & \text{if } U_{iqt} \geq 0 \\ -\theta^0_{exp,i} & \text{if } U_{iqt} < 0 \end{cases} \tag{19}$$

In transportation, one would expect that if an individual uses a certain mode he is more likely to choose it again. Hence, the hypothesis is that if the respondent used the mode in a most recent trip, he will scale the (dis)utility for that mode in such a way that it becomes (smaller) larger, so that $0 < \theta_{exp,i} < 1$. If the opposite is true, where an individual is less likely to choose the same alternative used in his most recent trip, then $-1 < \theta_{exp,i} < 0$. However, the absolute value of $\theta_{exp,i}$ has to be smaller than 1, otherwise it would imply a change in the sign of the utility when it is negative.

The model form is consistent with the overall utility maximisation assumption with the conditioning component (referred to as heteroscedastic conditioning) derived from information associated with the variance of the unobserved effects, effectively delimiting an observable and systematic way of accommodating heterogeneous scaling (see Swait and Adamowicz, 2001; Hensher and Ho, 2016 for details). This approach was preferred over including experience as an additional variable in the utility function, as it suggests that experience affects all the attributes that describe the experienced mode, and is not independent to them.

The model formulation also introduces correlation between choice sets, taken into consideration the panel nature of the stated choice data. This treatment relaxes the assumption of independence across choice sets within a respondent. As this conditioning is a form of scaling of utility, it will not change the distribution of the error term(s).

3.4 Probabilistic Decision Process (PDP)

One estimation method that allows for process heterogeneity is a latent class model. Every class can be defined to represent a different processing strategy (hence the reference to a probabilistic decision process), and every sampled individual is associated with each class up to a probability. The assigned probability can be specified as a function of other characteristics, such as the socioeconomic characteristics of respondents. However, the modeller implicitly assumes that each sampled individual only uses one decision process strategy. Several choice studies have used this approach to include multiple processing strategies (Swait and Adamowicz, 2001; Hensher and Collins, 2011; Campbell et al., 2012; Hess et al., 2012; Weller et al., 2014). With LPAA and two heuristics included in the current study, the model equations for each class have been prescribed as heuristic-specific utility expressions. Parameter estimates associated with each attribute are then specific to each class; with fixed parameters under process heterogeneity and preference homogeneity. When including behavioural refinements and experience in the PDP model, the risk attitudes and perceptual conditioning are included in the LPAA class only and experience is included in every class.

3.5 Conditioning by Random Process Heterogeneity (CRPH)

Our primary focus is on investigating the possible interaction between process heuristics and the random parameters obtained for an LPAA utility specification in order to see whether this increases our understanding of preference heterogeneity which we refer to as Conditioning by Random Process Heterogeneity (CRPH). The approach recognises that the parameters defined under LPAA may be conditioned by alternative process strategies. It analyses the degrees of potential substitution or complementarity between the non-systematic representation of preference heterogeneity through random parameters and a systematic representation through a conditioning of the heterogeneous preference distribution, where the latter may offer up a behaviourally richer (and statistically improved) explanation of the choice process. Theoretically, the CRPH approach represents a decision made using multiple process strategies simultaneously while allowing for differences between respondents. This proposal is very novel since the PDP approach – which has been used to include multiple process strategies in decision making - considers that each individual uses one process strategy up to a probability. Other approaches, such as the hybrid model (Chorus et al., 2013; Hess et al., 2014) consider that different subsets of attributes are evaluated using a different process strategy. Therefore, the idea that a respondent might use more than one process strategy to evaluate each attribute – with the intention of replicating the complex process that respondents actually use when reaching a decision – has not been undertaken before in discrete choice modelling, to the best of our knowledge.

The CRPH methodology takes into account the possibility that the mean and standard deviation of each attribute under LPAA with behavioural refinements is conditioned by multiple process strategies, thus including process heterogeneity. When there are multiple process strategies influencing the mean estimate, this approach is saying that – on average – respondents do not use a process strategy by itself but rather a combination of different strategies to a certain level. In terms of the standard deviation, the CRPH approach is saying that different respondents might use a different combination of process strategies and this partly explains the random component of the estimates. Therefore, this approach is testing the relationship between process and (LPAA linked) preference heterogeneity. All the models in this paper were estimated using PythonBiogeme (Bierlaire, 2016). For the CRPH models, the parameters were estimated using MML with random parameters under an LPAA framework as starting values.

The mixed logit model that considers random parameters, as was presented previously, decomposes parameter θ_{in} in its mean and standard deviation. To incorporate process heuristics using the CRPH approach, the mean and standard deviation of attribute x_{in} under an LPAA mixed logit model, are written as a function of each of the process heuristics. The utility expression can be written as follows:

$$U_{iqt} = \sum_{n} \left(\left[\begin{array}{l} \theta_{in}^{m} + \lambda_{VL,in}^{m} \cdot VL\left(x_{inqt}\right) + \lambda_{RAM,in}^{m} \cdot RAM\left(x_{inqt}\right) \\ + \left[\sigma_{in} + \lambda_{VL,in}^{s} \cdot VL\left(x_{inqt}\right) + \lambda_{RAM,in}^{s} \cdot RAM\left(x_{inqt}\right) \right] \cdot v \end{array} \right] \cdot x_{inqt} \right) + \varepsilon_{iqt} \quad (20)$$

where $VL\left(x_{inqt}\right)$ represents the transformation of x_{inqt} for the VL heuristic; $RAM\left(x_{inqt}\right)$ for the RAM heuristic; $\lambda_{VL,in}^{m}$ represents the relationship between the mean estimate and VL; $\lambda_{RAM,in}^{m}$ represents the relationship between the mean estimate and RAM; $\lambda_{VL,in}^{s}$ is the relationship between the standard deviation estimate of the random parameter distribution and VL; and $\lambda_{RAM,i}^{s}$ is the relationship between the standard deviation and RAM.

The parameters estimated in RAM are equivalent to the ones in LPAA, since both are related to the direct values of the attribute: x_{inqt}. However, they are not equivalent to the parameters estimated in a VL heuristic since they represent the difference between the attribute level and the reference level: $\left(x_{inqt} - ref_{nqt}\right)^{\varphi}$. Therefore, the parameters θ_{in}^{m} can be considered common between LPAA and RAM, but not for VL, which will have its own parameters θ_{in}^{VL}. The transformations of x_{in} associated with VL and RAM are as follows:

$$VL\left(x_{inqt}\right) = \theta_{in}^{VL} \cdot \left(x_{inqt} - ref_{nqt}\right)^{\varphi} \quad (21)$$

$$RAM\left(x_{inqt}\right) = \theta_{in} \cdot x_{inqt} + \sum_{j \in S} R(i,j) \quad (22)$$

Merging equation (20) with (21) and (22), the expression for CRPH that includes VL and RAM results in the following form:

$$U_{iqt} = \sum_{n} \left(\left[\begin{array}{l} \theta_{in}^{m} + \lambda_{VL,in}^{m} \cdot \left(\theta_{in}^{VL} \cdot \left(x_{inqt} - ref_{nqt}\right)^{\varphi} \right) + \lambda_{RAM,in}^{m} \cdot \left(\theta_{in}^{m} \cdot x_{inqt} + \sum_{j \in S} R(i,j) \right) \\ + \left[\sigma_{in} + \lambda_{VL,in}^{s} \cdot \left(\theta_{in}^{VL} \cdot \left(x_{inqt} - ref_{nqt}\right)^{\varphi} \right) + \lambda_{RAM,in}^{s} \cdot \left(\theta_{in}^{m} \cdot x_{inqt} + \sum_{j \in S} R(i,j) \right) \right] \cdot v \end{array} \right] \cdot x_{inqt} \right) + \varepsilon_{iqt} \quad (23)$$

The part of the equation that considers the VL transformation includes two parameters: $\lambda_{VL,in}$ and θ_{in}^{VL}. As both are alternative and attribute specific, it is not possible to estimate them separately in this form (the issue of identification). Therefore, a parameter representing both of them together will be estimated, equal to:

$$\lambda_{VL,in}^{'} = \lambda_{VL,in} \cdot \theta_{in}^{VL} \quad (24)$$

Therefore, equation (23) can be written as:

$$U_{iqt} = \sum_{n} \left(\begin{bmatrix} \theta_{in}^{m} + \lambda_{VL,in}^{'m} \cdot \left(x_{inqt} - ref_{nqt} \right)^{\varphi} + \lambda_{RAM,in}^{m} \cdot \left(\theta_{in}^{m} \cdot x_{inqt} + \sum_{j \in S} R(i,j) \right) \\ + \left[\sigma_{in} + \lambda_{VL,in}^{'s} \cdot \left(x_{inqt} - ref_{nqt} \right)^{\varphi} + \lambda_{RAM,in}^{s} \cdot \left(\theta_{in}^{m} \cdot x_{inqt} + \sum_{j \in S} R(i,j) \right) \right] \cdot v \end{bmatrix} \cdot x_{inqt} \right) + \varepsilon_{iqt} \quad (25)$$

λ_{in} can be unique or identical amongst the attributes between the process heuristics. If they are considered identical, then the relationship between the alternative process strategies and the mean or standard deviation estimate for a random parameter will be the same for all the attributes. An appeal of the CRPH approach is that the λ_{in} parameters can be attribute-specific (i.e., they depend on *n*) to allow for individuals to use alternative process heuristics for some attributes, but not for all of them. An example of how this could work in decision-making is an individual that compares the choice set travel times with a reference level but not for the other attributes. If this is the situation, the attributes that are not being influenced by a process heuristic would have $\lambda_{in}^{m} = \lambda_{in}^{s} = 0$ (in its mean and standard deviation). This form also allows process strategies to have an influence over the mean but not the standard deviation of an attribute, with $\lambda_{in}^{m} = 0$ and $\lambda_{in}^{s} \neq 0$ or, oppositely, over its standard deviation but not over its mean, with $\lambda_{in}^{m} \neq 0$ and $\lambda_{in}^{s} = 0$.

Another appeal of this approach is that behavioural refinements and experience can be independent of the process strategies. That is, risk attitudes and perceptual conditioning will only influence x_{in} and overt experience will affect the entire utility function.

These models are complex, but one can think of this specification along similar lines of a standard mixed logit model where a specified random parameter is conditioned on sources of systematic variation. These can include socioeconomic characteristics and reinterpretations of the role of specific attributes in choice making. If, for example, a particular attribute is in the LPAA formulation and in the interaction (via a process heuristic) with the LPAA parameter, then we are effectively creating a non-linearity and this gets reflected in the overall form of input into a utility expression as well as in the WTP formula which as we show is relatively complex and highly non-linear.

It is important to note that the CRPH approach estimates more than one parameter related to the mean and more than one related to the standard deviation (assuming there are significant influences of alternative process strategies). Therefore, the parameters estimated are not behaviourally interpretable by themselves but only in terms of their significance level - and should be removed if not significantly different from zero. However, they can be grouped to analyse the overall influence over the mean and over the standard deviation by calculating the marginal (dis)utilities of each attribute, which is the derivate of the utility function relative to a certain attribute x_{inqt}:

$$\frac{\partial U_{iqt}}{\partial x_{inqt}} = \begin{bmatrix} \theta_{in}^{m} + \lambda_{VL,in}^{m} \cdot VL\left(x_{inqt}\right) + \lambda_{RAM,in}^{m} \cdot RAM\left(x_{inqt}\right) \\ + \left[\sigma_{in} + \lambda_{VL,in}^{s} \cdot VL\left(x_{inqt}\right) + \lambda_{RAM,in}^{s} \cdot RAM\left(x_{inqt}\right) \right] \cdot v \end{bmatrix} + \begin{bmatrix} \lambda_{VL,in}^{m} \cdot \frac{\partial VL\left(x_{inqt}\right)}{\partial x_{inqt}} + \lambda_{RAM,in}^{m} \cdot \frac{\partial RAM\left(x_{inqt}\right)}{\partial x_{inqt}} \\ + \left[\lambda_{VL,in}^{s} \cdot \frac{\partial VL\left(x_{inqt}\right)}{\partial x_{inqt}} + \lambda_{RAM,in}^{s} \cdot \frac{\partial RAM\left(x_{inqt}\right)}{\partial x_{inqt}} \right] \cdot v \end{bmatrix} \quad (26)$$

If the purpose of a study is to compare the CRPH attribute estimates with a traditional model, such as an LPAA, then it would have to compare the attribute's marginal (dis)utilities derived from each model, as the CRPH parameters estimates have a different formulation than traditional models. Nevertheless, the willingness to pay estimates are one of the most important outcomes in choice studies so this study will focus on the comparison of them instead of the marginal (dis)utilities.

3.6 Willingness to Pay Estimates

The willingness to pay estimates (WTP) represent how much is a person willing to pay for a one unit increase (decrease) in an attribute x_{inqt} expressed as follows:

$$WTP\left(x_{inqt}\right)=\frac{\partial U_{iqt}/\partial x_{inqt}}{\partial U_{iqt}/\partial x_{\$.iqt}} \tag{27}$$

where $x_{\$,iqt}$ represents the cost attribute; and $\partial U_{iqt}/\partial x_{inqt}$ the marginal (dis)utilities associated with the attribute of interest.

In some cases, as will be seen later, cost is represented by more than one attribute. For example, the cost of using a car be described by fuel and parking costs. To account for this, a weighted average for the marginal (dis)utilities of the cost attributes is used to calculate the denominator of WTP (Hensher et al., 2012; Hensher et al., 2013a; Leong and Hensher, 2015):

$$\frac{\partial U_i}{\partial x_{\$,i}}=\frac{\partial U_i/\partial x_{\$1,i}\cdot x_{\$1,i}+\partial U_i/\partial x_{\$2,i}\cdot x_{\$2,i}}{x_{\$1,i}+x_{\$2,i}} \tag{28}$$

The majority of the models estimated are non-linear and, hence the WTP estimates are subject to the value of the attributes and other relevant parameters. We estimate a WTP for each individual given the attribute levels presented to them and sum up to obtain moments for the sample.

