PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

IDENTIFYING DISCRETE CHOICE MODELS WITH MULTIPLE CHOICE HEURISTICS

FELIPE A. GONZALEZ VALDES

Thesis submitted to the Office of Research and Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor in Engineering Sciences

Advisor:

JUAN DE DIOS ORTUZAR SALAS

Santiago de Chile, June, 2018

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

IDENTIFYING DISCRETE CHOICE MODELS WITH MULTIPLE CHOICE HEURISTICS

FELIPE ANTONIO GONZÁLEZ VALDÉS

Members of the Committee:

JUAN DE DIOS ORTÚZAR

LUIS RIZZI

RICARDO HURTUBIA

CRISTIÁN GUEVARA

BENJAMIN G. HEYDECKER Benjamin Heydecker

JORGE VÁSQUEZ

Thesis submitted to the Office of Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor in Engineering Sciences

Santiago de Chile, August, 2018

To all the people who dream

of a fairer city

ACKNOWLEDGMENTS

In first place, I would like to thank everyone that contributed somehow in allowing me to have the opportunity to participate in this program. Thanks Juan Enrique for convincing me to take this step. And to anyone that feels that contributed in my formation, thanks for helping me having the chance to develop myself. Furthermore, I would like to thank who was myself in the past; I do not know if nowadays I would be brave enough to embark in this five years journey.

I would like to thank Alejandra, my partner, who has always been there to support me, hold me back, and make me focus on the important things. Thanks to her, I have kept (or even gain) the touch with reality. She has picked me up when I have failed. And probably, the most important thing, she was been my partner throughout this journey.

Thanks to my supervisor Juan de Dios. He has always trusted in my capacities and skills, probably overestimating me. I would like to thank him for his attitude, even when he has been collapsed with his duties and overwhelming work load, he always finds time to help me. It has been an honour to work with Juan de Dios. Thanks for guide me to London, where I grew as a person.

I would like to thank my friends in the Transport and Logistic Department. Some have been with me since the undergrad and some I have met them in the postgraduate program; to all of them, thanks. Because of you, I have survived an arid research project that at the beginning of it, the applications where scarce. Thanks to you, I have finished my program with delay, but really enjoying the journey. Special thanks to Sebastián Muñoz, whose friendship was an important support.

I thank my parents. To my father who has always guided me through my life. To my mother, whose unconditional support has been key in giving me the confidence to live calmed. To you both, thanks.

I would like to thank Ben Heydecker. Thanks for believing in my project and receiving me in London, where I grew in a tremendous way. Moreover, your support and interest in this thesis meant important improvements and developments.

Thanks Sebastián Raveau, who has always shown interest in my research and my findings. Thanks to you, this thesis has developed an applied wing. Furthermore, thanks for being a friend with whom I have transited in this journey.

Thanks to my thesis committee. The rejection of my upgrade in the first instance allowed me to develop adequately my projects. Thanks for being so responsible in the task you have accepted.

To my friends in the student centre. Thanks to you, I developed myself in the program in a sphere that I always dreamed. Thanks VIVE CAI.

I would like to thank my out of Transport Department friends. Thanks Toby's Club for bringing happiness to my journey; you have been there in the darkest hours. To my undergrad friends Catalina and Trinidad, you have always been proud of myself; now also I am. To my friend Concha, who has always overestimated me, at least now I deserve your best thoughts.

To the people I met at London and Leeds. Because of you my friends, my stay at the United Kingdom was less hard. I hope to meet you again in life.

Thanks to the researchers that have always had a good disposition towards my questions, like Ricardo Alvarez Daziano. Thanks to the researchers that have showed interest in my work, like Stephan Hess, and to David Palma for materializing this interest.

Finally, if you feel somehow that contributed in my journey, thanks for being part of it.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

IDENTIFICACION DE MODELOS DE ELECCION DISCRETA CONSIDERANDO MECANISMOS DE ELECCION HETEROGENEOS

Tesis enviada a la Dirección de Investigación y Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

FELIPE A. GONZALEZ VALDES

RESUMEN

Entender y predecir el comportamiento de las personas es clave in diferentes áreas, tales como políticas públicas y marketing. Los modelos de elección discreta son una de las herramientas diseñadas para entender estos comportamientos. El núcleo de estos modelos es la heurística de elección; ella representa una forma en que se procesan las alternativas. Su correcto entendimiento es crucial para representar adecuadamente los comportamientos.

Numerosas heurísticas se han propuesto en la literatura. Con el objetivo de categorizarlas y entender su relación, hemos creado un marco teórico que las analiza en tres dimensiones: evaluación absoluta/relativa, simplificación de alternativas y simplificación de atributos. De estas heurísticas, hemos seleccionado cuatro para participar en nuestros experimentos: Maximización de la Utilidad Aleatoria (RUM por su nombre en inglés), Minimización del Remordimiento Aleatorio (RRM), Eliminación por Aspectos (EBA) y *Satisficing*.

Hemos desarrollado dos contribuciones respecto a Satisficing y EBA. Para Satisficing, construimos el modelo *Stochastic Satisficing* (SS), que es el primer modelo que implementa completamente la teoría usando datos típicamente disponibles. Para EBA, proponemos un enfoque analítico para acelerar su estimación.

Entendiendo que en una población pueden coexistir distintas heurísticas, se han propuesto modelos con múltiples heurísticas. Lamentablemente, su estimación – que usa clases latentes- ha mostrado problemas de identificabilidad. Para entender este problema, hemos estudiado analíticamente la identificabilidad, concluyendo que está gobernada por la diferencia de comportamiento de las heurísticas en la muestra; finalmente, obtuvimos una métrica simple e interpretable para dicha diferencia.

Habiendo estudiado la identificabilidad teóricamente, comprobamos sus alcances en la práctica. Estudiamos el impacto de distintas heurísticas, tamaños muestrales y grados de correlación entre los factores que afenta la elección de la heurística y la alternativa. Concluimos que, para nuestro contexto, RRM no es identificable de RUM, SS lo es para muestras grandes (40.000) y EBA siempre es identificable de RUM.

Dada la posibilidad de identificar estos modelos, proponemos una metodología que, mediante nuestro Modelo de Heurísticas Mixtas (MHM), facilita la búsqueda de las heurísticas subyacentes y su formulación. El MHM es un modelo de clases latentes con función de pertenencia de clase mixta. Permite encontrar las heurísticas presentes con mayor precisión que el modelo de clases latentes tradicional. Así, sin modelar la función de pertenencia de clase, las heurísticas subyacentes pueden ser encontradas y formuladas.

Una vez que modelos candidatos son encontrados, se debe aplicar algún criterio para seleccionar al mejor. Mediante la experimentación con un par de modelos candidatos, concluimos que, si el objetivo es entender el fenómeno, entonces criterios sobre la base de estimación que penalicen débilmente los parámetros adicionales deberían ser usados. Si el objetivo es predecir, entonces se debieran preferir criterios sobre una base de validación.

En esta tesis hemos mostrado que es factible estimar modelos con múltiples heurísticas. También, desarrollamos metodologías para encontrar los modelos más explicativos y seleccionar el más útil. Si bien, esta tesis disminuye la dificultad de estimar estos modelos, se requiere más investigación para obtener conclusiones generales.

Miembros de la Comisión de Tesis Doctoral

Juan de Dios Ortúzar Salas

Luis Ignacio Rizzi Campanella

Ricardo Hurtubia González

Cristián Angelo Guevara Cue

Benjamin G. Heydecker

Jorge Vásquez Pinillos

Santiago, Agosto, 2018

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

IDENTIFYNG DISCRETE CHOICE MODELS WITH MULTIPLE CHOICE HEURISTICS

Thesis submitted to the Office of Research and Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor in Engineering Sciences by

FELIPE A. GONZÁLEZ VALDÉS

ABSTRACT

Understanding and forecasting the behaviour of individuals is key in different realms of society such as public policy and marketing. Discrete choice models are an important econometric tool for understanding these behaviours. The kernel of a discrete choice model is its choice heuristic, which represents the way individuals process the alternatives. A correct representation of their heuristics is key to successfully represent their behaviour.

Several heuristics have been proposed in the literature. We have created a framework that allows to organise them and understand their similarity across three dimensions: absolute/relative evaluation, simplification of attributes and simplification of alternatives. We have selected four heuristics for our experiments: Random Utility Maximization (RUM), Random Regret Minimization (RRM), Elimination by Aspects (EBA), and Satisficing. For the last two of these, we made a particular contribution. For Satisficing, we created the

Stochastic Satisficing (SS) model, which is the first model that wholly implements satisficing theory using normally available data. For EBA, we proposed an analytical approach to increase its estimation speed.

Admitting that not every individual in a population may use the same heuristic, multiple heuristics models have been proposed in the literature. Unfortunately, their estimation using latent classes has given rise to identifiability issues. We studied the problem analytically under a maximum likelihood framework. We concluded that identifiability is closely related with the behavioural differences among the heuristics in the data; we obtained a readily and interpretable measure of this difference.

We tested the theoretical findings in a quasi-real transport context. We simulated fictitious individuals choosing real alternatives under our three experimental dimensions. We concluded that, for our context, RRM was non-identifiable from RUM, SS was identifiable from RUM for the larger samples (40.000 individuals), and EBA was always identifiable.

Given that it is possible to identify multiple heuristics, we proposed a methodology that, by using our Mixed Heuristics Model (MHM), facilitates finding the heuristics present in a sample and their formulation. The MHM is a latent class model with a mixed class membership function that allows to find the most likely used heuristics with higher accuracy than a non-random latent class model. This way, without modelling the class membership function, the underlying choice heuristics can be found and modelled with a traditional approach.

Once several models are available, a criterion must be used to select the best one. We tested competing pairs of heuristics with different degree of identifiability. We concluded that if the objective is understanding the underlying phenomena, in-sample criteria that do not penalise heavily additional parameters should be used, promoting more explicative models. Conversely, if the objective is forecasting, out-of-sample validation might be the best approach to promote more robust models.

Through our work, we showed that it is feasible to estimate multiple heuristics models. We also provide tools that allow finding the most explicative models and, among them, choose the most useful one. Therefore, after this thesis, the complexity surrounding the use of multiple heuristics models should decrease. Nonetheless, more research is needed to understand the degree of identifiability of these models in different contexts, so that general conclusions regarding the selection of heuristics can be obtained.

Members of the Doctoral Thesis Committee:

Juan de Dios Ortúzar

Luis I. Rizzi

Ricardo Hurtubia

Cristián A. Guevara

Benjamin G. Heydecker

Jorge Vásquez

Santiago, August, 2018

1. INTRODUCTION

Understanding and forecasting the decision of individuals is essential in different realms of society. For public policy, for example, it is necessary to value people's preferences and predict their reaction towards policies affecting their behaviour. Alike, but in the private sphere, in most markets it is important to understand consumers' choices for designing new products and foresee their reaction towards different companies' marketing strategies. One of the available techniques for studying people's preferences and forecasting their behaviour is the use of econometric models.

Discrete choice modelling is an econometric tool designed to model the responses of decision makers (DMs) when confronting discrete choice sets. In principle, every output obtained from a discrete set of possible outcomes may be modelled using this tool. These models are particularly useful for representing people's choices from these sets because of their behavioural interpretation of the individual. Thus, discrete choice models have been widely used to model people's preferences in different contexts such as the purchase of items (Fader and McAlister, 1990; Adamowicz and Swait, 2013; Beck et al., 2013; Hensher et al., 2013; Palma et al., 2016), housing and location decisions (Martínez et al., 2009; Greene et al., 2017), opinion toward public policy (Araña et al., 2008), and transport mode and route choice (Vovsha, 1997; Ortuzar and Willumsen, 2011).

The kernel of an econometric model is its functional form, which is assumed by the analyst to represent the phenomenon. In discrete choice models, the functional form represents a choice heuristic, which is a representation of a plausible way of choosing. Initially, only simple behaviours could be correctly modelled –as we detail in Chapter 2– but later, different models have been developed to treat increasingly complex phenomena.

Among the complexities faced, one is how to model populations with behaviours of a different nature. Several choice heuristics have been proposed that explains different kind of behaviour. These choice heuristics have been materialized, for example, in the random utility maximization – RUM – model (Lancaster, 1966; McFadden, 1973), the elimination by aspects – EBA – model (Tversky, 1972a, 1972b), and the Satisficing model (Simon, 1955) as the most iconic ones.

Another issue is that within a given choice heuristic some decisions may be more complex to model. For example, the popular RUM model's basic formulation assumes independent and homoscedastic errors and cross-sectional data. More complex versions of this model enables to work with correlated alternatives, panel data, heteroscedasticity, and taste variations (Ortuzar and Willumsen, 2011).

A final complexity –that is the target of this thesis– is modelling a population where DMs may use different choice heuristics. In the presence of multiple choice heuristics, a model with a single choice heuristic could misunderstand the phenomenon and fail in prediction.

Among the several proposed heuristics in the literature, RUM is the most widely used. Numerous reasons could explain its popularity such as the concordance with classical microeconomic theory, a set of well-developed and desirable properties, and a logical representation of DMs. Despite its popularity, in reality, the high psychological burden implied in using RUM in complex decisions could trigger the use of different choice heuristics by DMs. Indeed, behavioural decision theory suggests that the way DMs choose is shaped by the interaction of their cognitive capabilities, personal goals, decision complexities, and the characteristics of the task (Newell and Simon, 1972; Bettman, 1979; Peterson et al., 1979; Denstadli et al., 2012). Therefore, the same context may trigger different heuristics in different DMs.

Some attempts have been previously made to model multiple choice heuristics (Araña et al., 2008; Hess et al., 2012; McNair et al., 2012; Adamowicz and Swait, 2013). When trying to characterize the decision of using a choice heuristic, identifiability problems have arisen (Leong and Hensher, 2012b; Hess and Stathopoulos, 2013). All studies found in the literature have tried to model real data, which has given few sparkles about the reason for the non-identifiability problem. This thesis tackles this problem by studying this phenomenon theoretically, by first analytically exploring the properties of the main modelling approach and then, empirically by using a controlled simulated environment given by the choice heuristics, proportions, and variables affecting decisions.

Among the several edges of the problem, this study aims to identify whether it is possible to identify different choice heuristics or the conditions that enable such identification. Once identifiability is attained, how to choose the different choice heuristic present in a single model from a wide variety of options is analysed. Finally, we address how to choose a model among several competing ones under weak and strong identifiability estimations.

1.1. Objectives

Our general objective is to study the modelling of a population endowed with multiple choice heuristics. As stated previously, the current state of the art has not been able to properly characterize DMs' decision processes nor demonstrate the reasons for such a failure. Thus, this thesis intends to give answers to these problems.

Given the failure of previous authors to identify explicative models for the selection of choice heuristic, the first main objective is related to the feasibility of identifying such models and may be stated as follows.

a) Analyse if it is possible to identify the choice heuristics used by a population and understand the conditions that enable their identification.

Given the identification at a population level, several ways to analyse it at an individual level may be proposed which are detailed in this thesis. Then the second main objective is to:

b) Analyse if it is possible to identify –probabilistically— the choice heuristics being used by a decision maker.

If the first and second objective are accomplished, a useful model that interprets DMs' behaviour at the population and individual levels will be obtained. Nonetheless, as identifiability issues are frequent, it is possible that several groups of heuristics can interpret the choices with similar fit performance. Therefore, our last main objective is to:

c) Analyse which statistical techniques are useful for selecting the most useful model depending of the objective: understand the underlying behaviour or forecast.

For the conclusions to be as general as possible, the findings must be analysed with different sample sizes, number and proportions of choices heuristics, and responses per decision maker.

1.2. Hypotheses

To study the above main objectives, we will test the following hypotheses:

a) It is possible to identify the different choice heuristic used by DMs even if these consider the same attributes. Therefore, identifiability is accomplished just by the different interpretation that heuristics give to alternatives' attributes.

b) For finite sample sizes, the probability of using a heuristic will be recovered – probabilistically and to a reasonable extent—both at the population level and at the individual level. This implies, that the model will be able to identify the probability that each individual has of using each choice heuristic.

c) The higher the number of observations per decision maker, the higher the accuracy of the estimation of individual probabilities.

d) When models with several choice heuristics are available, in-sample techniques may systematically fail to identify the real underlying model. Conversely, out-of-sample techniques will identify correctly the real choice heuristic.

1.3. Methodology

Given the objectives of the thesis and the hypotheses considered, most of the analyses in this thesis are empirical. Nevertheless, to be able to analyse the results, first a theoretical framework is developed. The theoretical findings enable us to better understand the dynamics of the model and, therefore, adequately interpret the findings obtained in the subsequent empirical experiments. To develop this framework, an analysis of the estimated model is made: we study the optimum first order condition in a maximum likelihood context and the conditions that enable us to obtain a useful covariance matrix.

Once the theoretical framework is developed, we test the hypotheses empirically; for this, it is required to have control of the choice heuristics. Therefore, we focus mainly in the use of synthetic populations. A synthetic population creates a testing environment to understand the model's properties and the impact of the controlled variables in them. Indeed, to test the properties of models in finite samples, the use of synthetic populations is frequently used (e.g. Godoy and Ortuzar, 2008; Raveau et al., 2010; Daly et al., 2012). We vary the experimental conditions and analyse the results of the estimated models under these changes; for example, we consider different sample sizes, choice heuristics and correlation structures.

To create the synthetic data, first a fictitious population is created, which behaves according to the desired rules. The fictitious DMs are represented by a set of parameters and available choice heuristics. In each experiment, the set of plausible parameters are based on a real dataset, but modified for the intended purpose.

Once the synthetic population is created, it is given a realistic choice set from which DMs choose. With the objective of having conclusions closely aligned to reality, quasi-real choice sets are used (i.e. real choice scenarios but presented to the fictitious DMs). These choice sets are obtained through random sampling of complete choice sets from a real dataset. Detail about these choice sets are given in Chapter 3.

Finally, to show the applicability of the developed methodologies and model's analysis, existing real dataset are also used. Here, it is not intended to make a deep analysis of the phenomena, but only to show a real application.

1.4. Scope of the Thesis

The definitions of choice heuristics in the literature do not present important variations, but the limits of what split a heuristic into two different ones is blur. Under our interpretation, certain phenomena that affect the sensitivity of DMs are considered by certain authors as standalone heuristics; we do not differentiate these from the utility maximisation principle and these will not be investigated further. Some authors consider hysteresis and habit as choice heuristics (Adamowicz and Swait, 2013). Other authors consider the phenomenon in which DMs build their preferences along the experiment as a stand-alone heuristic (Hensher and Collins, 2011; Balbontin et al., 2017). Similarly, several choice models have been developed to specifically account for prospect theory (Kahneman and Tversky, 1979). This thesis does not work with prospect theory directly, but makes use of some choice heuristics that could indirectly account for it; the same is the case of modelling certain specific effects such as the decoy effect (Fukushi, 2015; Guevara and Fukushi, 2016). In conclusion, all these phenomena are considered within a choice heuristic, rather than standalone mechanisms.

Even though we develop several theoretical analyses, this thesis focuses on analysing the multiple-choice heuristics' problem in an empirical way. Therefore, although the empirical analysis is done as broad as possible in an attempt to obtain the most general conclusions, some of the conclusions attained are not generalizable for every choice dataset.

We do not intend to characterise the reason or variables affecting the selection of the choice heuristics of DMs in real experiments. We rather study the model's capacity to capture these phenomena –if existent at all– and understand the model's behaviour.

1.5. Thesis Structure

This thesis starts with an analysis of the literature, which is summarised in the following points:

1. First, we analyse two model estimation methods: maximum likelihood estimation and Bayesian estimation. Advantages of these two methods are explored, and both are used throughout the thesis.

2. We analyse different choice heuristics coming from different fields of study such as transport engineering, marketing, and psychology; in every case the most representative choice heuristics are selected.

3. Techniques for model selection are deeply analysed. We study hypothesis test developed in classical statistics, information criteria coming from information theory, and we end by analysing out of sample validation techniques.

4. Finally, we discuss the experience reported in the literature regarding the modelling of multiple choice heuristics.

We use two datasets in this thesis. With the aim of not having to describe them in each chapter, we analyse them in Chapter 3. Moreover, we detail there the way the datasets are used to create pseudo-synthetic scenarios.

After the literature analysis, we identified potential contributions regarding two heuristics. Therefore, in Chapter 4 we further developed an important contribution regarding the Satisficing heuristic (Simon, 1955) and a smaller one in relation to the Elimination by Aspects heuristics (Tversky, 1972a, 1972b).

In Chapter 5, we conduct the theoretical analysis of the multiple heuristics model. First, we analyse a binary case that helps to enlighten the underlying phenomena; later we generalise it to the multivariate case. In both cases, we start by analysing the optimum first order conditions on the likelihood function. Then, we explore the second order conditions and their relationship to the covariance matrix.

In Chapter 6, we start the empirical analysis of multiple heuristics in this thesis. We analyse the identifiability of a latent class multiple heuristics model at the population level. Here we explore several dimensions that may affect model identifiability. Within a two heuristics model we analyse the type of choice heuristics involved in the model, sample size, number of alternatives in the choice set, and the proportion of each heuristic in the sample. We end this chapter by analysing a three heuristics case.

Chapter 7 is based on the findings of the previous chapter. Here, we analyse the identifiability of those heuristics where identifiability was more probable at an individual level. We finally apply the findings to a real sample.

In Chapter 8, we attain our last objective: we analyse the case where different pairs of heuristics can represent the same in-sample behaviour. We analyse in sample and out of sample indicators in cases where weak and strong identifiability is achieved and discuss their use.

Finally, conclusions of this thesis are addressed in Chapter 9. We analyse the main findings and its implications for choice modelling. We also discuss the limitations of this study and how it may affect the results. We hypothesise about the impact of the variation on the imposed experimental conditions. Finally, we end by addressing different avenues that this thesis has opened for future research.

2. REVIEW AND ANALYSIS OF LITERATURE

In this chapter we analyse three main topics. First, estimation techniques, particularly those used in this thesis (maximum likelihood estimation and Bayesian estimation). Then, we describe several choice heuristics and organise them in a form suitable for discussion. Finally, we study techniques for model selection.

2.1. Methods for Estimating Discrete Choice Models

In this thesis, we attempt to estimate models of which identification is not guaranteed; therefore, the estimation technique chosen must consider this issue. We examine two estimation techniques: maximum likelihood estimation, which is the most popular alternative in transport engineering, and Bayesian estimation, which is interesting because there is no maximization problem involved. Both are explained as follows.

2.1.1. Maximum likelihood estimation

In maximum likelihood estimation, we assume that the vector of outputs x and model parameters θ have a joint density function $f(x,\theta)$ given by (2.1). In it, f represents the model.

$$\Pr(x,\theta) = f(x,\theta) \tag{2.1}$$

The likelihood function measures the plausibility that the parameter θ of a model acquires a certain value in light of the data (2.2).

$$\mathcal{L}(\theta|x) = f(x|\theta) = \Pr(x|\theta) \tag{2.2}$$

Specifically, the likelihood function describes the probability of occurrence of the observed outcome x conditional on a model structure and model parameters θ. The maximum likelihood estimation tries to find the model parameters that maximise the observed outcome (2.2). If all i observations are independent, then (2.2) is reduced to (2.3).

$$\mathcal{L}(\theta|x) = \prod_i \Pr(x_i|\theta) \tag{2.3}$$

Given that the multiplied term in (2.3) is a probability –which $\in$ $[0,1]$– and that usually many DMs are considered, it is numerically hard to maximise (2.3) due to its small magnitude. Then, the expression maximised is the logarithm of the likelihood or log-likelihood given by (2.4).

$$l(\theta|x) = \sum_i \log(Pr(x|\theta)) \tag{2.4}$$

If the model is correctly specified, the model parameters distribute asymptotically Normal (Ortuzar and Willumsen, chap. 8, 2011; Train, chap. 8, 2009), as in:

$$\hat{\beta} \xrightarrow{d} N(\beta, -\boldsymbol{H}^{-1}) \tag{2.5}$$

In (2.5), $\boldsymbol{H}$ represent the hessian matrix of the log-likelihood function as stated in (2.6). The negative of $\mathbf{H}$ is also called the *information matrix* (Ortuzar and Willumsen, 2011, chap. 7; Train, 2009, chap. 8). When the model is not correctly specified the Hessian matrix in (2.6) is replaced by the *robust Hessian matrix* (Train, 2009, chap. 8).

$$\boldsymbol{H} = \frac{\partial^2 l(\theta)}{\partial \theta^2} \tag{2.6}$$

Frequently, the maximum likelihood can hardly be calculated analytically and numerical maximisation is needed. This kind of optimisation is suitable for single heuristic models, like those in Chapter 4. Unfortunately, it can pose a problem for weakly identifiable models, as those being characterised in Chapters 5 to 8.