3.7 Willingness to Pay Confidence Intervals

To compare the willingness to pay estimates between models, confidence intervals are required, which are somewhat complex for WTP estimates derived from more than one parameter estimate in a non-linear form.

We use the Delta method to calculate the standard error of the WTP estimates (Oehlert, 1992; Scarpa and Rose, 2008; Bliemer and Rose, 2013; Hensher et al. 2015). Considering that $\hat{\theta}$ represents a vector of (maximum likelihood) estimates for the unknown parameters θ_{in}, the delta method states that if $\hat{\theta}$ is asymptotically distributed, then a function of $f(\hat{\theta})$ is asymptotically normally distributed with a mean of $f(\theta)$ and a variance of:

$$\nabla_{\theta} f(\theta)^{\mathrm{T}}\Omega_{\theta}\nabla_{\theta} f(\theta) \tag{29}$$

where $\nabla_{\theta} f(\theta)$ denotes the Jacobian of $f(\theta)$. Bliemer and Rose (2013) use the Delta method to obtain confidence intervals in linear mixed logit models. To implement this method, we have to re-write the parameter estimates $\hat{\theta}$ equivalent to equation (5) in terms of the distributional parameters $\theta^{m}{}_{in}$ and σ_{in}, which together will be referred to as ϑ, and a parameter-free distribution v, as follows:

$$\theta_{in} = \theta(v_{in} \mid \vartheta_{in}) = \theta^{m}{}_{in} + \sigma_{in} \cdot v_{in} \tag{30}$$

We assume that all random parameters are normally distributed[5]. The Jacobians (first derivatives) for normal distributed parameters are the following:

$$\nabla_{\vartheta}\theta_{in} = \begin{pmatrix} \nabla_{\theta^{m}{}_{in}}\theta_{in} \\ \nabla_{\sigma_{in}}\theta_{in} \end{pmatrix} = \begin{pmatrix} 1 \\ v_{in} \end{pmatrix} \quad \text{and} \quad \nabla_{v}\theta_{in} = \sigma_{in} \tag{31}$$

In a linear mixed logit model (LPAA) where both the cost attribute ($\theta_{i\$}$) and attribute *n* ($\theta_{in}$) are random, the WTP estimate for attribute n is:

$$WTP\left(x_{inqt}\right) = w_{in}(v_{in}, v_{i\$} \mid \vartheta_{in}, \vartheta_{i\$}) = \frac{\theta_{in}(v_{in} \mid \vartheta_{in})}{\theta_{i\$}(v_{i\$} \mid \vartheta_{i\$})} \tag{32}$$

which is asymptotically normally distributed around the true value of the WTP with a mean and a variance:

$$\hat{w}_{in}(v_{in}, v_{i\$}) \sim N\left(w_{in}, \begin{pmatrix} \nabla_{\vartheta_{in}} w_{in} \\ \nabla_{\vartheta_{i\$}} w_{in} \\ \nabla_{v_{in}} w_{in} \\ \nabla_{v_{i\$}} w_{in} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Omega_{\vartheta} & 0 \\ 0 & diag(1,...,1) \end{pmatrix} \begin{pmatrix} \nabla_{\vartheta_{in}} w_{in} \\ \nabla_{\vartheta_{i\$}} w_{in} \\ \nabla_{v_{in}} w_{in} \\ \nabla_{v_{i\$}} w_{in} \end{pmatrix} \right) \tag{33}$$

where $\nabla_{\vartheta_{in}} w_{in}$ and $\nabla_{\vartheta_{i\$}} w_{in}$ represent the first derivatives (Jacobian) of the WTP estimate of attribute n relative to ϑ_{in} and $\vartheta_{i\$}$, respectively; and $\nabla_{v_{in}} w_{in}$ and $\nabla_{v_{i\$}} w_{in}$ are relative to v_{in} and $v_{i\$}$, respectively. Ω_{ϑ} is the submatrix of the variances and covariances of the distributional parameters ϑ_{in} and $\vartheta_{i\$}$; 0 represents a matrix with zeros and magnitude of $\left(\#\vartheta_{in} + \#\vartheta_{i\$}\right) \times \left(\#v_{in} + \#v_{i\$}\right)$, where # represent the number of elements of ϑ_{in}, $\vartheta_{i\$}$, v_{in} and $v_{i\$}$; and diag(1,…,1) is a diagonal matrix with ones and magnitude of $\left(\#v_{in} + \#v_{i\$}\right) \times \left(\#v_{in} + \#v_{i\$}\right)$.

The Jacobians can be calculated as follows:

$$\begin{aligned}
\nabla_{\vartheta_{in}} w_{in} &= \nabla_{\theta_{in}} w_{in} \cdot \nabla_{\vartheta_{in}} \theta_{in} = \frac{1}{\theta_{in}} \cdot \nabla_{\vartheta_{in}} \theta_{in} = \frac{1}{\theta_{in}} \cdot \begin{pmatrix} 1 \\ v_{in} \end{pmatrix} \\
\nabla_{\vartheta_{i\$}} w_{in} &= \nabla_{\theta_{i\$}} w_{in} \cdot \nabla_{\vartheta_{i\$}} \theta_{i\$} = -\frac{\theta_{in}}{\theta_{i\$}{}^{2}} \cdot \nabla_{\vartheta_{i\$}} \theta_{i\$} = -\frac{\theta_{in}}{\theta_{i\$}{}^{2}} \cdot \begin{pmatrix} 1 \\ v_{i\$} \end{pmatrix} \\
\nabla_{v_{in}} w_{in} &= \nabla_{\theta_{in}} w_{in} \cdot \nabla_{v_{in}} \theta_{in} = \frac{1}{\theta_{in}} \cdot \nabla_{v_{in}} \theta_{in} = \frac{1}{\theta_{in}} \cdot \sigma_{in} \\
\nabla_{v_{i\$}} w_{in} &= \nabla_{\theta_{i\$}} w_{in} \cdot \nabla_{v_{i\$}} \theta_{i\$} = -\frac{\theta_{in}}{\theta_{i\$}{}^{2}} \cdot \nabla_{v_{i\$}} \theta_{i\$} = -\frac{\theta_{in}}{\theta_{i\$}{}^{2}} \cdot \sigma_{i\$}
\end{aligned} \tag{34}$$

Given (34), equation (33) can be re-written as:

[5] Different distribution assumptions were tested (e.g., triangular and lognormal), but we found that the normal distribution gave the best fit and behaviourally most appealing results.

$$\hat{w}_{in}(\nu_{in},\nu_{i\$}) \sim N\left(w_{in}, \begin{pmatrix} \frac{1}{\theta_{in}} \\ \frac{1}{\theta_{in}}\cdot\nu_{in} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}}\cdot\nu_{i\$} \\ \frac{1}{\theta_{in}}\cdot\sigma_{in} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}}\cdot\sigma_{i\$} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Omega_{\vartheta} & 0 \\ 0 & diag(1,...,1) \end{pmatrix} \begin{pmatrix} \frac{1}{\theta_{in}} \\ \frac{1}{\theta_{in}}\cdot\nu_{in} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}}\cdot\nu_{i\$} \\ \frac{1}{\theta_{in}}\cdot\sigma_{in} \\ -\frac{\theta_{in}}{\theta_{i\$}^{2}}\cdot\sigma_{i\$} \end{pmatrix} \right) \tag{35}$$

If both parameters are fixed, then $\nu_{in} = \nu_{in} = \sigma_{in} = \sigma_{i\$} = 0$. The expected WTP estimate $\hat{w}_{in}$ is:

$$\hat{w}_{in} = \int_{\nu_{in}} \int_{\nu_{i\$}} \hat{w}_{in}(\nu_{in},\nu_{i\$})\,\mathrm{d}\,F_{in}(\nu_{in})\,\mathrm{d}\,F_{i\$}(\nu_{i\$}) \tag{36}$$

where $F_{in}(\nu_{in})$ and $F_{i\$}(\nu_{i\$})$ are the cumulative distribution functions of the standard distributed ν_{in} and $\nu_{i\$}$, respectively. Since normal distributions are defined on the complete domain of $(-\infty,+\infty)$, the willingness to pay estimate would be undefined when $\theta_{i\$} \approx 0$. Daly et al. (2012) demonstrate that for the moments to be finite, the probability of observing $\theta_{i\$} \approx 0$ should be zero as in, for example, a log-normal distribution.

Bliemer and Rose (2013) suggest the use of the median instead of the mean, as it would represent a more robust estimator across different models, considering that the mean does not exist when the cost attribute is normally distributed. When estimating mixed logit models in this study, all the attributes are considered to be normally distributed; hence the results are presented using the median WTP estimates and median standard errors.

The expected WTP can be approximated by Monte Carlo simulation as:

$$\hat{w}_{in} \approx \frac{1}{R}\sum_{r=1}^{R} \hat{w}_{in}(\nu_{in}{}^{(r)},\nu_{i\$}{}^{(r)}) \tag{37}$$

where $r = 1,\ldots,R$ are pseudo random draws such as Halton sequences, to ensure more uniform coverage over the distribution (Train, 1999). The approximation is expected to be more accurate under a larger number of draws. We use 25,000 pseudo random draws to calculate the median and the variance (standard errors).

In non-linear models, as presented in this study, the function $\hat{w}_{in}(\nu_{in}{}^{(r)},\nu_{i\$}{}^{(r)})$ involves sample data (e.g., attribute levels). They can be evaluated using the mean levels of the data, or they can be averaged over the observations (Hensher et al. 2015). Both methods were used to verify that the results do not change significantly, although the results presented will report findings using the first method. The levels for the expected WTP with a confidence level of α, where $se(\hat{w}_{in})$ is the standard error, are as follows:

$$\left(\hat{w}_{in} - t_{1-\alpha/2} \cdot se(\hat{w}_{in}), \hat{w}_{in} + t_{1-\alpha/2} \cdot se(\hat{w}_{in})\right) \tag{38}$$

In the following two subsections will present the formulae for the confidence intervals when using the PDP and CRPH methods, both of which are highly non-linear.

3.7.1 PDP method

The PDP approach considers a 'latent class' structure, where each class represents a pre-defined heuristic. Applying the Delta method in a similar way to above, we have that:

$$\hat{w}_{in}(v_{in}, v_{i\$}) \sim N\left(w_{in}, \begin{pmatrix} \nabla_{\theta_{in|C1}} w_{in} \\ \nabla_{\theta_{in|C2}} w_{in} \\ \nabla_{\theta_{in|C3}} w_{in} \\ \nabla_{\alpha_n} w_{in} \\ \nabla_{\gamma_n} w_{in} \\ \nabla_{\theta_{i\$|C1}} w_{in} \\ \nabla_{\theta_{i\$|C2}} w_{in} \\ \nabla_{\theta_{i\$|C3}} w_{in} \\ \nabla_{\alpha_\$} w_{in} \\ \nabla_{\varphi} w_{in} \\ \nabla_{\beta_{i,\exp|C1}} w_{in} \\ \nabla_{\beta_{i,\exp|C2}} w_{in} \\ \nabla_{\beta_{i,\exp|C3}} w_{in} \\ \nabla_{Cte_{C1}} w_{in} \\ \nabla_{Cte_{C2}} w_{in} \\ \nabla_{Cte_{C3}} w_{in} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Omega_{\theta,\alpha,\gamma,\varphi,\beta_{i,\exp},Cte_{Class}} & 0 \\ 0 & diag(1,\ldots,1) \end{pmatrix} \begin{pmatrix} \nabla_{\theta_{in|C1}} w_{in} \\ \nabla_{\theta_{in|C2}} w_{in} \\ \nabla_{\theta_{in|C3}} w_{in} \\ \nabla_{\alpha_n} w_{in} \\ \nabla_{\gamma_n} w_{in} \\ \nabla_{\theta_{i\$|C1}} w_{in} \\ \nabla_{\theta_{i\$|C2}} w_{in} \\ \nabla_{\theta_{i\$|C3}} w_{in} \\ \nabla_{\alpha_\$} w_{in} \\ \nabla_{\varphi} w_{in} \\ \nabla_{\beta_{i,\exp|C1}} w_{in} \\ \nabla_{\beta_{i,\exp|C2}} w_{in} \\ \nabla_{\beta_{i,\exp|C3}} w_{in} \\ \nabla_{Cte_{C1}} w_{in} \\ \nabla_{Cte_{C2}} w_{in} \\ \nabla_{Cte_{C3}} w_{in} \end{pmatrix} \right) \tag{39}$$

The classes consider specific and fixed parameters for each attribute, as shown in section 3.4. Since the parameters are fixed, $\theta_{in} = \theta^{m}{}_{in}$. $\theta_{in|C1}$, $\theta_{in|C2}$, and $\theta_{in|C3}$ represent the estimate for attribute *n* alternative *i* for the RAM heuristic (class 1), VL heuristic (class 2) and LPAA (class 3), respectively; α_n and $\alpha_\$$ represent the risk attitudes for attribute *n* and for the cost attribute, respectively, for the LPAA heuristic (class 1); γ_n represents perceptual conditioning for attribute *n* (it is different from zero when it refers to the travel time attribute) for the LPAA heuristic (class 1); φ represents the concavity factor for the VL heuristic (class 2); $\beta_{i,\exp|C1}, \beta_{i,\exp|C2}, \beta_{i,\exp|C3}$ represents the parameter estimated for experience on mode *i* for each class; and $Cte_{C1}, Cte_{C2}, Cte_{C3}$ represent the class specific constants for the class assignment.

3.7.2 CRPH method

The CRPH approach considers interactions between the process heuristics and the mean and/or standard deviation of the estimates associated with the LPAA form with behavioural refinements, as presented in section 3.5. Applying the Delta method while assuming the attribute estimates are all normally distributed, we have:

$$\hat{w}_{in}(v_{in}, v_{i\$}) \sim N\left(w_{in}, \begin{pmatrix} \nabla_{\theta_{in}^m} w_{in} \\ \nabla_{\sigma_{in}} w_{in} \\ \nabla_{\alpha_n} w_{in} \\ \nabla_{\gamma_n} w_{in} \\ \nabla_{\lambda_{VL,in}^m} w_{in} \\ \nabla_{\lambda_{RAM,in}^m} w_{in} \\ \nabla_{\lambda_{VL,in}^s} w_{in} \\ \nabla_{\lambda_{RAM,in}^s} w_{in} \\ \nabla_{\theta_{i\$}^m} w_{in} \\ \nabla_{\sigma_{i\$}} w_{in} \\ \nabla_{\alpha_\$} w_{in} \\ \nabla_{\lambda_{VL,i\$}^m} w_{in} \\ \nabla_{\lambda_{RAM,i\$}^m} w_{in} \\ \nabla_{\lambda_{VL,i\$}^s} w_{in} \\ \nabla_{\lambda_{RAM,i\$}^s} w_{in} \\ \nabla_{\varphi} w_{in} \\ \nabla_{v_n} w_{in} \\ \nabla_{v_\$} w_{in} \\ \nabla_{\beta_{i,\exp}} w_{in} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \Omega_{\theta,\alpha,\gamma,\lambda,\varphi,\beta_{i,\exp}} & 0 \\ 0 & diag(1,...,1) \end{pmatrix} \begin{pmatrix} \nabla_{\theta_{in}^m} w_{in} \\ \nabla_{\sigma_{in}} w_{in} \\ \nabla_{\alpha_n} w_{in} \\ \nabla_{\gamma_n} w_{in} \\ \nabla_{\lambda_{VL,in}^m} w_{in} \\ \nabla_{\lambda_{RAM,in}^m} w_{in} \\ \nabla_{\lambda_{VL,in}^s} w_{in} \\ \nabla_{\lambda_{RAM,in}^s} w_{in} \\ \nabla_{\theta_{i\$}^m} w_{in} \\ \nabla_{\sigma_{i\$}} w_{in} \\ \nabla_{\alpha_\$} w_{in} \\ \nabla_{\varphi} w_{in} \\ \nabla_{\lambda_{VL,i\$}^m} w_{in} \\ \nabla_{\lambda_{RAM,i\$}^m} w_{in} \\ \nabla_{\lambda_{VL,i\$}^s} w_{in} \\ \nabla_{\lambda_{RAM,i\$}^s} w_{in} \\ \nabla_{v_n} w_{in} \\ \nabla_{v_\$} w_{in} \\ \nabla_{\beta_{i,\exp}} w_{in} \end{pmatrix} \right) \quad (40)$$

where α_n and $\alpha_\$$ represent the risk attitudes for attribute *n* and for the cost attribute, respectively; γ_n represents perceptual conditioning for attribute *n* (it is different from zero when it refers to the travel time attribute); $\lambda_{VL,in}^m$ and $\lambda_{VL,i\$}^m$ represents the relationship between the mean estimate and VL for attribute *n* and the cost attribute; $\lambda_{RAM,in}^m$ and $\lambda_{RAM,i\$}^m$ represent the relationship between the mean and RAM for attribute *n* and the cost attribute; $\lambda_{VL,in}^s$ and $\lambda_{VL,in}^s$ defines the relationship between the standard deviation and VL for attribute *n* and the cost attribute; $\lambda_{RAM,in}^s$ and $\lambda_{RAM,i\$}^s$ defines the relationship between the standard deviation and RAM for attribute *n* and the cost attribute; φ represents the concavity factor for the VL heuristic; and $\beta_{i,\exp}$ represents the parameter estimated for experience in using mode *i.*

4 Datasets

The first dataset used in this study, referred to as Metro Rail, was collected to evaluate the New South Wales government proposal to build a new Metro rail system for Sydney (Hensher et al., 2011). The survey included four alternatives: bus, metro, train and car. Each of them was described by access, main mode and egress attributes. The travel times for the car and bus are described using three attributes: slowest trip time, quickest trip time and travel time on average. Figure 2 presents an illustrative choice experiment screen. For more information the reader is referred to Hensher et al. (2011) and Balbontin (2018).

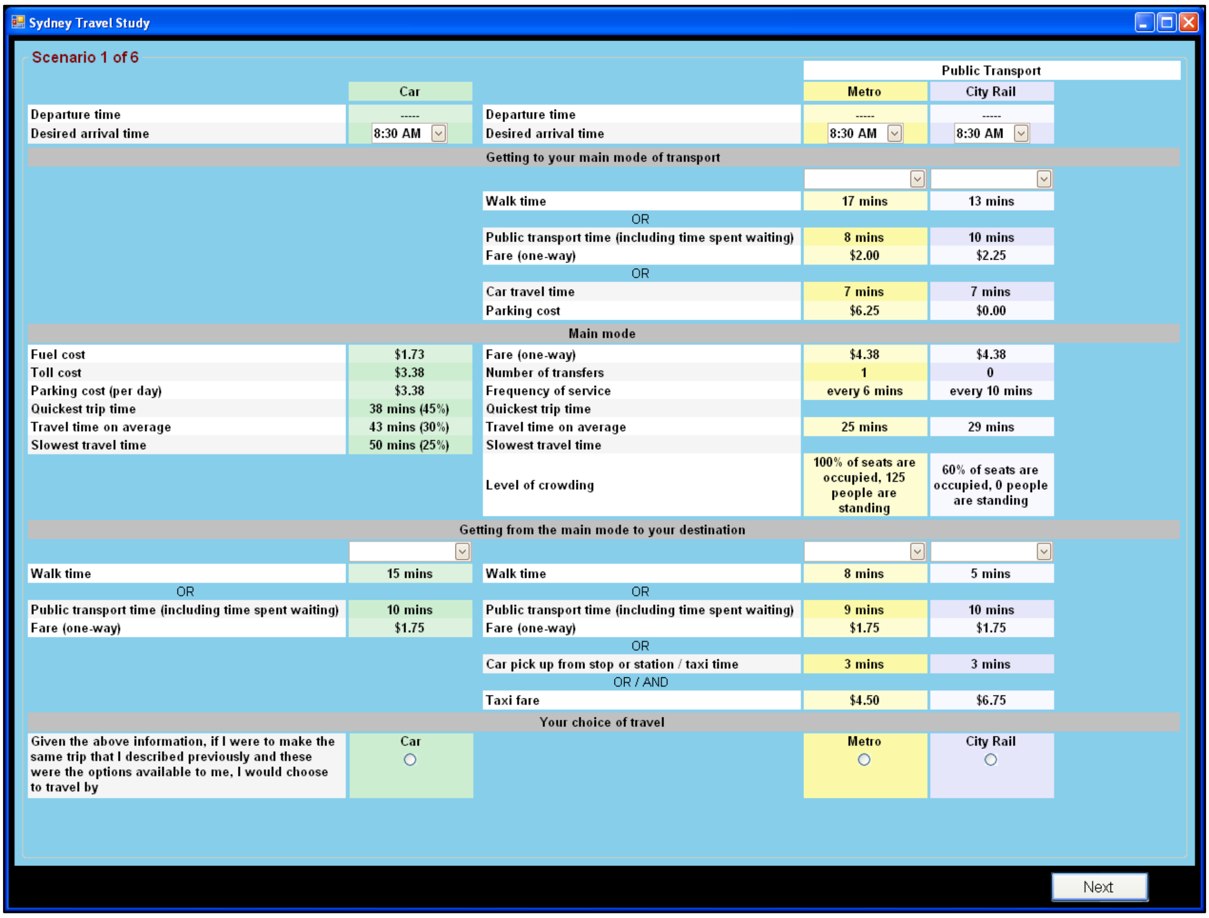
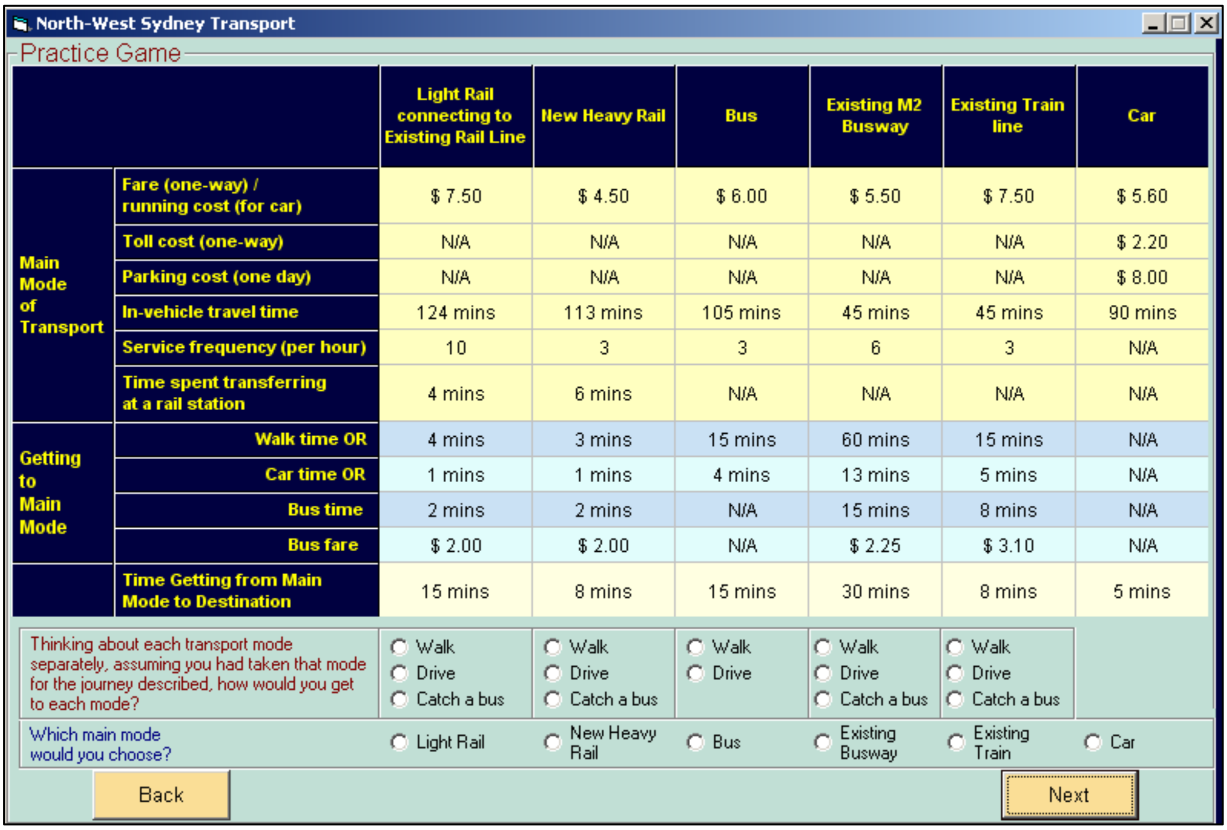


Figure 2: Illustrative screenshot of Metro Rail Sydney choice experiment

The second dataset, referred to as Northwest, was collected as part of a larger study to evaluate public transport investment options (train and bus) in the north west of Sydney, one of the fastest growing areas in Sydney (Hensher and Rose, 2007). The projects under consideration included variations of new heavy rail systems, new light rail and dedicated busway systems along the same corridor. The sample covered residents that made trips within the region (intra-regional) and outside of the region (inter-regional). If an individual made intra-regional trips, the survey presented three public transport modes: new light rail, new heavy rail and bus, plus a car alternative if it was available for him. If an individual made inter-regional trips, the survey included five public transport modes: new light rail, new heavy rail, bus, existing M2 busway and existing train line, plus a car alternative if available. Each alternative was described by access, egress and main mode attributes. Figure 3 presents an illustrative choice screen. For more information the reader is referred to Hensher and Rose (2007) and Balbontin (2018).



		Light Rail connecting to Existing Rail Line	New Heavy Rail	Bus	Existing M2 Busway	Existing Train line	Car
Main Mode of Transport	Fare (one-way) / running cost (for car)	$ 7.50	$ 4.50	$ 6.00	$ 5.50	$ 7.50	$ 5.60
	Toll cost (one-way)	N/A	N/A	N/A	N/A	N/A	$ 2.20
	Parking cost (one day)	N/A	N/A	N/A	N/A	N/A	$ 8.00
	In-vehicle travel time	124 mins	113 mins	105 mins	45 mins	45 mins	90 mins
	Service frequency (per hour)	10	3	3	6	3	N/A
	Time spent transferring at a rail station	4 mins	6 mins	N/A	N/A	N/A	N/A
Getting to Main Mode	Walk time OR	4 mins	3 mins	15 mins	60 mins	15 mins	N/A
	Car time OR	1 mins	1 mins	4 mins	13 mins	5 mins	N/A
	Bus time	2 mins	2 mins	N/A	15 mins	8 mins	N/A
	Bus fare	$ 2.00	$ 2.00	N/A	$ 2.25	$ 3.10	N/A
	Time Getting from Main Mode to Destination	15 mins	8 mins	15 mins	30 mins	8 mins	5 mins
Thinking about each transport mode separately, assuming you had taken that mode for the journey described, how would you get to each mode?		Walk / Drive / Catch a bus	Walk / Drive / Catch a bus	Walk / Drive	Walk / Drive / Catch a bus	Walk / Drive / Catch a bus	
Which main mode would you choose?		Light Rail	New Heavy Rail	Bus	Existing Busway	Existing Train	Car

Figure 3: Illustrative screenshot of North West Sydney choice experiment

The choice models associated with the two surveys used labelled choice data (obtained by a D-optimal design that accounted for random parameters) and included new and existing modal alternatives. One of the main differences between the two designs is that the Metro Rail dataset included reliability (trip time variability) and crowding attributes that were not considered in the Northwest data. Reliability allows for perceptual conditioning, which cannot be tested in Northwest. Moreover, the Northwest data considered seven different alternatives, while the Metro Rail data only four. Even though both datasets were collected in Sydney, they do represent different geographical catchment areas. This will provide an opportunity to study preferences in settings that are sufficiently different as a way to inform the extent to which there are common behavioural traits (i.e. replicability) in travel choice making.