Throughout this thesis, we assess the degree of identifiability of the models. In maximum likelihood, non-identifiability is detected by means of a non-invertible hessian or negative standard deviation estimates. However, there are several reasons that could explain these phenomena: i) poor asymptotic behaviour of the estimates (2.5), ii) numerical approximation algorithms are not at the optimal point, and iii) the model is, indeed, non-identifiable. Even though the first two problems can be tackled, they are non-dismissible.

To estimate the models via maximum likelihood, the *R* software (R Core Team, 2016) with the *Maxlik package* (Henningsen and Toomet, 2011) are used in this research. In the Maxlik package –as well as in other packages–, all the optimisers calculate the robust hessian matrix. Most of the available optimization methods are algorithms that work on a continuous space solution; while one of the methods, *Simulation Annealing*, works in a non-continuous space solution.

The algorithms for non-continuous space optimisation are normally not suitable for maximum likelihood estimation. Preliminary simulations performed by us indicate that although Simulation Annealing fails in finding the optima, the solutions provided are acceptable. Two problems with this algorithm are that the “optimal point” identified changes from estimation to estimation and, because the point is suboptimal, the hessian matrix may be non-invertible. The latter problem implies an impossibility of obtaining a covariance matrix for the estimates. Yet, if the space of the likelihood is non-continuous, like in the estimation of thresholds in an Elimination by Aspects model, the only available method in this package is Simulation Annealing.

The algorithms for continuous space optimisation are normally more suitable than non-continuous space algorithms. However, they tend to be captured in local optima, which is a frequent phenomenon in latent class models, which are the main models used in this thesis.

Then, maximum likelihood poses two problems in the context studied in this thesis. First, the impossibility of the inversion of the hessian matrix or the presence of negative standard deviation estimates cannot be uniquely linked to the non-identifiability of the model as a cause. And second, the most suitable algorithms for maximum likelihood tend to be captured by local optima. Therefore, an alternative estimation procedure is considered as an option for multiple heuristic models.

2.1.2. Bayesian estimation

Bayesian estimation identifies the most probable distribution of the parameters or posterior distribution in light of the data given a distribution of plausible values or prior. It is based on the Bayes theorem expressed in (2.7).

$$\Pr(\theta|y) = \frac{\Pr(\theta)\Pr(y|\theta)}{\Pr(y)} \tag{2.7}$$

The first component $\Pr(\theta)$ is a probability distribution of the parameters; its density function incorporates all available knowledge or prior belief. The second element $\Pr(y|\theta)$ is the likelihood function, as stated in (2.2), which relates the probability of generating the data under the model and parameters. The last component $\Pr(y)$ is the data density and is usually omitted since the model can be calculated with (2.8).

$$\Pr(\theta|y) \propto \Pr(\theta)\Pr(y|\theta) \tag{2.8}$$

Despite the existence of simple cases where the convolution of the likelihood and the prior have a closed form –in this case known as conjugate prior–, in most cases (2.8) cannot be calculated analytically (Gelman et al., 2013, cap. 12). Normally, to obtain useful statistics such as mean and variance or the empirical shape of the posterior, numerical integration is needed. Several techniques for calculating this convolution exists, such as grid-approximation, maximum *a posteriori*, and Markov Chain Monte Carlo (McElreath, 2012). The latter is the most general technique and is the one used in this thesis.

Markov Chain Monte Carlo (MCMC) is a stochastic method for numerical integration (Gelman et al., 2013). This technique samples from the desired distribution ($\Pr(\theta|y)$) to obtain an approximation of the parameter distribution.

There are several algorithms that implement MCMC, such as Metropolis Hastings, Hamiltonian sampling, and Gibbs sampling. In this thesis, we use Gibbs sampling[1] as implemented in the JAGS software (Plummer, 2003) connected to R through the RJags package (Plummer, 2016).

Gibbs sampling sequentially samples one parameter of the model conditional on the last values sampled for rest of the parameters (Gelman et al., chap. 11, 2013). Therefore, the parameter distribution depends exclusively on the current value of the rest of the parameters and not upon its history; hence, the sequence of sampled parameters is a Markov chain. The sampler sequentially obtains observations for each parameter of the model until the desired number of iterations or convergence is reached.

For the Gibbs sampler –or any other sampler–to be able to approximate (2.7), the Markov chain must be in a stationary state. Depending on the structure of the Markov chain and the initial points, different iterations must be performed before the chain reaches a high joint density function point, which is usually named burn-out period (Godoy and Ortuzar, 2008).

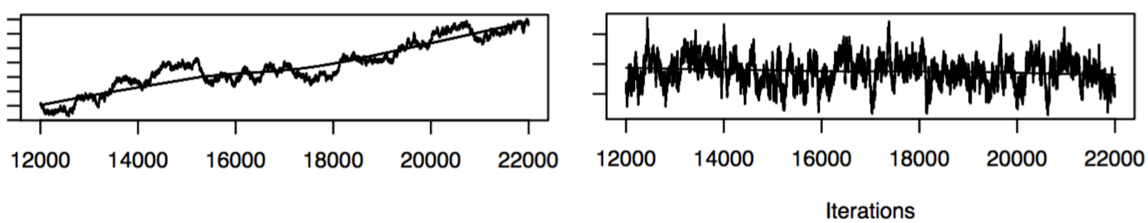
---

[1] We also tested the STAN software (Stan Development Team, 2015) that implements Hamiltonian sampling. For simple probability functions, STAN is faster than JAGS; however, if every individual has a structurally different probability function, like in the EBA model, the compilation time can be extremely large.

Once the parameters are in the high density function point, all sampled values may be used (Kass et al., 1998).

To analyse convergence, several tools may be used. Even though semi-automatic tools exist, the manual inspection of the Markov chains is recommended (Lunn et al., 2012). Convergence is analysed using a trace plot of the estimates and a density graph of the parameters. The trace plot of a parameter shows the value sampled for it in each iteration. Two examples are presented in Figure 2-1.

Figure 2-1 Trace plot of a Markov Chain



The trace plot is useful to determine whether the estimates of the Markov chain are stable in mean and variance. The left panel represents a parameter that does not have a stable mean; whereas the right panel shows parameter with stable mean and variance. A stable variance is represented by a non-varying dispersion of the trace plot. Shocks within the trace plot are expected, since low density events should be also represented, but no tendency should be obtained.

The density graph of the parameters plots the approximate density function obtained by sampling. Given the structure simulated in this thesis with unique underlying "real" parameters in each simulation, we expect a soft and unimodal density function (i.e. with only one peak). If the density function is not soft, then more sampling is needed. If the density function is not unimodal, then the Markov chain has not reached convergence or has a mixture of convergence and non-convergence sampling.

2.2. Choice Heuristics

The kernel of a discrete choice model is its choice heuristic, that is the procedure that DMs are supposed to follow to choose an alternative from a choice set. In other words, the choice heuristic is the set of rules that links certain sensitivities or parameters to a specific choice. To represent different behaviours, different heuristics have been proposed in the literature. Indeed, some heuristics try to represent a specific psychological theory (Simon, 1955; Kahneman and Tversky, 1979), while others represent different degrees of rationality (Chorus et al., 2008; McFadden, 1973).

In the literature, the terms of heuristic and choice model have been used normally as equivalent terms. Even though this is correct when a single heuristic exists, when multiple heuristics are embedded in the same model, a distinction is needed. In this thesis, and consistent with some literature (e.g. Cantillo and Ortúzar, 2005; Kivetz et al., 2004), the heuristic will be taken as the specific set of rules or mechanisms that DMs follow to choose, while the model, is a mathematical representation of one or more choice heuristics.

We analyse several choice heuristics that have been proposed and organise them using three criteria. Even though not all choice heuristics present in the literature are analysed, an extensive number of them is considered.

The first criterion is whether the choice heuristic evaluates alternatives in a comparative (*relative evaluation*) or in an absolute (*absolute evaluation*) fashion. Heuristics in the former group are contextually dependant and generally try to apply some concepts of Prospect theory (Tversky and Kahneman, 1991). Heuristics in the latter group are *economically rational* and exhibit the property of independence of irrelevant alternatives.

The second criterion analyses if all alternatives' attributes are evaluated (*total evaluation*) or if only a subset of them (*limited evaluation*) is considered. Heuristics in the limited evaluation cluster implement strategies that operate in contexts where a total evaluation strategy would involve high cognitive burden due to the number of attributes involved.

The last criterion is designed to verify if the individual chooses applying several processes or a single straightforward one. Implicit choice set heuristics or two-stage heuristics first reduce the size of the choice set (i.e. screens alternatives) and later chooses with an explicit choice set heuristic. An explicit choice set heuristic selects an alternative from the choice set using a single procedure. Moreover, even though any explicit choice set heuristic can be transformed into an implicit choice set heuristic by adding a screening stage, to the best of our knowledge, no relative evaluation two-stage choice heuristic has been proposed.

The choice heuristics presented are organised according to these criteria in Tables 2-1 and 2-2. These tables enable to visualise the kind and degree of similarity between the analysed heuristics and, furthermore, offer a framework to analyse/organise other choice heuristics. This framework allows us to select a subset of choice heuristics to test in the experiments.

Table 2-1 Organisation of choice heuristics with explicit choice sets

Explicit choice set	Total evaluation		Limited evaluation
Absolute evaluation	RUM		EBA LEX
	RAM LA RUM-RRM	MCD	
Relative evaluation	RRM CC		

Table 2-2 Organisation of choice heuristics with implicit choice sets

Implicit choice set	Total evaluation		Limited evaluation
Absolute evaluation	STS	HTS	Satisficing
Relative evaluation			

Finally, the following conventions will be used in this thesis. First, the modeller will be referred to as female, and the decision makers as males. Second, because the choice heuristics come from different bodies of knowledge (e.g. marketing, psychology, statistics and engineering), the nomenclature used is not straightforward and is stated in Appendix A.

2.2.1. Random Utility Maximization (RUM)

Random Utility Maximization (RUM) is a choice heuristic where DMs choose the alternative that maximise their utility. The utility of a given alternative i for individual q (U_{qi}) is a construct used by individuals to summarise all the characteristics of an alternative into a single figure of merit. When applied, the utility maximization problem (or consumer problem) is solved and an indirect utility function is obtained, which is usually the expression estimated (for further details see McFadden,1981). The modeller assumes that there is a part of the utility that she can identify (V_{qi}) –which is frequently linear and additive in parameters– while the other part is considered a stochastic error (ϵ_{qi}) as in (2.9).

$$U_{qi} = V_{qi} + \epsilon_{qi} \tag{2.9}$$

Depending on the distribution of the error terms, the model obtained varies. A Normal error term generates a Probit model (Daganzo, 1979). Independent and identical distributed Gumbel errors generate the Multinomial Logit (MNL) model (McFadden, 1973), which has been for many years the most popular formulation (2.10):

$$\epsilon_i \sim Gumbel(0, \sigma^2)$$

$$P_{qi} = \frac{\exp\,(\lambda \cdot V_{qi})}{\sum_{j \in I} \exp\,(\lambda \cdot V_{qj})} \tag{2.10}$$

$$\lambda = \frac{\pi}{\sigma\sqrt{6}}$$

Because the MNL model has several assumptions that are easily violated (e.g. homoscedasticity and independence of observations and errors), several alternative models have been proposed. A more flexible generalisation of the MNL, the nested logit (Williams, 1977; Daly and Zachary, 1978), works with partially correlated alternatives. A further but less popular improvement is the cross-correlated or cross-nested logit (Vovsha, 1997; Bhat, 1998; Ben-Akiva and Bierlaire, 1999). Furthermore, computing development has enabled the estimation of even more flexible models like the mixed logit model (Cardell and Reddy, 1977; Train, 1998). The mixed logit model enables a general covariance structure (Train, 2009) and can approximate any discrete choice model derived from RUM as closely as one pleases (McFadden and Train, 2000).