5 Analysis of Results

As explained in the introduction of section 3, different models were estimated: (a) three MNL models: under LPAA, VL and RAM framework; (b) three MML models: under a LPAA, VL and RAM framework; (c) a PDP model; and (d) a CRPH model. The parameter estimates for the final models are presented in Appendix A. The acronyms for the models that will be used in this section are presented in Table 1 and Table 2.

5.1 Comparison of the LPAA or process homogeneity and process heterogeneity models

The first question that arises is whether preferences are better represented (in a statistical sense) when considering multiple decision process strategies. The Akaike's Information Criterion (AIC) was proposed by Akaike (1974) and it can be used to compare the models in regards to their overall fit. It takes into account the log likelihood $l(\theta)$ of a model while penalising the number of parameters estimated, $\#Params$. The indicator also takes into account the number of observations $\#Obs$ and it is calculated as follows:

$$AIC = \frac{-2 \cdot l(\theta) + 2 \cdot \#Params}{\#Obs} \tag{41}$$

Table 3**Error! Reference source not found.** presents the difference between the AIC indicators for the models with LPAA or process homogeneity and process heterogeneity.

Table 3: Comparison of AIC indicators for models with process homogeneity and LPAA or process heterogeneity

	Models		Difference AIC (LPAA or process homogeneity model - Process heterogeneity model)
	LPAA or Process Homogeneity Model	Process Heterogeneity Model	
Metro Rail	LPAA_MNL	PDP	-0.014
	LPAA_MML	CRPH	0.007
	VL_MNL	PDP	0.003
	VL_MML	CRPH	0.024
	RAM_MNL	PDP	-0.014
	RAM_MML	CRPH	0.007
Northwest	LPAA_MNL	PDP	-0.12
	LPAA_MML	CRPH	0.048
	VL_MNL	PDP	-0.08
	VL_MML	CRPH	0.088
	RAM_MNL	PDP	-0.14
	RAM_MML	CRPH	0.028

*Colour scale represents a detriment (red) or improvement (green) in the AIC indicator when considering process heterogeneity relative to process homogeneity. A darker colour tone represents a higher detriment or improvement.

The red shading scale represents a detriment in the AIC indicator when allowing for process heterogeneity relative to LPAA or process homogeneity, and a darker (lighter) red shading represents a larger (lower) detriment. The green colour shading represents an improvement in the AIC indicator, where a larger (lower) improvement is represented by darker (lighter) green. As can be seen, the PDP approach used to integrate multiple process strategies has a worse AIC value than most of the LPAA or process homogeneity models (with the exception of the VL_MML in the Metro Rail dataset). Contrarily, the CRPH models used to account for process heterogeneity have a much improved AIC reinforces a position the importance and impact of random parameters when understanding preferences, and raises the question of what role random parameters play in capturing process heterogeneity when specific processing heuristics are also accommodated.

5.2 Comparison of the experience and behavioural refinement results

It is of interest to establish the differences in these models attributable to experience and behavioural refinements, and see how they interact with LPAA or process homogeneity and heterogeneity, given both are statistically significant in capturing preference variations, Table 4 summarises the parameters that are statistically significant in the estimated models. The last column in the table shows the percentage of the parameters that were statistically significant out of all that were tested. In the LPAA_MNL model, which represents a relatively simple model with fixed parameters and including one – and the most commonly used -

process strategy, 50% of the parameters tested were statistically significant in the Metro Rail dataset, and 82% in the Northwest dataset. The notable difference between these percentages is in perceptual conditioning, which was only tested in the Metro Rail dataset on travel time reliability but excluded for the Northwest data because no attribute was presented with levels of variation within an alternative and choice scenario. Encouragingly, both data sources found the experience parameter for all the modes and more than half of the possible risk attitudes to be statistically significant. When adding random parameters to the LPAA (LPAA_MML) model, we see a decrease in the number of parameters found to be statistically significant: in the Metro Rail dataset from 50% to 43%, and in Northwest data from 82% to 36%. The larger decrease for the Northwest data relative to the Metro Rail is due primarily to the experience parameters where half were not statistically significant, whereas in the Metro Rail dataset all remained statistically significant. The models that considered VL or RAM as the only process strategy, only included experience and not behavioural refinements. In the Metro Rail dataset 2/3 of the experience parameters were statistically significant in the VL models, and all were significant in the RAM models. In the Northwest dataset, all the experience parameters were statistically significant in the VL and RAM models. There was no difference in the number of significant behavioural refinements and/or experience parameters when adding random parameters in the VL or RAM models for both datasets.

The last two rows in each dataset in Table 4 summarise the evidence for the process heterogeneity models. There is a large decrease in the percentage of behavioural refinements and experience parameters found to be statistically significant for these models.

For the PDP model in the Metro Rail dataset, only a 25% of these parameters were statistically significant and 22% in the Northwest dataset. Even though the PDP model tested many more experience parameters (three times more, representing each class), approximately only 1/4 of them were statistically significant. For the CRPH model, these percentages decrease even more: in the Metro Rail dataset, to 21%, and in the Northwest dataset, to 18%.

It is also relevant to interpret the behavioural refinement results. In the Metro Rail dataset, the LPAA_MML evidence suggests risk aversion towards bus cost and risk taking attitude towards parking cost. In the PDP approach, the LPAA class had a significant risk taking attitude towards the parking cost, similarly for the LPAA_MML model. Perceptual conditioning was only statistically significant in the LPAA_MML model for car. In the Northwest dataset, the LPAA_MML model results show a risk aversion towards all public transport fares (for the currently available modes and the new modal investments); and the PDP model, the LPAA class showed a significant risk averse attitude towards the costs of the currently available modes. The other models did not include risk attitudes and/or perceptual conditioning.

What does this mean for model selection? The finding that the process heterogeneity models identified statistically significantly fewer behavioural refinements and experience parameters is an important finding. This suggests that the importance of including additional behavioural components is reduced in the presence of process heterogeneity (i.e., the PDP and CRPH models) relative to LPAA or process homogeneity (i.e., VL or RAM models), and that the risk of confoundment may increase with more complex model forms. This statement suggests that by including process heterogeneity, other behavioural refinements are not as important. It could certainly be because process heterogeneity is to a greater extent interacted with the behavioural refinements, but we cannot be certain of this. We can only argue that less attention is required on other behavioural refinements when including process heterogeneity.

Table 4: Summary of the behavioural refinements and experience significant in the preferred models

Dataset	Preferred Models	Model Characteristics		Behavioural Refinements and Experience Parameters				
		Random Parameters	Multiple decision process strategies	Experience*	Risk attitudes*	Perceptual conditioning*	Number of Parameters	% of Total Possibilities
Metro Rail	LPAA_MNL	No	No	Yes (3/3)	Yes (3/4 travel time + 1/5 cost)	No (0/2)	7	50%
	LPAA_MML	Yes	No	Yes (3/3)	Yes (0/4 travel time + 2/5 cost)	Yes (1/2)	6	43%
	VL_MNL	No	No	Yes (2/3)	-	-	2	67%
	VL_MML	Yes	No	Yes (2/3)	-	-	2	67%
	RAM_MNL	No	No	Yes (3/3)	-	-	3	100%
	RAM_MML	Yes	No	Yes (3/3)	-	-	3	100%
	PDP	No	Yes	Yes (4/9)	Yes (0/4 travel time + 1/5 cost)	No (0/2)	5	25%
	CRPH	Yes	Yes	Yes (3/3)	No (0/4 travel time + 0/5 cost)	No (0/2)	3	21%
North-west	LPAA_MNL	No	No	Yes (4/4)	Yes (2/3 travel time + 3/4 cost)	-	9	82%
	LPAA_MML	Yes	No	Yes (2/4)	Yes (0/3 travel time + 2/4 cost)	-	4	36%
	VL_MNL	No	No	Yes (4/4)	-	-	4	100%
	VL_MML	Yes	No	Yes (4/4)	-	-	4	100%
	RAM_MNL	No	No	Yes (4/4)	-	-	4	100%
	RAM_MML	Yes	No	Yes (4/4)	-	-	4	100%
	PDP	No	Yes	Yes (4/16)	Yes (0/3 travel time + 1/4 cost)	-	5	22%
	CRPH	Yes	Yes	Yes (2/4)	No (0/3 travel time + 0/4 cost)	-	2	18%

*The numbers in parenthesis represent how many parameters are significant out of the total of parameters that were tested.

5.3 Comparison of the process heterogeneity methods

Another objective of this paper is to compare the process heterogeneity methods used. The PDP approach (as a latent class form) suggests, from the perspective of the decision maker, that there is no statistically significant *interaction* between two or three process strategies; rather that an individual respondent will select an independent process strategy, in the presence of other process heuristics, up to a probability. In contrast, the CRPH approach hypothesises that the assessed decision process strategies interact with each other with individuals using more than one rule to evaluate a specific attribute. This interaction between process strategies is included in the mean parameter estimate across the sample and in the standard deviation (capturing preference heterogeneity). The CRPH approach is an innovative approach to test for preference heterogeneity at the attribute level in a setting of process heuristics compared to the PDP approach, the most common approach used in the transport literature (and other literatures) to integrate multiple heuristics in choice making models.

Table 5 summarises the AIC indicators for the preferred PDP and CRPH models for each dataset. There is a significant improvement in the AIC for the CRPH approach: in the Metro Rail dataset this improvement is 0.021 and in the Northwest dataset it is 0.169. These differences are statistically significant and show that the CRPH provides an improved representation of preferences in both datasets. This supports the position that process strategies can and often do interact with each other in preference revelation.

Table 5: AIC indicator of preferred PDP and CRPH models

Dataset	Model	Number of Parameters Estimated	Log Likelihood at convergence	Log likelihood at zero	AIC
Metro Rail	PDP	47	-5,007.53	-13,125.44	1.068
	CRPH	35	-4,922.41		1.047
Northwest	PDP	44	-4,864.47	-7,838.25	2.167
	CRPH	32	-4,494.74		1.999

The interactions that are statistically significant for the preferred CRPH models (CRPH) are shown in Table 6 with different colours for each dataset, where yellow represents statistically significant interactions in the preferred CRPH model for both datasets. Four interactions between the standard deviation and the process strategies were present in both datasets: access time with RAM; fare public transport with both VL and RAM; and travel time public transport with VL. There was only one interaction between the mean and process strategies present in both datasets: headway public transport and VL. The attributes transfer in public transport, % seating probability and density were only available in the Metro Rail dataset.

There are both similarities and differences between the models' interactions, with the evidence suggesting dataset specific effects. In both datasets, several interactions were found significant between the mean and standard deviation estimates – traditionally defined under an LPAA assumption - with both the VL and RAM heuristic. This is not surprising as there is a sense that more complex model forms typically reveal differences in evidence between datasets that is not observed in simpler models because of the absence of additional sources of behavioural variance. What this suggests is that if the behaviourally richer models are an improved explanation of choice making, then it becomes more likely that data must be collected in the setting in which a study is focussed, limiting the ability to make inferential statements about the portability of evidence. This may not be such good news for practitioners who are looking for evidence of WTP estimates that can be taken from one context and used into another context. Intuitively we have uncovered further support for the view that accounting for more sources of variability in preferences of a sample of choice reduces the ability to transfer evidence.

Table 6: Process strategies' interactions in CRPH preferred models for the datasets

	Mean		Standard deviation	
	VL	RAM	VL	RAM
Access Time	Northwest			Both
Fare Public Transport			Both	Both
Fuel + Toll Cost Car			Metro Rail	Northwest
Parking Cost Car	Metro Rail		Metro Rail	
Travel Time Public Transport		Northwest	Both	
Travel Time Car	Northwest		Northwest	
Egress Time				
Transfer Public Transport				
Headway Public Transport		Both		
% Seat Public Transport	Metro Rail			
Density Public Transport				Metro Rail

Legend: Northwest (blue), Metro Rail (green), Both (yellow)

The CRPH method proposes that there is a relationship between the preferences identified through process heterogeneity and those identified through the simpler LPAA representation of taste heterogeneity by including interactions between the parameters' mean and standard deviation as well as different process strategies. The interactions presented in Table 6 support the existence of a relationship between sources of preference heterogeneity identified through the process heuristic treatment and the LPAA treatment, both included in an interaction form. In both datasets there are numerous statistically significant interactions between these two sources of preference heterogeneity. The results show that these relationships are statistically significant and contribute behaviourally appealing evidence in representing choice making in both datasets. Moreover, the findings suggest that the relationship between these two sources of preference heterogeneity is attribute-specific and should not be considered common between all the attributes. This is shown in Table 6 where each attribute is represented by a different combination of interactions.

5.4 Willingness to Pay Estimates and Confidence Intervals

The median WTP is the main focus (in contrast to mean estimates) of the analysis because when estimating all parameters as random, the mean WTP is highly dependent on the draws, and thus is not very stable; in fixed parameters models the mean and median WTP are equivalent as was explained in section 3.7. We compare the main findings for the attributes present in both datasets: travel times, access time, egress times, and headway. Table 7 presents the median WTP estimates for the attributes on each dataset relative to year 2009[6], where the last column represents the percentage difference between the datasets.