Developments over the MNL model have not only addressed the error term. Modifications of the MNL's utility function have been applied also, for example to create the dogit model (Gaundry and Dagenais, 1979), which allows to handle captives users. Further improvements in the modelling of non-measurable (latent) variables have been also some important developments (Ben-Akiva et al., 2002, 2012; Raveau et al., 2010). Finally, latest development of the RUM heuristic allows to model DMs considering the simultaneous decision of choosing a discrete alternative and, conditional on it, evaluate a continuous consumption level (Bhat, 2005, 2008; Pinjari and Bhat, 2010; Calastri et al., 2017b, 2017a).

Even though the MNL model requires strong assumptions, given that it is the most popular and simple model, it is the one used that will be used in this thesis. Thus, from now on the MNL model and the RUM choice heuristic will be used unequivocally.

2.2.2. Elimination by Aspects (EBA)

Elimination by Aspects –EBA– (Tversky, 1972b, 1972a) is a choice heuristic applicable in complex situations where DMs face an overwhelming amount of information. EBA is a non-independent random utility model (Tversky, 1972a) that sequentially discards alternatives based on their attributes. Even though in its general form the EBA is able to model a diversity of situations, having as special cases the MNL, the Nested Logit model, and the Cross Nested Logit model (Kohli and Jedidi, 2015, 2017; Aribarg et al., 2017), in this thesis we consider that heuristics derived from random utility theory as stand-alone heuristics differ from EBA.

In EBA, alternatives are completely described by aspects, defined as discrete desirable[2] characteristics that are directly mapped from the alternative's attributes. If the attribute associated to an aspect is discrete, then, a discrete set of aspects directly describes the attributes. However, if the attribute is continuous, there is no straightforward representation

---

[2] Tversky (1972b) postulates that an alternative model could consider an heuristic with undesirable attributes where regret is involved. However, this formulation is out of the scope of this thesis.

of the aspects. The most common representation considers a threshold that divides the attribute into acceptable (or desirable) and unacceptable (or undesirable). Unfortunately, the methods to estimate such threshold are scarce and are further addressed in Section 4.1.

In the EBA heuristic, DMs are supposed to choose an aspect from the set of available aspects and eliminate all alternatives that do not have the it. The process continues until only one alternative is available, which is then chosen.

The inspection order, or ranking of aspects, could be deterministically determined for the individual. However, to accommodate different decision profiles and uncertainty, the modeller assumes that DMs choose stochastically the inspected attribute. Therefore, the stochastic nature of the model lies in the stochasticity associated with the inspection order.

The aspects are the only elements involved in the decision process. The importance of each aspect is given by a positive continuous variable $w \in W$ named *weight,* that is positively correlated with the probability of choosing the aspect. The most common formulation of EBA selects one aspect with a probability proportional to the weight of the available aspects in the current choice set (2.11). To freely[3] estimate the weight (i.e. unconstrained in $\mathbb{R}$), the log-weight (α) is normally estimated as in (2.12).

---

[3] This is needed in maximum likelihood estimation since the estimators distribute asymptotic Normal. Additionally, it is desirable in Bayesian estimation in order to have a higher variety of priors available.

$$P_i = \frac{w_i}{\sum_{\forall j \in availableA} w_j} \tag{2.11}$$

$$\mathrm{w_k} = \exp(\alpha_k) \tag{2.12}$$

Different formulations of the aspects and the probability of selecting them generate different types of EBA models. The Hierarchical EBA heuristic, or PRETREE, is used to model the selection of aspects that are not independent (Tversky and Sattath, 1979). The Elimination by Dimensions heuristic (Gensch and Ghose, 1992) and Elimination by Cut-offs (Manrai and Sinha, 1989) try to tackle the problem of EBA's thresholds on continuous attributes. Even though the EBA can accommodate several behaviours, only the most common version –and the most popular one– will be used (2.11) here and will be identified unequivocally as EBA.

2.2.3. Random Regret Minimization (RRM)

Random Regret Minimization –RRM– (Chorus et al., 2008) is a heuristic where DMs value alternatives relatively. It is based on the concept of anticipated regret, that is, the feeling triggered when the individual imagines how would have been the situation if he had taken another choice (Simonson, 1992). Indeed, the relationship between the anticipated regret and choice is a well-studied problem in psychology (e.g. Zeelenberg, 1999; Zeelenberg and Pieters, 2007).

The first formulation of the RRM model (2.13) is a direct interpretation of the economic principle of minimising the maximum loss (Savage, 1951). If the error term is Gumbel distributed, the model adopts the logit model structure (2.14).

$$\begin{aligned} \min_i R_i &= \min_i \max_{i \neq j} R_{ij} + \epsilon_i \\ &= \min_i \max_{i \neq j} \left\{ \sum_m max\left(0, \beta_m\left(x_{jm} - x_{im}\right)\right) \right\} + \epsilon_i \end{aligned} \tag{2.13}$$

$$P_i = \frac{\exp\left(-\lambda R_i\right)}{\sum_{\forall j \in J} \exp\left(-\lambda R_j\right)} \tag{2.14}$$

Even though (2.13) exhibits a simple structure, model estimation is not straightforward due to the max operator. Then, typically an approximate function is used (Chorus, 2010), which is shown in (2.15). This is the most popular version of the RRM model, also known as RRlog (Jang et al., 2017).

$$R_i \approx \sum_{\forall j \in J \neq i} \sum_{\forall k \in K} \ln\left(1 + exp\left(\beta_m\left(x_{jm} - x_{im}\right)\right)\right) \tag{2.15}$$

Several other versions of the RRM model have been proposed, one of them, the μ-RRM (van Cranenburgh et al., 2015), enables to further exploit the flexibility of the model (2.16). The RRlog model is not scale-invariant; hence, the scale may be estimated. The μ-RRM exploits such feature and estimates the scale at the expense of additional parameters. Even though, several other random regret models have been proposed, such as the generalized-random regret model (Chorus, 2014) or the pure-random regret model (van Cranenburgh et al., 2015), they will not be further analysed since they do not represent a significant –or any–

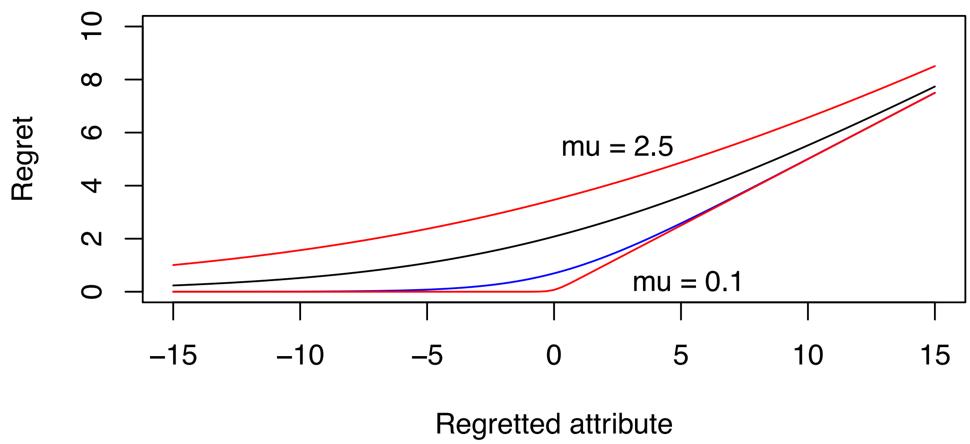
improvement over the μ-RRM model and only enable further understanding of the regret function.

$$R_i \approx \sum_{\forall j \in J \neq i} \sum_{\forall k \in K} \mu_m \ln\left(1 + exp\left(\frac{\beta_m}{\mu_m}\left(x_{jm} - x_{im}\right)\right)\right) \tag{2.16}$$

The μ-RRM has two variables per attribute-regret function which allows to change the degree of regret associated to each attribute. The μ parameter controls the shape of the regret and the β parameter controls its magnitude.

The behaviour of the μ parameter is shown in Figure 2-2. The blue line represents the traditional RRM model. The red lines, for which the shape parameter is indicated, represent the two most extreme cases where the μ parameter takes the higher and smaller values. For high values, the μ-RRM model exhibit a behaviour where the penalisation of the losses compared to the gains is not extreme (in the RUM model the penalisation is null). On the other hand, for small values the model exhibits a high profundity of regret (van Cranenburgh et al., 2015), where the penalisation of the losses are extreme and the valuation of the gains are non-existent. Finally, the black line denotes an intermediate behaviour between the higher value of the shape parameter and RRM.

Figure 2-2 Regret function for different scale parameters



The RRM has become popular in several domains. Applications are mainly found in transport, dealing with route choice (Leong and Hensher, 2012a; Chorus and Bierlaire, 2013), vehicle purchase (Beck et al., 2013), searching of parking space (van Cranenburgh et al., 2015), manoeuvre in car accidents (Kaplan and Prato, 2012), and transport mode (Chorus, 2010). Other applications can be found in health economics (Boeri et al., 2013) and marketing, to analyse online dating (Chorus and Rose, 2012).

2.2.4. Lexicographic behaviour (LEX)

A Lexicographic choice heuristic privileges a single attribute over the rest. Only when the differences over the sought attribute are not noticeable, another attribute may be considered (Tversky, 1969). It has been argued that complex choices could trigger this choice heuristic (Ortuzar and Willumsen, 2011), and it has even been reported in simple decision tasks (Tversky et al., 1988). This choice heuristic has the particularity that no utility function can accommodate it (Debreu, 1954) and in some specific context it can outperform compensatory structures (e.g. Jedidi and Kohli, 2008).

The LEX choice heuristic is not intended to model all choice situations, thus, applications are not extensive. Moreover, distinguishing lexicographic behaviour from compensatory structures is not straightforward, since the response pattern of LEX could also be formed by compensatory responses (Saelensminde, 2006).

It can be argued that a lexicographic behaviour is a reduced version of an EBA heuristic. The EBA heuristic categorises alternatives into desirable aspects, therefore, if the only sought aspect is *if the alternative has the best value of the analysed attribute*, then, EBA is consistent with lexicographic behaviour. Hence, LEX may be interpreted as an extreme version of the EBA heuristic. Thus, only EBA will be used for experimentation in this thesis.

2.2.5. Models for dealing with Prospect Theory behaviour

Prospect theory departs from the classic rational economical behaviour by anchoring in three key elements: (i) preferences are context-dependent, (ii) losses loom larger than gains, and (iii) DMs may perceive biased probabilities. The theory assumes that although DMs are utility maximisers, they perceive attributes subjectively.

In sections 1.4 and 2.2 we have defined the limits of a choice heuristic. Prospect theory heuristics are in the blur limit between being contained in the RUM paradigm and being stand-alone. Although we consider that DMs following prospect theory behaviour are still utility maximisers, we consider this separately given its importance.

There are a family of models that try to capture the principles of Prospect theory (Kahneman and Tversky, 1979). Among the concepts applied, we describe the ones that focus on loss aversion (i.e. gains are valued different than losses). Gains and losses are compared against a reference; different reference points and utility functions gives birth to different models. Most of these models keep the RUM structure; indeed, they could be interpreted as RUM models with a different utility function. However, they are analysed as a different choice heuristic since they value alternatives relative to a reference point.

2.2.5.1. Loss Aversion (LA)

The loss aversion heuristic (Tversky and Kahneman, 1991) values gains differently than losses compared to the *statuo quo*. The utility function is piece-wise defined linear and additive (2.17):

$$V_i = \beta_0 + \sum_m \begin{cases} \beta_m(x_{im} - \overline{x_{jm}}), & x_{im} \geq x_{jm} \\ (\beta_m + \lambda_m)(x_{im} - \overline{x_{jm}}), & x_{im} < x_{jm} \end{cases} \tag{2.17}$$

In (2.17), one branch represents the gain domain and the other the loss domain. If the attribute is desirable, the first branch represents gains with respect to the *status quo* valued at a β_m rate; whereas, the other branch, represents the loss domain that is penalised higher through the loss aversion parameter (λ_m). If the attribute is undesirable the relationship is inverse.

In the LA model, the relationship between the RUM heuristic and Prospect Theory is evident: the latter provides further insights about the way that DMs value attributes, but do not challenge the principle of utility maximisation.

2.2.5.2. Contextual Concavity (CC)

The CC heuristic (Kivetz et al., 2004) implements the concept of context dependence and loss aversion. It values each attribute against the best one available in the choice set (2.18).

$$U_i = \beta_0 + \sum_m \left(\beta_k \left(x_{im} - \max_j x_{jm}\right)\right)^{\phi_m} + \epsilon_i \tag{2.18}$$

In (2.18) the ϕ parameter generates a concavity or convexity in the utility function depending of its value. When the model is concave, i.e. $\phi \in (0,1)$, losses are valued more than gains. The *max* operator induces context dependence and positions the concavity point. Again, as in the LA model, the relationship between RUM and Prospect Theory is evident.

2.2.6. Models of relative evaluation

The objective of the following heuristics is to evaluate alternatives relatively. DMs build a representation of the alternative based on the alternative performance compared to other alternatives. The individual evaluates the ratio of gains and losses of the various alternatives considered. Among several formulations that model this feature, there are two that have been recently tested in transport research: the Relative Advantage Maximisation model and the Majority of Confirming Dimensions model.

2.2.6.1. Relative Advantage Maximisation (RAM)

RAM (Tversky and Simonson, 1993) attempts to value context independent and context dependent features into a single unit of merit (2.19). Context independent features are modelled as a RUM-like structures (V_i), whereas context dependent features are modelled as a relative measure of advantage (A_m) and disadvantages (D_m) over other options (2.20).

$$U_i = \beta_v V_i + \beta_r R_{iJ} + \epsilon_i \tag{2.19}$$

$$R_{iJ} = \sum_{\forall j \neq i} R_{ij} = \sum_{\forall j \neq i} \sum_{m} \frac{A_m(i,j)}{A_m(i,j) + D_m(i,j)} \tag{2.20}$$

Equation (2.20) indicates how the relative performance of an alternative is linked to relative advantages and disadvantages. Tversky and Simonson (1993) suggests defining advantages and disadvantages asymmetrically, so that disadvantages are more valued than advantages. Even though this definition aligns with Prospect Theory (Kahneman and Tversky, 1979), no statistical evidence supports such proposal for this particular model yet (Kivetz et al., 2004).

Finally, Equation (2.21) indicates the relationship between the attributes of the compared alternatives and the corresponding advantage:

$$A_m(i,j) = \begin{cases} \beta_k(x_{im} - x_{jm}), & \beta_m(x_{im} - x_{jm}) > \tau^m \\ 0, & other\ case \end{cases} \tag{2.21}$$

Specifically, β_k values a desirable difference of an attribute of the selected alternative (i) compared to another one (j). The variable τ corresponds to a minimum perception threshold.

Yet, the application of a minimum perception threshold is only theoretical in this model since it has not been tested. Nonetheless, the minimum perception threshold has been applied in other models (e.g. Cantillo and Ortúzar, 2005).

2.2.6.2. Majority of Confirming Dimensions (MCD)

Similarly to RAM, the MCD (Russo and Dosher, 1983) is a choice heuristic that compares alternatives relatively. Its main objective is to reduce the cognitive burden involved in a decision. In MCD, individuals compare alternative attributes and choose the option with the biggest number of *winning attributes*. It has been implemented as a complement to RUM, within the same utility function, as an additional variable (Hensher and Collins, 2011) or as a standalone heuristic (Leong and Hensher, 2012b).

2.2.7. Hard two-step (HTS) heuristics

HTS heuristics are composed of two stages. First the choice is reduced by strictly (hard) applying some criteria and then, over this reduced choice set, a total evaluation heuristic is used for choice. Even though these heuristics do not restrict the type of total evaluation heuristic used, only RUM has been applied.

In HTS heuristics, the screening criterion is hard because if the sought aspect is not found, the alternative is immediately discarded. For binary variables, the criterion is straightforward: the sought aspect is either present or not. Whereas for continuous variables, a threshold needs to be estimated or imposed –like in the EBA model. Since these heuristics are hard, no tolerance over the thresholds is accepted. The stochastic nature of the model lies in one of two ways: estimation of such thresholds or, over imposed thresholds, estimation of the probability of usage of each threshold.

The basic HTS heuristics are the Disjunctive (Gilbride and Allenby, 2004) and Conjunctive (Jedidi and Kohli, 2005) rules. The difference between them is the way they deal with multiple criteria. For the Disjunctive rule, an alternative is acceptable if one of the criteria is met, whilst for the Conjunctive rule, an alternative is acceptable if all criteria are met. The Conjunctive heuristic is equivalent to a two-step EBA-RUM (Gilbride and Allenby, 2006). Mathematically, let A_i be 1 if the alternative is acceptable and 0 otherwise; let a_{ik} be 1 if alternative i satisfies the condition k and 0 otherwise. Then, the Disjunctive and Conjunctive rules are given by (2.22) and (2.23) respectively.

$$A_i = \begin{cases} 1, & \sum_{\forall k \in K} a_{ik} \geq 1 \\ 0, & otherwise \end{cases} \tag{2.22}$$

$$A_i = \begin{cases} 1, & \prod_{\forall k \in K} a_{ik} = 1 \\ 0, & otherwise \end{cases} \tag{2.23}$$

A generalisation of the Disjunctive and Conjunctive heuristics is the Subset Conjunctive heuristic (Jedidi and Kohli, 2005). This considers an alternative acceptable if at least n out of K conditions are met (2.24). The optimal number n may be estimated through maximum likelihood simultaneously with the rest of the model. This heuristic is equivalent to the Disjunctive rule when $n = 1$ and to the Conjunctive rule when $n = K$.

$$A_i = \begin{cases} 1, & \sum_{\forall k \in K} a_{ik} \geq n \\ 0, & otherwise \end{cases} \tag{2.24}$$

A model that further analyses the Conjunctive heuristic is the Economic Screening Rule (Gilbride and Allenby, 2006). This choice heuristic considers that if an alternative is not inspected, then both the expected maximum utility and the cognitive burden decreases. This choice heuristic analyses which attributes are screened by the Conjunctive rule by linking them to the expected maximum utility loss that could be experienced if some alternatives are not further analysed. Therefore, this model considers that if an attribute is inspected, a cognitive effort is involved; if not, the cognitive burden decreases, but so does the expected maximum utility.

Let $\mathbb{E}(\mathrm{Max}(U)_I)$ be the expected maximum utility of the choice set I –calculated as the *logsum*[4]. Let the set $I - \{k\}$ be the choice set remaining after screening alternatives with a certain criterion in the k[th] attribute. Then, the maximum utility of this reduced choice set is given by $\mathbb{E}\left(Max(U)_{I-\{k\}}\right)$ and attribute k is screened if condition (2.25) is true; this indicates that an attribute is screened if the expected utility loss (left hand expression) is smaller than the willingness to lose utility that each individual q has (γ_q).

$$\mathbb{E}(Max(U)_I) - \mathbb{E}\left(Max(U)_{I-\{k\}}\right) < \gamma_q \tag{2.25}$$

---

[4] In a RUM model, the expected maximum utility with Gumbel errors –$\mathbb{E}(Max(U_i))$– is given by $\log(\sum_{i \in I} exp(U_i))$, popularly known as logsum.

The model then applies a RUM heuristic to choose among all acceptable alternatives. The objective is to estimate simultaneously the RUM's parameters and the distribution of γ_q in the population.

2.2.8. Soft two-step (STS) heuristics

In a STS heuristic DMs first reduce the size of the choice set by screening alternatives probabilistically and then choose using a total evaluation heuristic. In these heuristics, the choice set for each person is diffuse or probabilistic. These models are based on the work of Manski (1977), where the probability (P_{iq}) that each individual q assigns to alternative i is the total probability over the conditional choice sets $c \in C$, as in:

$$P_{iq} = \sum_{\forall c \in C} P_{iqc} P_c \tag{2.26}$$

Even though any combination of screening heuristic and choice heuristic could be used, only conjunctive-RUM heuristics have been applied. One example is the Semi-Compensatory heuristic (Cantillo and Ortúzar, 2005). In the first step, alternatives are acceptable if all the sought conditions are met –like in the Conjunctive heuristic–. These conditions are modelled as a probability of acceptance of the analysed attributes around estimated thresholds. Then, in the second step, a RUM heuristic is applied over the remaining choice set.

Given that the number of possible choice sets could be extensive, different approximations of (2.26) have been proposed (Cascetta and Papola, 2001; Martínez et al., 2009). The most popular approximation, the Constrained multinomial logit (Martínez et al., 2009), has been used with exogenous and endogenous thresholds (Castro et al., 2013). However, Bierlaire et al. (2010) have shown it to be a poor first order approximation of (2.26). Notwithstanding, higher order approximations –fifth order shows to be good enough– have shown to behave well (Paleti, 2015).

2.2.9. Satisficing

Satisficing has been commonly defined as *choosing the first satisfactory alternative*. Even though the notion of Satisficing was clearly expressed in the seminal work of Simon (1955), few researches have been able to wholly implement its principles.

Among the principles of Satisficing stands the consideration of partial pay-off functions. Following Satisficing, individuals are supposed to have problems in mixing attributes of a different nature into a single unit of merit; however, this feature is frequently ignored in most applications (e.g. Araña et al., 2008; Radner, 1975; Richardson, 1982).

Recently, a model in the marketing literature was proposed which fully applies the Satisficing principles (Stüttgen et al., 2012). Nevertheless, it requires eye-tracking data for estimation, which is yet not possible in most choice settings. So, a model that wholly applies the Satisficing principles is not available in practice. Therefore, to fill this gap in the literature, we propose a model that applies the three main Satisficing principles and explain it in Section 4.2.

2.3. Techniques for Model Selection

When several models are available a criterion may be needed to choose the best among them. Several techniques allow to characterise different properties of the models and their performance. This way, information is provided to help select the preferred model. Among the several techniques for characterising model performance, we describe here the use of hypothesis tests, information criteria and out of sample validation.

2.3.1. Hypothesis tests

The statistical tests reported in this subchapter deliver indicators based on the model using exclusively the estimation sample. We describe two tests that may indicate the preference of one model over another non-nested model –as those to be compared in this thesis; however, they are restricted to maximum likelihood estimation contexts.