There are statistically significant differences between the datasets, especially in the access times, egress times and headway. The egress times were presented very differently in the surveys, where the Metro Rail survey presented more detailed egress time information (i.e., walk time, public transport time, car pick up or taxi time), and the Northwest survey presented it only as the total time (i.e., time getting from main mode to destination). This we believe is contributing to differences in evidence between the datasets. The travel times for bus and car were also presented differently in the Metro Rail survey, with different levels of variation. The

[6] Calculated using the annual inflation rate provided by the Reserve Bank of Australia. Respondents were asked for the interval that best represented their income, so a uniform distribution was assumed within each intervals to include inflation.

remaining attributes were relatively similar in both surveys. The results show that there are statistically significant differences in the median WTP estimates between the datasets.

Table 8**Error! Reference source not found.** presents the percentage change in the median WTP estimates when considering process heterogeneity relative to LPAA or the different process homogeneity models. In the Metro Rail dataset, there is a clear increase in the median WTP when using the PDP (latent class) approach to allow for process heterogeneity versus all the LPAA or process homogeneity models, and a decrease when using the MML CRPH approach (except for the car travel time relative to the VL model and the public transport egress time and headway relative to the LPAA model, where there are minor increases).

In the Northwest dataset, there is an increase in the majority of the median WTP estimates when considering process heterogeneity relative to all the LPAA or process homogeneity models, with a few exceptions. There appears to be no common pattern between the datasets in the median WTP estimates when assessing process heterogeneity versus LPAA or process homogeneity. This suggests, given the proviso of data differences, that the proposed CRPH method will not always lead to higher or lower WTP estimates; however they provide a statistically improved and behaviourally more appealing way to represent preferences that underlie choice making.

Table 7: Median WTP estimates for the different models in each dataset relative to year 2009

Attribute	Model	WTP median estimates		
		Metro Rail	Northwest	% change (Metro Rail - Northwest)
Public Transport Travel Time ($ per person hour)	LPAA_MML	5.52	7.68	-39%
	VL_MML	6.43	6.33	2%
	RAM_MML	6.36	5.09	20%
	PDP	8.41	12.11	-44%
	CRPH	5.20	11.76	-126%
Car Travel Time ($ per person hour)	LPAA_MML	15.26	11.06	28%
	VL_MML	13.94	17.95	-29%
	RAM_MML	8.80	33.87	-285%
	PDP	23.60	31.03	-31%
	CRPH	13.74	22.94	-67%
Public Transport Access Time ($ per person hour)	LPAA_MML	5.98	7.61	-27%
	VL_MML	6.46	5.18	20%
	RAM_MML	7.47	4.07	46%
	PDP	11.31	9.64	15%
	CRPH	3.98	7.13	-79%
Public Transport Egress Time ($ per person hour)	LPAA_MML	7.28	6.09	16%
	VL_MML	9.89	3.64	63%
	RAM_MML	9.03	3.18	65%
	PDP	14.37	7.43	48%
	CRPH	7.70	8.79	-14%
Car Egress Time ($ per person hour)	LPAA_MML	18.08	7.81	57%
	VL_MML	26.96	10.41	61%
	RAM_MML	14.74	15.41	-5%
	PDP	32.89	15.65	52%
	CRPH	14.44	13.86	4%
Headway ($ per person minute)	LPAA_MML	0.03	0.06	-61%
	VL_MML	0.04	0.11	-178%
	RAM_MML	0.04	0.08	-74%
	PDP	0.06	0.06	-6%
	CRPH	0.04	0.10	-181%

*Colour scale represents a decrease (red) or increase (green) in the WTP estimates in the Metro Rail data models relative to the Northwest data models. A darker colour tone represents a larger increase or decrease.

Table 8: Influence on the median WTP estimates when considering process heterogeneity versus LPAA or process homogeneity

Attribute	LPAA or Process Homogeneity Model	% change in median WTP estimates when considering process heterogeneity			
		Metro Rail		Northwest	
		PDP	CRPH	PDP	CRPH
Public Transport Travel Time ($ per person hour)	LPAA_MML	52%	-6%	58%	53%
	VL_MML	31%	-19%	91%	86%
	RAM_MML	32%	-18%	138%	131%
Car Travel Time ($ per person hour)	LPAA_MML	55%	-10%	181%	107%
	VL_MML	69%	-1%	73%	28%
	RAM_MML	168%	56%	-8%	-32%
Public Transport Access Time ($ per person hour)	LPAA_MML	89%	-33%	27%	-6%
	VL_MML	75%	-38%	86%	38%
	RAM_MML	51%	-47%	137%	75%
Public Transport Egress Time ($ per person hour)	LPAA_MML	97%	6%	22%	44%
	VL_MML	45%	-22%	104%	141%
	RAM_MML	59%	-15%	134%	177%
Car Egress Time ($ per person hour)	LPAA_MML	82%	-20%	101%	78%
	VL_MML	22%	-46%	50%	33%
	RAM_MML	123%	-2%	2%	-10%
Headway ($ per person minute)	LPAA_MML	63%	6%	8%	85%
	VL_MML	46%	-5%	-44%	-4%
	RAM_MML	26%	-18%	-23%	32%

*Colour scale represents a decrease (red) or increase (green) in the median WTP when considering process heterogeneity relative to LPAA or process homogeneity. A darker colour tone represents a larger increase or decrease.

Given that the value of travel time savings (VTTS) is the most researched and important user benefit in transport appraisal, we take a closer look at the WTP evidence across the models by presenting and discussing the confidence levels around the median estimates. Confidence intervals obtained when considering preferences under LPAA or process homogeneity, LPAA heterogeneity and process heterogeneity for VTTS are given in Figure 4. We graph the median VTTS and 95% confidence intervals for the three preferred models with LPAA or process homogeneity (LPAA_MML, VL_MML, and RAM_MML) and with process heterogeneity (CRPH; and PDP) for both datasets. As indicated above, the standard errors for the PDP models are lower because it does not estimate any parameter as random, contrary to all the other models.

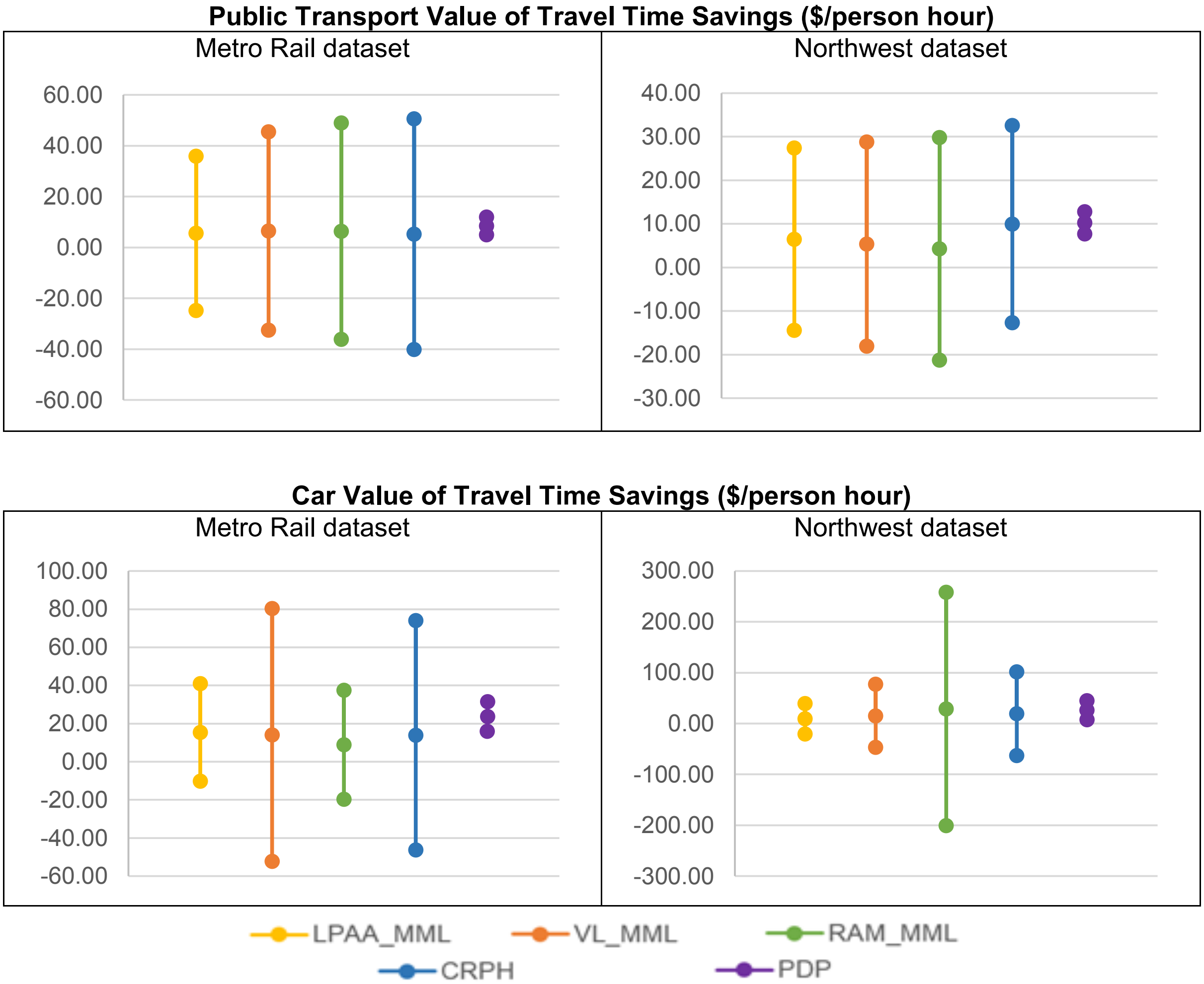


Figure 4: VTTS median and 95% confidence intervals for the models with LPAA or process homogeneity versus process heterogeneity

In the Metro Rail dataset, the lowest median VTTS for public transport is estimated using the CRPH model ($5.20 per person hour), followed by the LPAA_MML model ($5.52 per person hour). The highest median VTTS for public transport is estimated using the PDP model ($8.41 per person hour). The standard error is lower in the PDP model, followed by the LPAA_MML model ($15.46) and by the VL_MML model ($17.79). The largest standard error is found in the CRPH model ($23.14). In the Northwest dataset, the LPAA or process homogeneity models have a lower VTTS median for public transport relative to the process heterogeneity models. The lowest is for the RAM_MML model ($4.30 per person hour) and the highest for the PDP model ($10.24). The standard error for the VTTS under the PDP approach is very low as it estimates fixed parameters. For the other models, the standard error for the public transport

VTTS is lower in the LPAA model ($10.65) followed by the CRPH model ($11.54), and larger in the RAM model ($13.02).

For the car median VTTS, the differences are larger than for public transport, as can be seen in Figure 4. In the Metro Rail dataset, the lowest car median VTTS is estimated using the RAM_MML model ($8.80 per person hour). The highest car median VTTS is estimated using the PDP model ($23.60 per person hour). The largest standard errors are found in the VL_MML and the CRPH models. In the Northwest dataset, the median for the car VTTS is lower in the LPAA preferred model ($9.35 per person hour), followed by the VL ($15.17) and then followed by the CRPH model ($19.39). The car VTTS median is larger under the PDP approach. The standard error is lower in the PDP models, followed by the LPAA model ($15.13 per person hour), VL model ($31.56), the CRPH ($41.88), and larger in the RAM model ($116.99).

Table 9 presents a comparison of the median VTTS using t-test of differences (including the mean and standard error estimates) for the preferred models with process heterogeneity and homogeneity, indicating which estimates are statistically significant from each other. The first three columns compare the CRPH model (which considers preference heterogeneity associated with the interaction between the LPAA form and process heuristics) with the three preferred models with LPAA or process homogeneity (VL and RAM), and the last two columns compare the PDP model (which considers preference heterogeneity through the process rule and preference homogeneity through the LPAA form) with the three preferred models. The results in the Metro Rail dataset show that the majority of the median VTTS estimates using the CRPH final model are statistically significantly different (absolute value larger than 1.96 with a 95% confidence level) from the ones estimated using the LPAA or process homogeneity models. The exceptions are for the bus median VTTS for the stand alone LPAA model, metro median VTTS for all the LPAA or process homogeneity models, and car median VTTS for the VL model, where there is not enough evidence to suggest they are statistically different to the CRPH preferred model. When comparing the PDP model with the LPAA or process homogeneity models, the median VTTS estimates are always larger for the PDP model. The results also show that the PDP median VTTS estimates are all statistically different from the LPAA or process homogeneity median VTTS estimates.

Table 9: Comparison of attributes' median VTTS for models with process homogeneity using the t-test of differences

Travel Time		CRPH vs.			PDP vs.		
		LPAA_MML	VL_MML	RAM_MML	LPAA_MML	VL_MML	RAM_MML
Metro Rail	Bus	↑ 0.07	↓ ***-3.27***	↓ ***-3.06***	↑ ***17.19***	↑ ***5.68***	↑ ***5.36***
	Train	↓ ***-2.24***	↓ ***-3.12***	↓ ***-2.89***	↑ ***10.65***	↑ ***7.49***	↑ ***7.04***
	Metro	↓ -0.56	↓ -1.85	↓ -1.47	↑ ***13.37***	↑ ***9.39***	↑ ***9.03***
	Car	↓ ***-2.73***	↓ -0.27	↑ ***8.70***	↑ ***36.72***	↑ ***16.99***	↑ ***58.66***
Northwest	New Light Rail	↑ ***8.25***	↑ ***9.67***	↑ ***12.97***	↑ ***22.28***	↑ ***21.18***	↑ ***24.18***
	New Heavy Rail	↑ ***6.01***	↑ ***7.61***	↑ ***10.81***	↑ ***19.77***	↑ ***19.10***	↑ ***22.06***
	New Busway	↑ ***17.26***	↑ ***21.80***	↑ ***23.59***	↑ ***9.19***	↑ ***15.92***	↑ ***18.40***
	Bus	↑ ***23.08***	↑ ***29.60***	↑ ***32.31***	↑ ***24.70***	↑ ***33.90***	↑ ***36.63***
	Busway	↑ ***15.35***	↑ ***20.98***	↑ ***23.78***	↑ ***21.88***	↑ ***29.56***	↑ ***32.22***
	Train	↑ ***3.66***	↑ ***5.35***	↑ ***9.22***	↑ ***32.51***	↑ ***28.98***	↑ ***31.87***
	Car	↑ ***14.08***	↑ ***5.02***	↓ ***-4.64***	↑ ***58.88***	↑ ***20.94***	↓ -1.28

Note: Bold and italic estimates are the ones significant at a 95% confidence level

In the Northwest dataset, the results show that all the median VTTS estimates are significantly different and higher in the CRPH and PDP model than in the LPAA or process homogeneity models, except for the car VTTS that is lower in the CRPH and PDP models than in the RAM model. However, the car VTTS difference between the CRPH and RAM model is significant, while it is not significant between the PDP and RAM model (with 95% confidence level).