2.3.1.1. Likelihood ratio test

This test has several formulations under different conditions. There are some versions that are only valid for multinomial logit models by making use of the *independent of irrelevant alternatives properties* (McFadden et al., 1977; Horowitz, 1982). The most popular formulation (2.26) is only valid if the general model is well specified. Expression (2.27) relates the likelihood of a general model and the likelihood of a restricted version of the general model.

$$LR = -2\left(l(\theta_{restricted}) - l(\theta_{general})\right) \sim \chi_r^2$$
$$with\ r = number\ of\ restrictions \quad (2.27)$$

Even though version (2.27) of the test can be useful, in the literature it is widely reported that when the model is not well specified –a common case in models with multiple heuristics, the LR statistic does not necessary distribute chi-squared (Foutz and Srivastava, 1977; Kent, 1982; White, 1982). A general version of the LR statistic, less popular however, enables its use under non-nested models (i.e. one is not a restricted version of the other) and under misspecification (Vuong, 1989). Let the two compared models be f and g; then (2.28) defines the terms under which the estimator is defined.

$$H_0\text{: } both\ models\ are\ equivalent$$
$$\omega^2 = \frac{1}{n}\sum_{t=1}^{n}\left[\log\left(\frac{p_f(Y_t|\theta_1)}{p_g(Y_t|\theta_2)}\right)\right]^2 \quad (2.28)$$

Expression (2.29) states how the estimator distributes. With the expression of H_0 a t-test may be performed that indicates the equivalence of the two models.

$$H_0\text{: } \frac{LR}{\omega\sqrt{n}} \sim N(0,1)$$
$$Under\ H_f\text{: } \frac{LR}{\omega\sqrt{n}} \to \infty \quad (2.29)$$
$$Under\ H_g\text{: } \frac{LR}{\omega\sqrt{n}} \to -\infty$$

2.3.1.2. Other tests for non-nested models

Other tests have been proposed to compare non-nested models by simultaneously estimating both models under analysis. The versions of Cox (1962) and Davidson and MacKinnon (1981), stated by (2.30) and (2.31) respectively, enable to compare two models by means of a t-test.

$$y = \alpha f(x) + \beta g(x) \tag{2.30}$$

$$y = (1 - \alpha) f(x) + \ \alpha f(x) \tag{2.31}$$

Even though the test –and the idea behind it–can be simple, it is difficult to apply it in our context. Multiple heuristic models are difficult to estimate and the risk of non-identifiability is high; therefore, the simultaneous estimation of two multiple heuristic models is implausible.

2.3.2. Information criteria

Information criteria deals with a phenomenon known as *bias-variance trade-off* (Hastie et al., 2001; McElreath, 2012). This relates the fit of the model (bias) with its performance in out of sample validation as the number of parameters increases (variance). From it, we can conclude that although sophisticated models may reduce the bias significantly, their predictive performance may be compromised; when this happens, the model is said to be *overfitted*.

Several information criteria try to approximate the additional forecasting bias produced by adding additional parameters. Most information criteria are based on the concept of *divergence*, that is the additional uncertainty induced by using a probability distribution to approximate another one (McElreath, 2012). The best known measure to quantify the divergence of a model is the Kullback-Leibler divergence –KL (Kullback and Leibler, 1951).

Approximating one distribution through another, produces an error measured in entropy units quantified by the divergence. The KL divergence quantifies the divergence as the difference in cross-entropy of the distribution used and the entropy of the distribution to approximate (2.32). Let $H(f)$ be the entropy of a distribution and $H(f,g)$ the cross entropy of two distributions. Random variable y_q comes from a f distribution. If a distribution g – or model– is used to approximate distribution f, then the KL divergence is defined as (2.32) and expands to (2.33) acquiring the form of an expectation. Note that the minimum value of the divergence KL is zero:

$$Divergence_{KL}(f,g) = H(f,g) - H(f) \tag{2.32}$$

$$Divergence_{KL}(f,g) = \sum_{\forall q \in Q} p_f(y_q) \cdot \left(\log\left(p_f(y_q)\right) - \log\left(p_g(y_q)\right)\right) \tag{2.33}$$

Akaike (1973, 1974) proves that if a researcher uses a KL loss function to compare two models, then the divergence need not be completely computed, but rather the difference among the deviances of the two models. Akaike proves that the difference of two KL Divergences is the difference of the loglikelihood of the models. To calculate the difference, let the deviance of a model be given by (2.34). Then, if two models g and e try to represent a generating process f, the difference of the two divergences, which is equivalent to the difference of the deviances, is stated by (2.35).

$$Deviance = -2 \sum_{\forall q \in Q} \log\left(p_g(y_q)\right) \tag{2.34}$$

$$\Delta Deviance = -2 \sum_{\forall q \in Q} \log\left(p_g(y_q)\right) - \log\left(p_e(y_q)\right) \tag{2.35}$$

The out of sample deviance of two models is a measure of model underfit/overfit. To calculate it, the deviance needs to be computed piecewise. However, to avoid using separate samples or to have an immediate indicator after model fitting, most information criteria try to approximate the out of sample deviance by using information of the estimation sample and the model structure.

The Akaike information criterion (AIC) tries to approximate the corresponding piece of the out of sample deviance of a model (Akaike, 1973). This estimator uses the log-likelihood of a model and penalizes it by the number of parameters k (2.36). This estimator is valid for both Bayesian estimated models and maximum likelihood estimated models.

$$AIC = -2 \sum_{\forall q \in Q} \log\left(p(y_q)\right) + 2k \tag{2.36}$$

AIC approximates correctly the deviance under several conditions. First, it requires the sample to be large. Second, it requires that the prior distribution of the parameters is relatively flat, which is always true in maximum likelihood estimation, but not necessarily under a Bayesian framework. And finally, it requires that the posterior distribution of the parameters is Normal. Each of these restrictions have been approached through other information criterion.

When the sample is not big enough, the AIC may be biased. The Corrected Akaike Information Criterion –CAIC (Hurvich and Tsai, 1989; Anderson et al., 1998) provides a correction for this:

$$CAIC = AIC + \frac{2(k+1)(k+2)}{|Q| - k - 2} \tag{2.37}$$

When the prior distribution of the parameters is not relatively flat, the Deviance Information Criterion –DIC (Lunn et al., 2012) is used. The DIC (2.38) changes the parameter penalty in the AIC formulation (2.36) considering the number of effective parameters.

$$DIC = -2\sum_{\forall q \in Q} \log\left(p(y_q)\right) + 2k_{eff\ DIC} \tag{2.38}$$

The number of effective parameters does not have a precise definition but can be estimated as in (2.39) and (2.40). Both estimators arise from a χ^2 distribution and are valid asymptotically (Gelman et al., 2013). To estimate the DIC, several Markov chains are run over the same dataset and stability is analysed over the different chains; here S is the set of samples from each chain.

$$k_{eff\ DIC} = \sum_{\forall q \in Q} \left(\log\left(p(y_q)\right) - \frac{1}{S} \sum_{\forall s \in S} \log\left(p(y_q|\theta_s)\right) \right) \tag{2.39}$$

$$k_{eff\,DIC} = \sum_{\forall q \in Q} Var\left(\log\left(p(y_q)\right)\right) \tag{2.40}$$

If the posterior distribution do not distribute normal, then the Widely Applicable Information Criterion or Watanabe Akaike Information Criterion –WAIC– (Watanabe, 2010) is used. WAIC modifies the DIC by adding further information about the different Markov chains as in (2.41) and (2.42). Note that if the posterior distribution is symmetrical in (2.41), the estimator for DIC and WAIC is the same.

$$k_{eff\,WAIC} = \sum_{\forall q \in Q} \left(\log\left(\frac{1}{S} \sum_{\forall s \in S} p(y_q|\theta_s) \right) - \frac{1}{S} \sum_{\forall s \in S} \log\left(p(y_q|\theta_s)\right) \right) \tag{2.41}$$

$$k_{eff\,WAIC} = \sum_{\forall q \in Q} Var\left(\log\left(p(y_q)\right)\right) \tag{2.42}$$

Even though DIC and WAIC have two estimators for the number of effective parameters, the second formulation –(2.40) and (2.42)– has shown a better performance in replicating the out of sample deviance (Gelman et al., 2013).

Another information criterion is the Bayesian Information Criterion –BIC (Hastie et al., 2001), which is based in the Schwarz criterion (Schwarz, 1978). This criterion is also valid for Bayesian and maximum likelihood estimated models. Unlike other estimators that try to assess the forecasting performance of a model, BIC tries to establish the posterior probability of a model over other models (akin to Equations 2.6 and 2.7). Equation (2.43) shows the BIC; as can be easily inspected, it penalises harder the likelihood of the model than the AIC as the sample grows.

$$BIC = -2 \sum_{\forall q \in Q} \log\left(p(y_q)\right) + \log(|Q|) \cdot k \tag{2.43}$$

Finally, the Minimum Description Length (Wallace, 2005; Grünwald, 2007) promotes more parsimonious models. This measure finds its origin in the theory of coding for data compression. Even though its origin differs from other information criteria, this approach gives a selection criterion formally identical to BIC (Hastie et al., 2001).

There are other information criteria, e.g. the Vapnik-Chervonenkis Dimension (Hastie et al., 2001), that are not addressed here. However, as these techniques do not have the popularity of AIC and BIC are less validated. Because the objective of this thesis is not an exhaustive analysis of different information criteria, their analysis is out of the scope of this thesis.

2.3.3. Out of sample validation

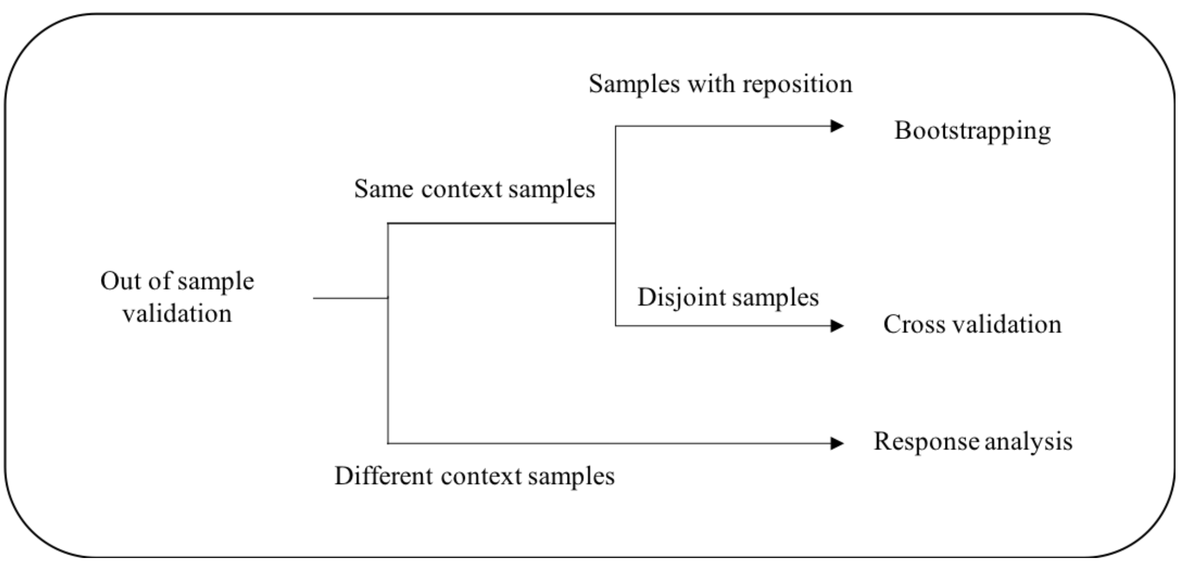
Once several satisfactory models are available (satisfactory in terms of its statistical properties and theoretical validity), we may desire to categorise them in terms of their predictive performance. Out of sample validation is one of the best tools to analyse the performance of a model as it allows to directly measure any desired property (like predictive performance). Indeed, information criteria try to approximate these techniques. If plenty of data is available, the predictive performance can be directly measured though out of sample validation and no approximation would be needed.

Several techniques use out of sample validation to measure the forecasting performance and degree of overfit of a model. We categorise the different techniques in two dimensions depending of the data involved in the validation and the statistical estimator calculated.

2.3.3.1. Type of data in out of sample validation

The first dimension considers the type of data used for the analysis. Validation data may come from the same type of experiment or context as the estimation data (which is the typical out of sample analysis) or it can come from another context; this is known as response analysis (Williams and Ortuzar, 1982a, 1982b). Moreover, when data comes from the same type of experiment, the whole sample may be used for estimation and validation, like in bootstrapping, or it may be used in disjoint fashion, like in cross-validation. This first dimension is summarised in Figure 2-3.

Figure 2-3 Classification of out of sample techniques



When data comes from the same context of the experiment, it has the advantage that no additional experiment must be performed for validation. The totality of the data may be used for estimation and validation, like in the case of bootstrapping (Efron and Tibshirani, 1997), or they may be disjoint sets, like in cross validation. In both techniques, the process may be repeated to obtain a mean and variance of the analysed feature.

Bootstrapping uses the same sample for estimation and validation. This technique first estimates the model with the totality of the sample. Then, the validation sample is obtained by sampling with reposition from the original dataset. This technique provides biased estimates –however consistent– of the measured properties. Nevertheless, for small samples, it provides smaller mean square errors due to the increase in the estimation and validation sample sizes.

Cross validation uses disjoint datasets for estimation and validation. This technique has the advantage of providing unbiased estimates. Depending on the proportion of sample used for estimation and validation, this technique is classified into two subgroups: K-fold cross validation and leave-one-out cross validation (Hastie et al., 2001). K-fold cross validation divides the sample into K groups; one group is used for validation and K-1 for estimation. Then, the group left out is alternated until every group has been left out once. Leave-one-out cross validation is the extreme case where the number of groups is the sample size. Larger estimation samples decrease the estimated bias, whereas, larger validation samples decrease the variance of the estimation. Studies present in the literature suggests that dividing the sample into five to ten parts offers a good compromise between bias and variance (Breiman and Spector, 1992; Kohavi, 1995).

Finally, response analysis or non-hold out sample uses validation data from a different context than the estimation data. It has the disadvantage of requiring another data source or additional experiments must be performed. The objective of response analysis is to understand how the model performs to a change in the choice context; e.g. in transportation it could be an increase in the fare of all modes. The underlying hypothesis is that a model that represents adequately the phenomenon should have a good performance if the context changes since it captures the underlying mechanisms (Keane and Wolpin, 2007). On the other hand, if a model is overfitted, it may perform adequately in the estimation sample, but poorly when the context changes. This technique can characterise the competing models deeper, however, it requires richer information.

2.3.3.2. Estimator calculated in out of sample techniques

Independent of the type of data used for analysis, a statistic must be calculated for the tested case. We will analyse three different statistics: the first preference recovery (FPR) or hit rate, the log-likelihood of the validation sample, and probability bands.

The FPR (Ortuzar and Willumsen, 2011, chap. 8) or hit rate corresponds to the proportion of the cases in which the model assigns the maximum probability to the chosen alternative. The expected value of the FPR, the expected recovery, is the average likelihood of the model. By using both estimates and the expected recovery of a null model it is possible to infer if the model is informative and reasonable. The FPR main advantage is being an intuitive estimator.

The average log-likelihood of the validation sample is a measure to express the performance of a model. It is based on the deviance of a model (2.34) that measure the difference between the model's performance and the real underlying model. The deviance has several useful properties that outperforms the FPR as a performance statistic. For example, it is obtained analytically from information theory. Furthermore, it penalises infinitely a deterministic model as a model for a stochastic process (for further information see McElreath, 2012). However, it could be a less intuitive statistic.

Probability bands is a technique that analyses in detail the performance of the model throughout the complete range of possible probabilities (Ortuzar and Willumsen, 2011, chap. 8). The analysis through probability bands compares the theoretical distribution of the model in the scope of the data range with the empirical distribution of the data. Then, the significance of this difference may be checked through a chi-squared test. This way, the competing models may be precisely analysed in the whole domain of probabilities.

2.3.4. Use of techniques of model selection in this thesis

As stated in Section 2.1, this thesis uses Bayesian and maximum likelihood estimation. To be able to compare Bayesian models, no hypothesis tests will be used to discriminate among competing models.

Information criteria may be used with maximum likelihood and Bayesian estimation. Among the information criteria exposed in subsection 2.3.2, the performance of DIC and BIC will be analysed regarding their capacity to choose between satisfactory models when multiple choice heuristics are involved. DIC is needed since some priors are not flat enough. Because of the large sample sizes, we will use no correction for the DIC. Our numerical simulation shows that for 15 parameters, a sample size of 1,000 observations is enough to exhibit a bias smaller than 0.5 log-likelihood points. Therefore, conventional DIC and BIC will be used.

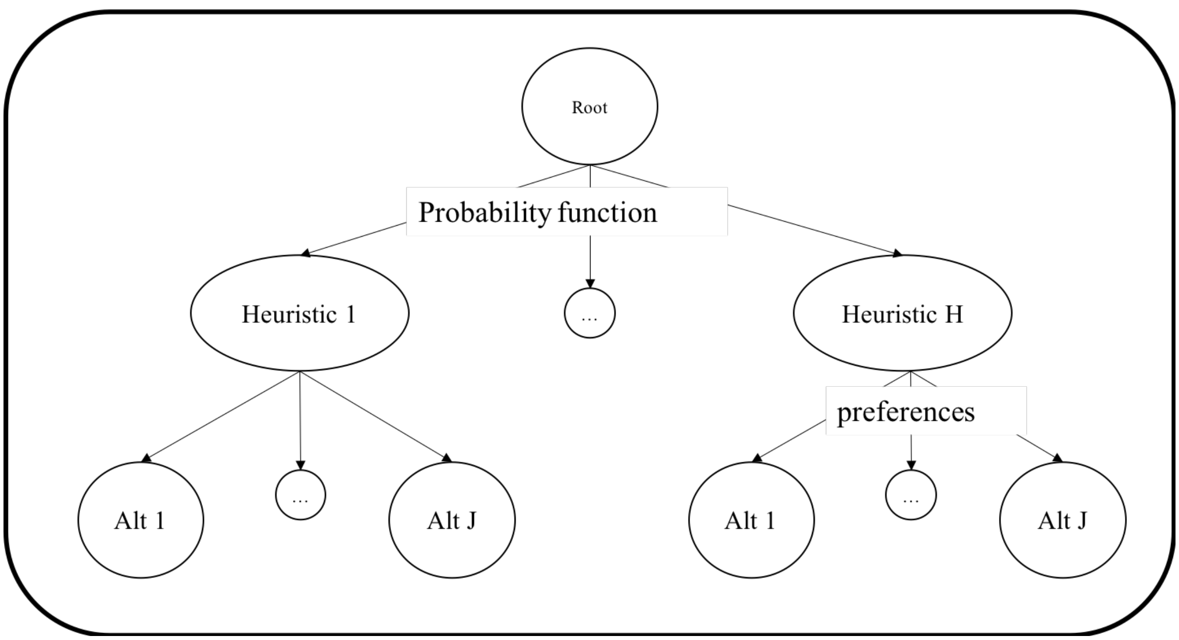
Finally, regarding out of sample techniques, this thesis considers the use of simulated data; therefore, the amount of data is not a restriction. This way, techniques like bootstrapping are not required. Indeed, cross validation is used with relatively large sample sizes. Moreover, in some cases, response analysis is used to further understand the model's properties. Regarding the estimator calculated, only out of sample log-likelihood is calculated, since it outperforms the FPR and the degree of detail that the probability bands provides is not needed.

2.4. Modelling Multiple Choice Heuristics

Models with multiple choice heuristics attempt to capture the behaviour of the population in which DMs choose using different heuristics. The main approach to model multiple choice heuristics is to use latent classes, where each class represents a heuristic. Despite other attempts to use flexible structures to represent more than one choice heuristic in a single formulation (Fiebig et al., 2010; van Cranenburgh et al., 2015), the use of latent classes has been the prevalent modelling technique. Therefore, we will refer to the multiple heuristic model unequivocally as the one using the latent class approach.

The logic behind latent classes is summarised in Figure 2-4. DMs choose every heuristic with a certain probability function –the class membership function. Then, once the heuristic is chosen, DMs choose one of the available alternatives with the corresponding heuristic. The model tries to estimate simultaneously the class membership function's parameters and each of the heuristics' parameters.

Figure 2-4 Structure of a latent class model for multiple heuristics



The probability assigned by the latent class model to each alternative i is given by (2.44). In it, the probability that individual q chooses an alternative is given by the total probability of choosing the alternative conditional on the choice heuristic h and the probability of choosing such heuristic $\pi(\theta)$.

$$P_{iq} = \sum_{\forall h \in H} P_{iqh}(\beta|h)\, \pi_h(\theta) \tag{2.44}$$

Some studies using multiple choice heuristics reported in the literature are summarised in Table 2-3. Analysing them is important to identify the issues concerning multiple choice heuristic models. As can be seen, most studies considering multiple choice heuristics use stated preference (SP) data. The sample size among studies vary, throughout the experiments we test with sample sizes ranging from 1,000 to 40,000 observations.

The main issue of multiple heuristics models is identifiability, which we will address in Chapters 5 and 6. The higher the number of heuristics used, the harder it is to identify the model. Furthermore, sophisticated class membership function that enables to characterise the decision process of individuals further complicate identifiability. That is the reason that from the experiments reported, all except two models use a constant class membership function, i.e. the probability of using a heuristic is the same across all individuals. The two models that estimate a "sophisticated" class membership function exhibit identifiability problems: McNair et al. (2012) had to normalise certain estimates, whereas Hess and Stathopoulos (2013) had to use latent variables. Therefore, identifiability is the main issue in multiple heuristic models.

Table 2-3 Studies reported in the literature using multiple choice heuristics

Study	Type of data	Individuals	Responses per individual	Total sample size	Simultaneous heuristics
Araña et al. (2008)	SP	225; 225	3	675; 675	4
McNair et al. (2012)	SP	290; 292	4	800; 872	3
Hess et al. (2012)	SP	1,676; 368; 996	8; 10; 8	13,408; 3,680; 7968	2
Leong and Hensher (2012)	SP	752	16	12,032	2
Hess and Stathopoulos (2013)	SP	368	10	3,680	2
Balbontin et al. (2017)[5]	SP	1,578	6	9,468	4
Adamowicz and Swait (2013)	RP	3,242; 262	5,4; 14,8	17,504; 3,885	3

Note: SP = stated preference; RP = revealed preference

Higher identifiability may be obtained if the choice heuristic uses different attributes. Indeed, latent class models using only RUM are frequent in the literature; these models are identified by using different attributes in the utility functions in each class. In this thesis we try to

[5] They consider attribute non-attendance as a stand-alone heuristic. We interpret attribute non-attendance as a RUM heuristic where an attribute is ignored.

reduce this phenomenon to the minimum, so that choice heuristics may be identified solely due to their difference in interpreting attributes or behavioural difference. This way, we guarantee no confounding between heuristic preference and sensitivity difference.

3. DATASETS USED IN THIS THESIS

This chapter addresses issues concerning the two datasets used in this thesis. Neither was collected within the specific context of this thesis; however, they provide useful contexts to apply some of the models.

3.1. Las Condes - CBD, San Miguel - CBD dataset

This dataset comes from a revealed preference survey and is used intensively throughout the simulation experiments. First, we explain the dataset and its general characteristics. Then, we explain how we obtained fictitious choice sets from it and how we manipulated its characteristics according to the design of the experiment.

3.1.1. The dataset

The "Las Condes - CBD, San Miguel CBD" dataset is a revealed preference transport mode choice dataset collected by Donoso and Ortuzar (1982). The "Las Condes - CBD" corridor survey was gathered in 1981 and the "San Miguel - CBD" corridor survey was gathered in 1983 (Ortuzar et al., 1983). This dataset has the feature of being well-tested as it has been used in several transport studies throughout the years (Ortúzar and Fernández, 1985; Ortúzar and Espinosa, 1986; Ortuzar and Ivelic, 1987; Gaudry et al., 1989; Jara-Díaz and Ortúzar, 1989; Guevara, 2016; Guevara et al., 2016). Its popularity is probably due to the extreme quality of measurement of all the alternative attributes

The dataset considers journey-to-work trips made by 1,374 DMs, with 697 of them living in the “Las Condes - CBD” corridor in Santiago de Chile and 677 of them living in “San Miguel - CBD”. DMs had between two and nine modes available which where endowed with the following attributes: cost, travel time, walking time and waiting time. Considering each available mode as an observation, their means and variances (expressed as the coefficient of variation for ease of inspection) are given in Table 3-1.

Table 3-1 Mean and coefficient of variation of the alternative attributes

Attribute	**Mean**	**Coefficient of variation**
Cost (CLP$[1])	45.4	0.72
Travel time (min)	16.9	0.43
Walking time (min)	6.6	0.62
Waiting time (min)	1.5	0.97

[1]In 1983, 1 US dollar was worth CLP$ 80.