The results show that there are many significant differences in the VTTS estimates under LPAA or process homogeneity in contrast to models that allow for process heterogeneity. This is an important finding, suggesting significant differences under alternative behavioural assumptions. Although there is no common pattern in the relative estimates of the median WTP estimates when assessing process heterogeneity and LPAA or process homogeneity between the datasets, there is a significant improvement in the statistical performance of the CRPH models. This suggests that in ongoing research, with a focus on identifying behaviourally relevant WTP estimates to be used in real applications, WTP estimates should be obtained from both to ensure that behavioural simplification is not at the expense of behavioural relevance.

6 Conclusions

Given the evidence, we are able to propose a revision of the preference process and choosing behavioural paradigm. This is summarised in Figure 5 as a conceptual framework, with our evidence supporting behaviourally relevant roles for multiple process strategies, behavioural refinements and experience, each in turn offering additional bases of understanding preferences.

The results using the Northwest dataset support a position where risk attitudes are present even where there is limited or no variability in the levels of the attributes, encouraging the assessment of behavioural refinements despite the characteristics of the design of the available choice experiments. Overt experience also has an important role to play in decision-making. The empirical evidence supports McFadden's (2001) call for more effort in building in process rules and experience in choice modelling. Specifically we offer new evidence on how experience, multiple decision process strategies and behavioural refinements all interact. When adding more behavioural relevance (often seen as modelling complexity) into discrete choice models through random parameters and process heterogeneity, the inclusion of additional behavioural refinements may not be necessary.

The preferred preference revelation model form, *conditioning by random process heterogeneity* (CRPH), supports a behavioural paradigm in which individuals use more than one process heuristic in decision making, supporting heterogeneity in processing information related to alternatives on offer. The impact on important behavioural outputs such as willingness to pay is profound, and has important policy relevance in project appraisal.

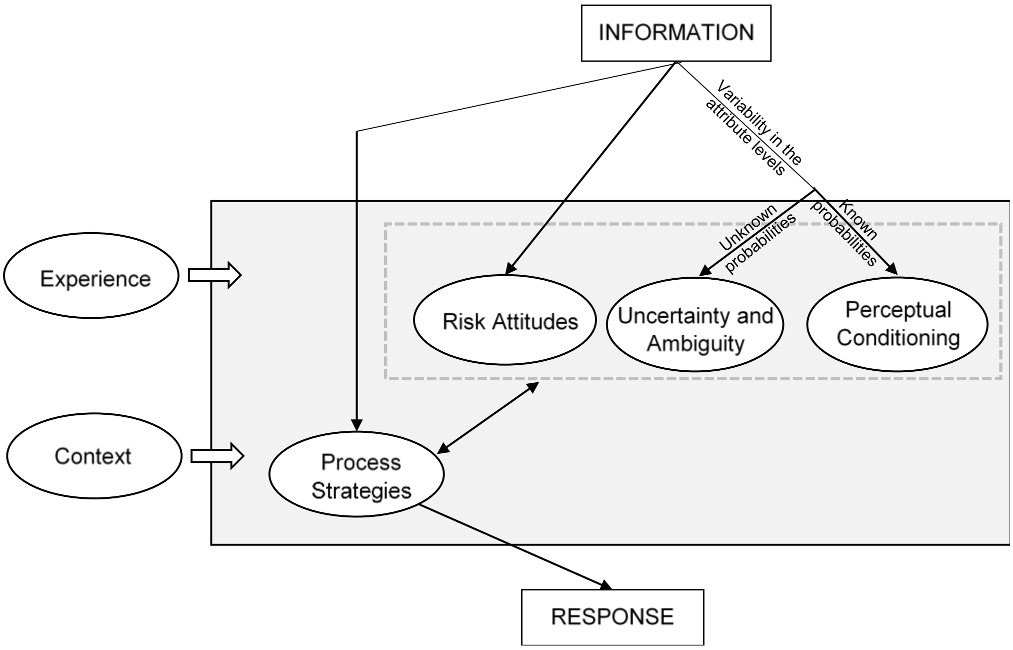


Figure 5: Proposed conceptual framework for decision making

In summary, we recommend that a choice study should allow for taste heterogeneity through the standard random parameter specification (under LPAA) *as well as* an overlay (or interaction) with one or more process heuristics, overt experience and behavioural refinements, especially with risk attitudes (we suggest to test for the inclusion of perceptual conditioning). What this tends to do is to provide a better representation of preferences which is translated into significantly different (higher or lower) median WTP estimates of key attributes such as the value of travel time savings. The dataset requirements to be able to include process heterogeneity are more demanding since greater variance in attribute levels is required to identify the various contributions.

The two datasets used in this paper were constructed through a Bayesian-efficient design using D-error as the optimality criterion. Other datasets required for further inquiry and practical application only need to have enough variability in the attribute levels. If any of the attributes is presented with more than one level of occurrence, then perceptual conditioning can and should be tested for those attributes. Otherwise, only risk attitudes can and should be tested. In terms of data collection, it is strongly advised to include questions regarding individual experience in the alternatives presented, since this is a way that experience can be included in the modelling. No questions regarding the process strategies used by respondents are necessary to estimate this type of models, although they could be asked to help guide the selection of process strategies, but this was not studied in this paper.

This research has proposed and tested the role that different choice model forms (from relatively simple to complex forms) might play in improving our understanding of sources of influence on preference revelation in choice making. For the first time, we have simultaneously integrated into a discrete choice model multiple decision process strategies, risk attitudes, perceptual conditioning and overt experience, and investigated the relationship between the richer behavioural paradigms design to reveal preferences embedded in process heterogeneity and LPAA specified heterogeneity (commonly defined through random parameters). Two datasets were used to provide a more generalised overview of the outcomes under these different behavioural assumptions. The datasets used do have a degree of commonality in terms of design and presentation which is a limitation when generalisation testing is of interest. However, they do have some differences in the design, such as the presentation of travel times and number of alternatives, and in terms of respondents' profile.

One of the most important findings is that when process heterogeneity is accounted for through specific heuristics such as value learning, behavioural refinements and overt experience may not be needed to be incorporated as explicitly influencing sources. This helps in identifying appealing parsimonious preference expressions in choice models. When preference heterogeneity captured through process rules is overlayed in more parsimonious models through random parameters, we find that the interaction between LPAA specified random parameters and processing heuristics adds new insights into the relationship between an increasing number of sources of preference heterogeneity. These phenomena are correlated, and hence behaviourally condition each other in important ways, supporting the use of the interaction mechanism. The evidence is strong enough to suggest empirically that there exists (in two datasets at least) a significant attribute-specific relationship between process strategies and random parameters associated with the LPAA model form. In future research it would be interesting to test how these models behave in a simulated dataset considering process and LPAA specified preference heterogeneity together with behavioural refinements, and to what extent the components of the CRPH functional form complement each other.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alberini, A., Kanninen, B., Richard, T., 2017. Modeling Response Incentive Effects in Dichotomous Choice Contingent Valuation Data Author (s): Anna Alberini , Barbara Kanninen and Richard T . Carson Published by : University of Wisconsin Press Stable URL : http://www.jstor.org/stable/3147170 Modelin 73, 309–324.

Balbontin, C., 2018. Integrating Decision Heuristics And Behavioural Refinements Into Travel Choice Models. PhD Thesis, ITLS, University of Sydney.

Balbontin, C., Hensher, D. a., Collins, A.T., 2017a. Integrating attribute non-attendance and value learning with risk attitudes and perceptual conditioning. Transp. Res. Part E Logist. Transp. Rev. 97, 172–191. https://doi.org/10.1016/j.tre.2016.11.002

Balbontin, C., Hensher, D.A., Collins, A.T., 2017b. Is there a systematic relationship between random parameters and process heuristics? Transp. Res. Part E Logist. Transp. Rev. 106, 160–177. https://doi.org/10.1016/j.tre.2017.07.013

Bateman, I.J., Burgess, D., Hutchinson, W.G., Matthews, D.I., 2006. Preference learning versus coherent arbitrariness: NOAA guidelines or a Learning Design Contingent Valuation (LDCV). Work. Pap. - Cent. Soc. Econ. Res. Glob. Environ. 55, 1–25. https://doi.org/10.1016/j.jeem.2007.08.003

Bateman, I.J., Carson, R.T., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P., 2008. Choice set awareness and ordering effects in discrete choice experiments. CSERGE Work. Pap. EDM 08-01.

Ben-Akiva, M., Morikawa, T., 1990. Estimation of switching models from revealed preferences and stated intentions. Transp. Res. Part A Gen. 24, 485–495. https://doi.org/10.1016/0191-2607(90)90037-7

Bierlaire, M., 2016. PythonBiogeme : a short introduction. Rep. TRANSP-OR 160706, Ser. Biogeme. Transp. Mobil. Lab. Sch. Archit. Civ. Environ. Eng. Ec. Polytech. Fédérale Lausanne, Switzerland.

Bliemer, M.C.J., Rose, J.M., 2013. Confidence intervals of willingness-to-pay for random coefficient logit models. Transp. Res. Part B Methodol. 58, 199–214. https://doi.org/10.1016/j.trb.2013.09.010

Camerer, C.F., Ho, T.H., 1994. Violations of the betweenness axiom and nonlinearity in probability. J. Risk Uncertain. 8, 167–196. https://doi.org/10.1007/BF01065371

Campbell, D., Hensher, D.A., Scarpa, R., 2014. Bounding WTP distributions to reflect the "actual" consideration set. J. Choice Model. 11, 4–15. https://doi.org/10.1016/j.jocm.2014.02.004

Campbell, D., Hensher, D.A., Scarpa, R., 2012. Cost thresholds, cut-offs and sensitivities in stated choice analysis: Identification and implications. Resour. Energy Econ. 34, 396–411. https://doi.org/10.1016/j.reseneeco.2012.04.001

Cantillo, V., Ortuzar, J.D., Williams, H., 2007. Modeling discrete choices in the presence of inertia and serial correlation. Transp. Sci. 41, 195–205. https://doi.org/10.1287/trsc.1060.0178

Carson, R.T., Wilks, L., Imber, D., 1994. Valuing the Preservation of Australia's Kakadu Conservation Zone. Oxf. Econ. Pap. 46, 727–749.

Chernev, A., 2004. Extremeness Aversion and Attribute-Balance Effects in Choice. J. Consum. Res. 31, 249–263. https://doi.org/10.1086/422105

Chorus, C., 2012. Random Regret Minimization: An Overview of Model Properties and Empirical Evidence. Transp. Rev. 32, 75–92. https://doi.org/10.1080/01441647.2011.609947

Chorus, C.G., 2010. A new model of random regret minimization. Eur. J. Transp. Infrastruct. Res. 10, 181–196.

Chorus, C.G., Arentze, T. a., Timmermans, H.J.P., 2008. A Random Regret-Minimization model of travel choice. Transp. Res. Part B Methodol. 42, 1–18. https://doi.org/10.1016/j.trb.2007.05.004

Chorus, C.G., Rose, J.M., Hensher, D. a., 2013. Regret minimization or utility maximization: It depends on the attribute. Environ. Plan. B Plan. Des. 40, 154–169. https://doi.org/10.1068/b38092

Collins, A.T., 2012. Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity. PhD Thesis, ITLS, Univ. Sydney.

Collins, A.T., Rose, J.M., Hensher, D. a., 2013. Specification issues in a generalised random parameters attribute nonattendance model. Transp. Res. Part B Methodol. 56, 234–253. https://doi.org/10.1016/j.trb.2013.08.001

Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation (Amst). 39, 19–31. https://doi.org/10.1007/s11116-011-9331-3

Day, B., Pinto, J.L., 2010. Ordering anomalies in choice experiments. J. Environ. Econ. Manage. 59, 271–285. https://doi.org/10.1016/j.jeem.2010.03.001

De Borger, B., Fosgerau, M., 2008. The trade-off between money and travel time: A test of the theory of reference-dependent preferences. J. Urban Econ. 64, 101–115. https://doi.org/10.1016/j.jue.2007.09.001

DeShazo, J.R., 2002. Designing Transactions without Framing Effects in Iterative Question Formats. J. Environ. Econ. Manage. 43, 360–385. https://doi.org/10.1006/jeem.2000.1185

Eeckhoudt, L., Gollier, C., Schlesinger, H., 2005. Economic and financial decisions under risk, in: Economic and Financial Decisions under Risk. Princeton University Press, pp. 3–25.

Gilboa, I., Pazgal, A., 2001. Cumulative Discrete Choice. Mark. Lett. 12, 119–130. https://doi.org/10.1023/A:1011134718403

Gilboa, I., Schmeidler, D., 1995. Case-Based Decision Theory. Q. J. Econ. 110, 605–639.

Goodwin, P.B., 1977. Habit and Hysteresis in Mode Choice. Urban Stud. 14, 95–98. https://doi.org/10.1080/00420987720080101

Hensher, D.A., 1975. Perception and Commuter Modal Choice - An Hypothesis. Urban Stud. 101–104. https://doi.org/10.1080/00420987520080091

Hensher, D.A., Balbontin, C., Collins, A.T., 2018. Heterogeneity in decision processes: embedding extremeness aversion, risk attitude and perceptual conditioning in multiple process rules choice making. Transp. Res. Part A 111, 316–325. https://doi.org/https://doi.org/10.1016/j.tra.2018.03.026

Hensher, D.A., Collins, A.T., 2011. Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? J. Choice Model. 4, 62–89. https://doi.org/10.1016/S1755-5345(13)70019-8
Hensher, D.A., Collins, A.T., Greene, W.H., 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. Transportation (Amst). 40, 1003–1020. https://doi.org/10.1007/s11116-012-9447-0
Hensher, D.A., Greene, W.H., Li, Z., 2011a. Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings. Transp. Res. Part B Methodol. 45, 954–972. https://doi.org/10.1016/j.trb.2011.05.023
Hensher, D.A., Ho, C.Q., 2016. Experience conditioning in commuter modal choice modelling – Does it make a difference? Transp. Res. Part E Logist. Transp. Rev. 95, 164–176. https://doi.org/10.1016/j.tre.2016.09.010
Hensher, D.A., Rose, J.M., 2007. Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. Transp. Res. Part A Policy Pract. 41, 428–443. https://doi.org/10.1016/j.tra.2006.09.006
Hensher, D.A., Rose, J.M., Collins, A.T., 2011b. Identifying commuter preferences for existing modes and a proposed Metro in Sydney, Australia with special reference to crowding. Public Transp. 3, 109–147. https://doi.org/10.1007/s12469-010-0035-4
Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis - Second Edition. Cambridge University Press. https://doi.org/10.1017/CBO9780511610356
Hensher, D.A., Rose, J.M., Greene, W.H., 2012. Inferring attribute non-attendance from stated choice data: Implications for willingness to pay estimates and a warning for stated choice experiment design. Transportation (Amst). 39, 235–245. https://doi.org/10.1007/s11116-011-9347-8
Hess, S., Beck, M.J., Chorus, C.G., 2014. Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives. Transp. Res. Part A Policy Pract. 66, 1–12. https://doi.org/10.1016/j.tra.2014.04.004
Hess, S., Daly, A., Batley, R., 2018. Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory Decis. 84, 181–204. https://doi.org/10.1007/s11238-017-9651-7
Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: An approach and four case studies. Transportation (Amst). 39, 565–591. https://doi.org/10.1007/s11116-011-9365-6
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–92. https://doi.org/10.2307/1914185
Kivetz, R., Netzer, O., Srinivasan, V., 2004. Alternative Models for Capturing the Compromise Effect. J. Mark. Res. 41, 237–257. https://doi.org/10.1509/jmkr.41.3.237.35990
Leong, W., Hensher, D. a., 2014. Relative advantage maximisation as a model of context dependence for binary choice data. J. Choice Model. 11, 30–42. https://doi.org/10.1016/j.jocm.2014.05.002
Leong, W., Hensher, D.A., 2015. Contrasts of Relative Advantage Maximisation with Random Utility Maximisation and Regret Minimisation. J. Transp. Econ. Policy 49, 167–186(20).
Leong, W., Hensher, D.A., 2012a. Embedding Decision Heuristics in Discrete Choice Models: A Review. Transp. Rev. 32, 313–331. https://doi.org/10.1080/01441647.2012.671195
Leong, W., Hensher, D.A., 2012b. Embedding multiple heuristics into choice models: An exploratory analysis. J. Choice Model. 5, 131–144. https://doi.org/10.1016/j.jocm.2013.03.001
Luce, R.D., 1959. Individual choice behaviour. John Wiley & Sons, Inc, New York.
McFadden, D., 2001. Economic Choices. Am. Econ. Assoc. 91, 351–378.
McFadden, D., 1974. Conditional Logit Analysis of Qualititative Choice Behaviour, P. Zarembk. ed. Frontiers in Econometrics. Academic Press, New York.
McNair, B.J., Bennett, J., Hensher, D.A., 2011. A comparison of responses to single and

repeated discrete choice questions. Resour. Energy Econ. 33, 554–571. https://doi.org/10.1016/j.reseneeco.2010.12.003

McNair, B.J., Hensher, D.A., Bennett, J., 2012. Modelling Heterogeneity in Response Behaviour Towards a Sequence of Discrete Choice Questions: A Probabilistic Decision Process Model. Environ. Resour. Econ. 51, 599–616. https://doi.org/10.1007/s10640-011-9514-6

Oehlert, G.W., 1992. A Note on the Delta Method. Am. Stat. 46, 27. https://doi.org/10.2307/2684406

Ortúzar, J.D.D., Willumsen, L., 2011. Modelling Transport - Fourth Edition. John Wiley and Sons, Chichester.

Plott, C.R., 1996. Rational individual behavior in markets and social choice processes: the discovered preference hypothesis, in: Arrow, K.J., Colombatto, E., Perlman, M., Schmidt, C. (Eds.), Rational Foundations of Economic Behavior. Palgrave Macmillan. https://doi.org/10.1057/9780230389724

Scarpa, R., Rose, J.M., 2008. Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why. Aust. J. Agric. Resour. Econ. 52, 253–282. https://doi.org/10.1111/j.1467-8489.2007.00436.x

Sharpe, K.M., Staelin, R., Huber, J., 2008. Using Extremeness Aversion to Fight Obesity: Policy Implications of Context Dependent Demand. J. Consum. Res. 35, 406–422. https://doi.org/10.1086/587631

Simonson, I., Tversky, A., 1992. Choice in Context : Tradeoff Contrast and Extremeness Aversion. J. Mark. XXIX, 281–296.

Swait, J., Adamowicz, W., 2001. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. J. Consum. Res. 28, 135–148. https://doi.org/10.1086/321952

Train, K., 1999. Halton Sequences for Mixed Logit. Work. Pap. 1–18.

Tversky, A., Kahneman, D., 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. Q. J. Sci. 106, 1039–1061.

Tversky, A., Simonson, I., 1993. Context- dependent Preferences. Manage. Sci. 39, 1179–1189.

van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transp. Res. Part A Policy Pract. 74, 91–109. https://doi.org/10.1016/j.tra.2015.01.008

Weller, P., Oehlmann, M., Mariel, P., Meyerhoff, J., 2014. Stated and inferred attribute non-attendance in a design of designs approach. J. Choice Model. 11, 43–56. https://doi.org/10.1016/j.jocm.2014.04.002

Appendix

Table 10: Metro Rail dataset process homogeneity models' estimates (t-values in parenthesis)

			LPAA_MNL_BRExp	VL_MNL_Exp	RAM_MNL_Exp	LPAA_MML_BRExp		VL_MML_Exp		RAM_MML_Exp	
Number of Parameters Estimated			**24**	**17**	**17**	**31**		**27**		**28**	
Log Likelihood at convergence			**-6,133.04**	**-5,944.52**	**-6,140.35**	**-4,958.17**		**-5,043.60**		**-4,963.32**	
Log likelihood at zero										**-13,125.44**	
AIC			**1.301**	**1.259**	**1.301**	**1.054**		**1.071**		**1.054**	
Parameters	**Acronym**	**Alternatives**	**Mean**	**Mean**	**Mean**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Alternative Specific Constant Bus	ASCBUS	Bus	1.46 (2.78)	0.39 (3.16)	0.43 (2.04)	1.47 (2.59)	-	1.19 (4.68)	-	1.90 (3.28)	-
Alternative Specific Constant Train	ASCTRAIN	Train	1.76 (3.75)	0.46 (4.37)	0.79 (4.68)	2.45 (5.10)	-	1.57 (6.91)	-	2.95 (6.07)	-
Alternative Specific Constant Metro	ASCMETRO	Metro	2.01 (4.12)	0.74 (7.39)	1.20 (8.57)	2.63 (5.78)	-	1.47 (6.98)	-	3.07 (6.25)	-
Alternative Specific Constant Car	ASCCAR	Car	-	-	-	-	-	-	-	-	-
Access Time	ACTIMEPT	Bus, Train, Metro	-0.05 (18.63)	-0.04 (9.53)	-0.05 (19.00)	-0.10 (10.83)	0.14 (12.95)	-0.09 (9.83)	0.15 (8.76)	-0.09 (10.22)	0.17 (14.12)
Fare Public Transport	COSTPT	Bus, Train, Metro	-0.23 (12.41)	-0.24 (13.65)	-0.21 (12.51)	-0.64 (13.88)	0.56 (11.59)	-0.56 (13.12)	0.51 (7.60)	-0.56 (13.48)	0.44 (9.11)
Fuel + Toll Cost Car	COSTCRTRC	Car	-0.05 (2.23)	-0.08 (4.90)	-0.07 (3.56)	-0.24 (3.80)	0.35 (6.03)	-0.21 (4.26)	0.39 (6.96)	-0.25 (4.43)	0.38 (7.36)
Parking Cost Car	COSTCRPC	Car	-0.23 (10.02)	-0.05 (7.50)	-0.06 (8.16)	-0.58 (5.08)	0.53 (5.73)	-0.21 (8.56)	0.20 (7.67)	-0.38 (10.64)	0.38 (11.62)
Travel Time Public Transport	TTPT	Bus, Train, Metro	-0.16 (2.33)	-0.01 (7.66)	-0.03 (14.80)	-0.09 (14.42)	0.08 (12.77)	-0.09 (15.59)	0.08 (12.88)	-0.08 (13.20)	0.05 (9.99)
Travel Time Car	TTCR	Car	-0.04 (11.80)	-0.02 (6.52)	-0.04 (12.10)	-0.16 (4.74)	0.06 (3.64)	-0.07 (7.11)	0.08 (7.92)	-0.07 (7.10)	0.07 (9.55)
Egress Time	EGTIME	All Alternatives	-0.06 (16.84)	-0.04 (9.23)	-0.05 (16.26)	-0.12 (10.06)	0.15 (9.63)	-0.14 (11.93)	0.15 (9.95)	-0.11 (10.65)	0.10 (4.98)
Transfer Public Transport	TRANPT	Bus, Train, Metro	-0.11 (3.88)	-0.09 (3.86)	-0.11 (4.07)	-0.29 (5.79)	0.38 (3.23)	-0.24 (5.24)	0.30 (2.65)	-0.24 (5.72)	0.11 (0.64)

			LPAA_MNL_BRExp	VL_MNL_Exp	RAM_MNL_Exp	LPAA_MML_BRExp		VL_MML_Exp		RAM_MML_Exp	
Headway Public Transport	FREQPT	Bus, Train, Metro	-0.01 (6.96)	-0.01 (5.90)	-0.01 (7.06)	-0.03 (9.50)	0.05 (12.65)	-0.03 (9.55)	0.05 (12.89)	-0.03 (9.89)	0.05 (12.89)
% Seat Public Transport	SEATPT	Bus, Train, Metro	0.44 (4.43)	0.30 (3.29)	0.39 (4.28)	1.07 (5.90)	1.60 (4.90)	0.99 (5.99)	1.84 (7.34)	0.93 (5.64)	0.91 (1.73)
Density Public Transport	STANDPT	Bus, Train, Metro	-0.19 (11.93)	-0.16 (11.89)	-0.18 (12.04)	-0.37 (11.98)	0.42 (9.12)	-0.37 (12.36)	0.39 (9.30)	-0.35 (12.13)	0.37 (8.62)
Experience Bus	EXPBS	Bus	0.30 (9.37)	-	0.37 (9.07)	0.34 (10.52)	-	0.29 (4.88)	-	0.42 (10.06)	-
Experience Train	EXPTR	Train	0.09 (3.70)	-0.11 (2.13)	0.11 (3.97)	0.07 (2.74)	-	-	-	0.09 (2.11)	-
Experience Car	EXPCR	Car	0.44 (16.36)	0.43 (8.99)	0.65 (23.54)	0.54 (15.22)	-	0.30 (3.24)	-	0.65 (13.97)	-
Risk Attitudes Travel Time Bus	ALPHABSTT	Bus	0.44 (3.73)	-	-	-	-	-	-	-	-
Risk Attitudes Travel Time Train	ALPHATRTT	Train	0.46 (3.72)	-	-	-	-	-	-	-	-
Risk Attitudes Travel Time Metro	ALPHAMTTT	Metro	0.60 (3.91)	-	-	-	-	-	-	-	-
Risk Attitudes Cost Bus	ALPHABSCS	Bus	-	-	-	-0.23 (2.53)	-	-	-	-	-
Risk Attitudes Cost Metro	ALPHAMTCS	Metro	-	-	-	-	-	-	-	-	-
Risk Attitudes Fuel+Toll Car	ALPHACRTRCS	Car	-	-	-	-	-	-	-	-	-
Risk Attitudes Parking Car	ALPHACRPCS	Car	0.65 (8.79)	-	-	0.20 (2.06)	-	-	-	-	-
Perceptual Conditioning Car	GAMMACR	Car	-	-	-	1.83 (2.96)	-	-	-	-	-
Concavity VL	CONC	All Alternatives	-	1.16 (5.35)	-	-	-	-	-	-	-

Table 11: Metro Rail dataset process heterogeneity models' estimates (t-values in parenthesis)

			PDP_BRExp			CRPHms_BRExp					
Number of Parameters Estimated			**47**			**35**					
Log Likelihood at convergence			**-5,007.53**			**-4,922.41**					
Log likelihood at zero			**-13,125.44**								
AIC			**1.068**			**1.047**					
Class Identification			**Class 1**	**Class 2**	**Class 3**	**-**					
Heuristic			RAM	VL	LPAA	LPAA, VL and RAM					
Behavioural Refinements			Y	N	N	Y					
Experience			Y	Y	Y	Y					
Class Membership (%)			37%	18%	45%	-					
Parameters	**Acronym**	**Alternatives**	**Mean**	**Mean**	**Mean**	**Mean**	**Std Dev**	**M_VL**	**M_RAM**	**S_VL**	**S_RAM**
Alternative Specific Constant Bus	ASCBUS	Bus	-2.18 (3.51)	-2.08 (4.23)	2.54 (5.15)	1.60 (2.63)	-	-	-	-	-
Alternative Specific Constant Train	ASCTRAIN	Train	-1.45 (2.78)	-1.03 (2.88)	2.86 (6.41)	2.44 (4.41)	-	-	-	-	-
Alternative Specific Constant Metro	ASCMETRO	Metro	0.14 (0.32)	-3.11 (7.88)	2.67 (6.5)	2.75 (5.27)	-	-	-	-	-
Alternative Specific Constant Car	ASCCAR	Car	-	-	-	-	-	-	-	-	-
Access Time	ACTIMEPT	Bus, Train, Metro	-0.05 (5.09)	-0.07 (5.45)	-0.06 (10.01)	-0.08 (8.93)	-0.22 (10.81)	-	-	-	-0.05 (3.63)
Fare Public Transport	COSTPT	Bus, Train, Metro	-0.25 (3.93)	-0.24 (4.15)	-0.42 (10.74)	-0.92 (10.36)	-0.82 (9.12)	-	-	-0.03 (6.84)	-0.05 (6.12)
Fuel + Toll Cost Car	COSTCRTRC	Car	-	-0.12 (2.13)	-0.23 (5.29)	-	0.18 (3.49)	-	-	-0.27 (5.53)	-
Parking Cost Car	COSTCRPC	Car	-0.34 (5.71)	-0.12 (4.80)	-0.24 (4.42)	-0.33 (6.77)	0.13 (3.55)	-0.002 (2.05)	-	-0.01 (4.74)	-
Travel Time Public Transport	TTPT	Bus, Train, Metro	-0.05 (6.57)	-0.02 (4.23)	-0.05 (11.29)	-0.10 (9.06)	0.03 (6.86)	-	-	-0.003 (3.53)	-
Travel Time Car	TTCR	Car	-0.14 (6.49)	-0.06 (5.03)	-0.03 (4.81)	-0.09 (7.47)	0.07 (8.10)	-	-	-	-

			PDP_BRExp			CRPHms_BRExp					
Egress Time	EGTIME	All Alternatives	-0.08 (6.05)	-0.04 (2.82)	-0.09 (10.51)	-0.13 (10.52)	0.12 (7.53)	-	-	-	-
Transfer Public Transport	TRANPT	Bus, Train, Metro	-	-0.18 (1.81)	-0.25 (4.86)	-0.26 (5.82)	-	-	-	-	-
Headway Public Transport	FREQPT	Bus, Train, Metro	-0.02 (3.04)	-0.01 (2.21)	-0.02 (5.41)	-0.03 (10.00)	0.09 (11.10)	-	-0.001 (5.09)	-	-
% Seat Public Transport	SEATPT	Bus, Train, Metro	0.58 (1.88)	0.80 (2.51)	0.63 (3.36)	1.59 (6.79)	-	-1.45 (3.88)	-	-	-
Density Public Transport	STANDPT	Bus, Train, Metro	-0.43 (7.40)	-0.19 (3.35)	-0.24 (7.65)	-0.33 (10.90)	0.53 (8.15)	-	-	-	0.26 (3.19)
Experience Bus	EXPBS	Bus	-	0.78 (6.88)	0.26 (4.59)	0.29 (9.94)	-	-	-	-	-
Experience Train	EXPTR	Train	-	-	-	0.09 (3.56)	-	-	-	-	-
Experience Car	EXPCR	Car	0.73 (20.61)	0.34 (2.85)	-	0.37 (7.49)	-	-	-	-	-
Risk Attitudes Parking Car	ALPHACRPCS	Car	-	-	0.75 (5.88)	-	-	-	-	-	-

Table 12: Northwest dataset process homogeneity models' estimates (t-values in parenthesis)

			LPAA_MNL_BRExp	VL_MNL_Exp	RAM_MNL_Exp	LPAA_MML_BRExp		VL_MML_Exp		RAM_MML_Exp	
Number of Parameters Estimated			**14**	**15**	**14**		**26**		**25**		**25**
Log Likelihood at convergence			**-6,170.72**	**-6,004.14**	**-6,169.89**		**-4,611.45**		**-4,702.90**		**-4,566.35**
Log likelihood at zero											**-7,838.25**
AIC			**2.731**	**2.657**	**2.730**		**2.047**		**2.087**		**2.027**
Parameters	**Acronym**	**Alternatives**	**Mean**	**Mean**	**Mean**	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Alternative Specific Constant New Light Rail	ASCLR	New Light Rail	1.76 (8.97)	1.37 (11.12)	2.10 (10.65)	5.48 (8.39)	-	1.76 (8.44)	-	4.74 (12.25)	-
Alternative Specific Constant New Heavy Rail	ASCNHR	New Heavy Rail	1.84 (10.83)	1.16 (12.10)	2.19 (12.48)	5.66 (8.89)	-	1.83 (10.63)	-	4.86 (13.19)	-
Alternative Specific Constant New Busway	ASCNBW	New Busway	0.63 (3.21)	-0.07 (0.51)	0.99 (4.85)	4.47 (6.83)	-	0.52 (2.53)	-	3.64 (9.44)	-
Alternative Specific Constant Bus	ASCBS	Bus	1.51 (8.86)	0.80 (8.03)	1.85 (10.39)	5.13 (7.98)	-	1.02 (5.83)	-	3.94 (10.57)	-
Alternative Specific Constant Busway	ASCBW	Busway	1.23 (7.19)	0.53 (5.43)	1.58 (8.94)	4.85 (7.51)	-	0.84 (4.72)	-	3.91 (10.71)	-
Alternative Specific Constant Train	ASCTRAIN	Train	1.31 (7.60)	0.66 (6.78)	1.68 (9.55)	4.15 (6.53)	-	0.53 (2.94)	-	3.38 (9.40)	-
Alternative Specific Constant Car	ASCCAR	Car	-			-	-				
Access Time	ACTIMEPT	Public Transport	-0.04 (13.76)	-0.02 (8.90)	-0.03 (13.97)	-0.07 (9.33)	0.11 (13.90)	-0.06 (7.93)	0.11 (13.77)	-0.04 (8.09)	0.09 (11.13)
Fare Public Transport	COSTPT	Public Transport	-0.25 (26.49)	-0.16 (13.50)	-0.17 (25.08)	-1.03 (11.62)	0.90 (15.05)	-0.47 (17.35)	0.50 (16.17)	-0.37 (17.46)	0.45 (14.99)
Fuel + Toll Cost Car	COSTCRTRC	Car	-0.17 (9.68)	-0.10 (8.49)	-0.11 (9.74)	-0.86 (12.40)	0.81 (19.05)	-0.41 (8.38)	0.50 (10.02)	-0.06 (2.25)	-
Parking Cost Car	COSTCRPC	Car	-0.03 (5.52)	-0.02 (4.45)	-0.02 (5.33)	-0.10 (5.88)	0.15 (8.43)	-0.05 (4.76)	-	-0.12 (5.86)	0.15 (5.70)
Travel Time Public Transport	TTPT	Public Transport	-0.04 (31.69)	-0.02 (8.65)	-0.03 (30.07)	-0.07 (22.46)	0.06 (19.39)	-0.07 (21.56)	0.07 (19.17)	-0.05 (17.79)	0.06 (16.24)
Travel Time Car	TTCR	Car	-0.03 (11.74)	-0.01 (5.04)	-0.02 (12.08)	-0.08 (11.02)	0.07 (15.58)	-0.07 (9.29)	0.07 (11.56)	-0.08 (10.33)	0.14 (12.86)

			LPAA_MNL_BRExp	VL_MNL_Exp	RAM_MNL_Exp	LPAA_MML_BRExp		VL_MML_Exp		RAM_MML_Exp	
Egress Time	EGTIME	All Alternatives	-0.04 (12.89)	-0.01 (7.28)	-0.02 (12.63)	-0.06 (7.11)	0.07 (6.10)	-0.04 (6.68)	0.08 (9.17)	-0.03 (7.06)	0.07 (8.37)
Headway Public Transport	FREQPT	Public Transport	-0.02 (2.10)	-0.04 (6.43)	-0.01 (2.01)	-0.03 (1.93)	0.19 (13.83)	-0.08 (4.59)	0.22 (14.93)	-0.05 (4.27)	0.19 (14.53)
Experience Bus	EXPBS	Bus	-	-	-	-	-	0.18 (3.00)	-	0.19 (4.27)	-
Experience Busway	EXPBW	Busway	-	-	-	-	-	0.21 (3.63)	-	0.11 (2.96)	-
Experience Train	EXPTR	Train	-	-	-	0.25 (9.46)	-	0.30 (5.77)	-	0.32 (8.98)	-
Experience Car	EXPCR	Car	-	-	-	0.17 (2.68)	-	-0.97 (4.59)	-	0.66 (11.43)	-
Risk Attitudes Cost Currently Available Modal Facilities	ALPHAEXCS	Bus, Busway, Train and Car	-	-	-	0.45 (9.54)	-	-	-	-	-
Risk Attitudes Cost New Modal Investments	ALPHANEXCS	New Light Rail, New Heavy Rail, New Busway	-	-	-	0.45 (9.24)	-	-	-	-	-
Concavity VL	CONC	All Alternatives	-	1.22 (7.57)	-	-	-	-	-	-	-

Table 13: Northwest dataset process heterogeneity models' estimates (t-values in parenthesis)

			PDP_BRExp			CRPHms_BRExp					
Number of Parameters Estimated			**44**			**32**					
Log Likelihood at convergence			**-4,864.47**			**-4,494.74**					
Log likelihood at zero						**-7,838.25**					
AIC			**2.167**			**1.999**					
Class Identification			**Class 1**	**Class 2**	**Class 3**	**-**					
Heuristic			RAM	VL	LPAA	LPAA, VL and RAM					
Behavioural Refinements			Y	N	N	Y					
Experience			Y	Y	Y	Y					
Class Membership (%)			15%	34%	50%	-					
Parameters	**Acronym**	**Alternatives**	**Mean**	**Mean**	**Mean**	**Mean**	**Std Dev**	**M_VL**	**M_RAM**	**S_VL**	**S_RAM**
Alternative Specific Constant New Light Rail	ASCLR	New Light Rail	-3.50 (3.90)	1.54 (6.49)	3.68 (4.76)	3.89 (7.69)	-	-	-	-	-
Alternative Specific Constant New Heavy Rail	ASCNHR	New Heavy Rail	-2.34 (4.42)	1.38 (7.13)	3.41 (4.24)	4.52 (9.35)	-	-	-	-	-
Alternative Specific Constant New Busway	ASCNBW	New Busway	-2.12 (3.06)	-0.64 (2.14)	3.12 (3.92)	2.96 (6.05)	-	-	-	-	-
Alternative Specific Constant Bus	ASCBS	Bus	1.36 (4.85)	-0.64 (2.23)	3.00 (3.74)	3.62 (7.65)	-	-	-	-	-
Alternative Specific Constant Busway	ASCBW	Busway	-0.26 (0.88)	-0.03 (0.14)	3.01 (3.81)	3.46 (7.24)	-	-	-	-	-
Alternative Specific Constant Train	ASCTRAIN	Train	-1.06 (2.45)	-1.40 (5.58)	3.17 (3.8)	2.70 (5.59)	-	-	-	-	-
Alternative Specific Constant Car	ASCCAR	Car	-	-	-	-	-	-	-	-	-
Access Time	ACTIMEPT	Public Transport	-0.10 (6.95)	-	-0.05 (7.99)	-0.03 (8.96)	-	-0.002 (7.85)	-	-	0.05 (14.98)
Fare Public Transport	COSTPT	Public Transport	-0.08 (5.85)	-0.06 (4.02)	-0.60 (14.89)	-0.64 (17.53)	0.47 (15.64)	-	-	-0.04 (21.39)	0.04 (18.63)

			PDP_BRExp			CRPHms_BRExp					
Fuel + Toll Cost Car	COSTCRTRC	Car	-0.04 (2.90)	-0.11 (4.45)	-0.09 (2.38)	-0.14 (6.11)	-	-	-	-	0.14 (8.25)
Parking Cost Car	COSTCRPC	Car	-0.03 (6.61)	-	-0.22 (4.77)	-0.21 (10.02)	-0.24 (13.90)	-	-	-	-
Travel Time Public Transport	TTPT	Public Transport	-0.02 (5.55)	-0.01 (2.92)	-0.08 (23.5)	-0.07 (12.62)	-0.08 (16.84)	-	0.0002 (2.65)	0.002 (2.28)	-
Travel Time Car	TTCR	Car	-	-0.01 (2.93)	-0.09 (5.02)	-0.12 (4.78)	0.08 (11.65)	-0.001 (2.84)	-	-0.01 (5.04)	-
Egress Time	EGTIME	All Alternatives	-	-0.01 (2.93)	-0.05 (7.14)	-0.06 (8.80)	0.06 (9.05)	-	-	-	-
Headway Public Transport	FREQPT	Public Transport	-	-0.03 (3.70)	-	-0.04 (2.68)	-0.34 (13.38)	-	0.01 (6.52)	-	-
Experience Bus	EXPBS	Bus	-	-	0.18 (3.57)	-	-	-	-	-	-
Experience Train	EXPTR	Train	-	-	0.19 (6.02)	0.29 (9.11)	-	-	-	-	-
Experience Car	EXPCR	Car	0.72 (5.81)	-	-0.28 (2.52)	0.19 (3.08)	-	-	-	-	-
Risk Attitudes Cost Currently Available Modal Facilities	ALPHAEXCS	Bus, Busway, Train and Car	-	-	0.11 (3.52)	-	-	-	-	-	-
Concavity VL	CONC	All Alternatives	-	1.38 (4.56)	-	-	-	-	-	-	-