The “Las Condes - CBD, San Miguel CBD” dataset is extensively used throughout the simulated experiments of this thesis with the objective of providing realistic choice set in which fictitious DMs can choose. Two dimensions are controlled throughout the experiments: the sample size and the number of alternatives. Both are explained below.

3.1.2. Creating fictitious choice contexts

The objective of using a real dataset in a simulated experiment is to provide realistic choice sets, while being able to control the conditions of the experiment. In this thesis two conditions regarding the datasets are manipulated: the sample size and the number of alternatives.

3.1.2.1. Sampling process

Through the sampling process we obtain datasets of the desired sample size. To create a synthetic choice set databank which is as realistic as possible, random observations (i.e. choice sets) are obtained from the real dataset. The objective is to preserve the correlation of each alternative's attributes and the correlation of the attributes between the different alternatives. To achieve it, rather than sampling individual alternatives, a whole random choice set is sampled from the real dataset. The sampling process must be done with replacement, since the original dataset size does not allow to obtain several synthetic datasets of the desired sizes without replacement.

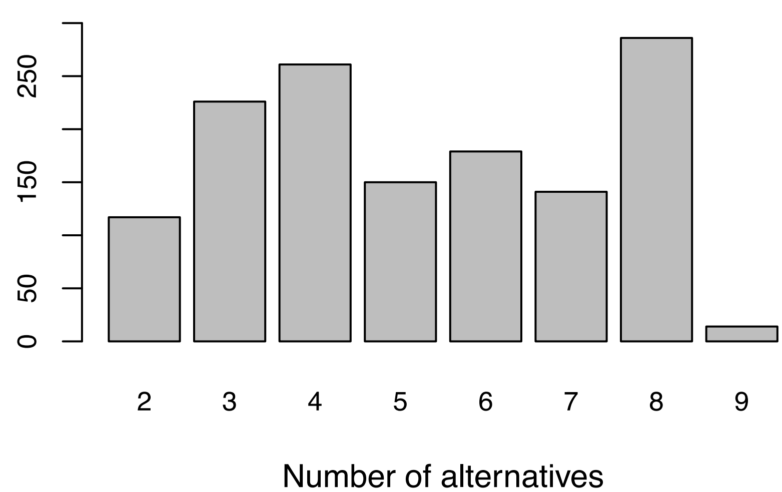
3.1.2.2. Adjusting the number of alternatives

The method described previously provides a realistic scenario for DMs to choose from. However, in some experiments we required that the dataset had some desired characteristics such as a specific size of the choice set. In this thesis, choice sets of size three and seven were used for simulation.

To have a choice set with the desired characteristic, two options may be followed: sample from the observations that meet the characteristics or adjust the sampled observation so that they meet the characteristic. The former solution has the problem of reducing the available sample size; whereas, the latter need further processing.

In the experiments, we wish to control for the size of the choice sets. Then, the total sample size of 1,374 observations may be categorised in relation to its choice set size as in Figure 3-1. We used sample sizes of up to 40,000 observations. Therefore, the choice sets of a specific sample size might not provide the necessary variety of choice sets for the estimation to be successful.

Figure 3-1 Choice set size distribution



The problem of the variety of alternatives could be tackled by modifying (and expanding) large choice sets (i.e. the second approach described before). For example, suppose there are choice sets with two to nine available alternatives and we require to have choice sets with eight alternatives; then every choice set with eight alternatives is useful plus every choice set with nine alternatives only by removing one alternative. Additionally, each nine alternative choice set can be expanded nine times by eliminating a different alternative in each of them.

Sampling was performed as follows. Let C_i be the choice sets of size i, $|C_i|$ its cardinality, and c_i an element of C_i. Let l be the minimum sample size or lower bound of acceptable choice set sizes and u the maximum. We would like to give every choice set the same probability of being chosen before being adjusted; then, the probability of sampling choice set c_i before being adjusted is given by (3.1).

$$P(c_i) = \frac{1}{|C_i|} \cdot \frac{|C_i|}{\sum_u^m |C_i|} \tag{3.1}$$

Once a choice set is sampled, alternatives are randomly removed until the desired choice set size is obtained. This way, there are a large variety of different choice sets available to be sampled. We used two choice set sizes, three alternative choice set with 28,477 available and size seven choice sets with 2,933 available.

3.2. Singapore Air Travel

The “Singapore air travel” dataset comes from a SP experiment conducted by Sebastián Raveau in 2015 and was answered by 52 MIT Singapore workers yielding 1,248 choices.

The data consists of travel itineraries from Singapore to 24 different destinations in Asia-Pacific and the Middle East (Figure 3-2). Trips were divided into 12 short-haul trips and 12 long-haul trips. All alternative itineraries correspond to real flight options for given travel dates: all short-haul itineraries correspond to a weekend-long trip from Thursday November 26, 2015 to Sunday November 29, 2015; while all long-haul itineraries correspond to a week-long trip from Saturday November 21, 2015, to Sunday November 29, 2015. DMs evaluated the alternatives in the context of a leisure trip considering only economy class seats.

Figure 3-2 Trips considered in the Singapore air travel survey

a) Short-haul trips

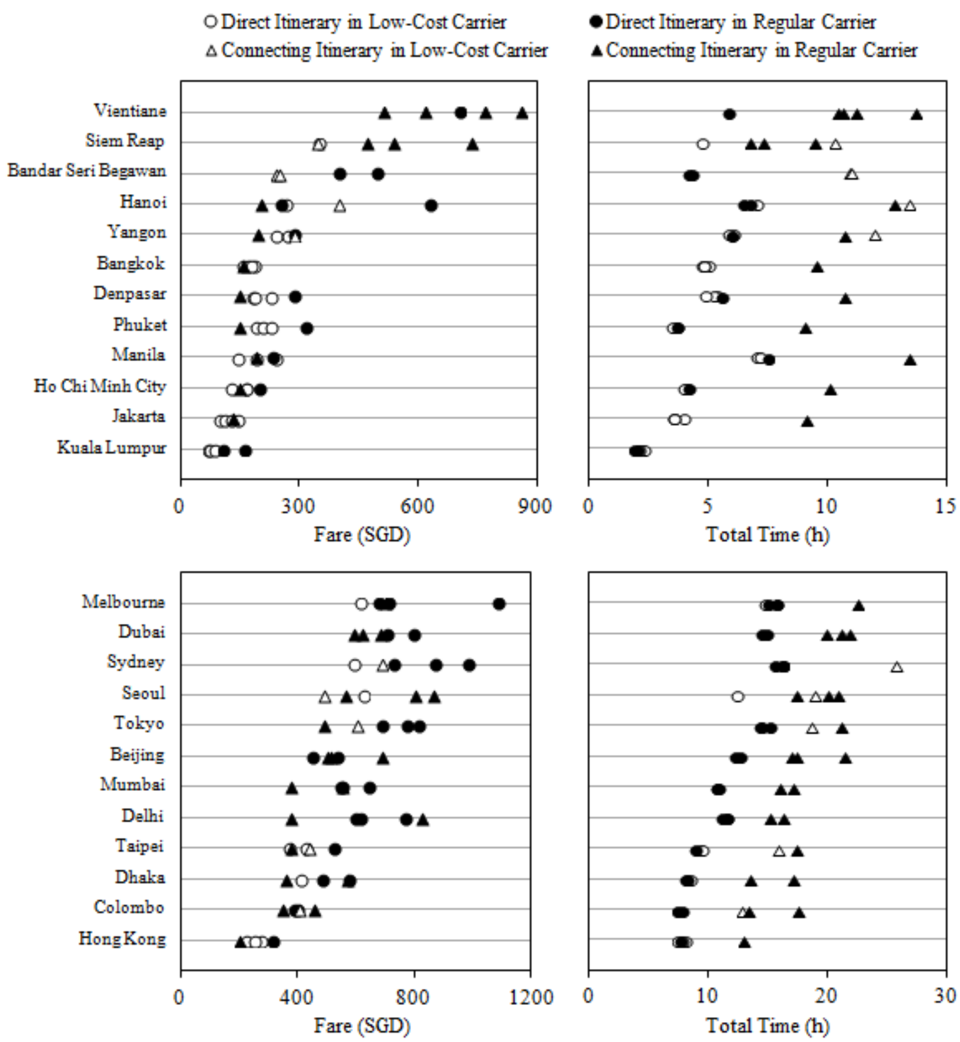
b) Long-haul trips

For 23 of the 24 choice scenarios, respondents had to choose between five real itineraries. The only exception was Bandar Seri Begawan (BWN), where only four reasonable itineraries existed (i.e. without excessively long detours and layovers). All respondents had to choose among the same alternatives. In total, 35 different airlines were part of the 119 itineraries, with an expectable bigger representation of carriers from South-East Asia. Of the 35 airlines, eight were low-cost carriers.

Each alternative itinerary was characterized according to six attributes: fare, total time, number of connections, connecting time (if any), if the airline was a low-cost carrier, and if the itinerary required to be at the airport at an "inconvenient" time (earlier than 9:00am and/or later than 9:00pm).

The aggregate attributes of the alternative itineraries for the 24 trips are presented in Figure 3-3. There is a strong dependency between connections and travel time, whereas there is no clear relationship between connections and fare. The relationship between fare and low-cost carrier is not as expected for several destinations, since even in the presence of low-cost alternatives, the cheapest itinerary for eight destinations corresponds to a regular-carrier.

Figure 3-3 Attributes of Singapore air travel survey alternatives



4. FURTHER DEVELOPMENTS OF CHOICE HEURISTICS

This chapter presents contributions related with two of the key choice heuristics examined in this thesis. First, an analysis of the estimation of the EBA model is performed, and second, the formulation of a model that wholly represents Satisficing theory is proposed.

4.1. Estimation of the Elimination by Aspects Model

The EBA model, as described in subsection 2.2.2, sequentially selects aspects and eliminates all alternatives not having the sought aspect. Each aspect has a certain weight and the probability of selecting it is proportional to this weight (2.10). Finally, the model estimates each aspect's weight.

The estimation of the EBA model involves two main issues: the generation of the alternatives' probabilities and the categorization of continuous attributes into aspects. Two contributions designed to decrease the importance of these issues are proposed in this thesis. First, an analytical approach to estimate the EBA model is proposed in subsection 4.1.1. This approach is discussed and compared with the current approach. Then, in subsection 4.1.2 we show that it is possible to estimate the thresholds and weights simultaneously under maximum likelihood and provide an example of it.

4.1.1. Estimation of weights in the EBA model

The estimation process of the EBA model is not straightforward since no general formula can represent the process heuristic. Even in simple configurations, the probability is hard to obtain analytically as the following example shows.

Consider three alternatives $A, B, and\ C$ represented by six aspects with weights w_i as shown in Table 4-1. Aspects one to three are shared between alternatives, whereas aspects four to six are exclusives to one alternative.

Table 4-1 Aspects defining alternatives on example of the EBA formula

Alternative	Aspects
A	$\{w_1, w_2, w_4\}$
B	$\{w_2, w_3, w_5\}$
C	$\{w_1, w_3, w_6\}$

Even in this simple case, the choice probabilities are not straightforward; for example, the probability of choosing A is given by (4.1). In it, alternative A can be chosen by selecting immediately aspect 4 or by selecting either 1 or 2 and then one of the remaining A's aspects.

$$P(A) = \frac{w_1}{\sum_{j=1}^{6} w_j} * \left(\frac{w_2 + w_4}{w_2 + w_3 + w_4 + w_6}\right) + \frac{w_2}{\sum_{j=1}^{6} w_j} * \left(\frac{w_1 + w_4}{w_1 + w_3 + w_4 + w_5}\right) + \frac{w_4}{\sum_{j=1}^{6} w_j} \tag{4.1}$$

The elimination process is recursive and so it is its formula. Therefore, there is no general formulation to easily estimate the EBA model. Three main strategies have been used to calculate this recursive formula:

i. For each alternative manually identify all paths that have as an outcome the inspected alternative chosen (Young et al., 1983; Manrai and Sinha, 1989; Fader and McAlister, 1990).

ii. Enumerate all paths and then identify which alternative is chosen in each path (Hess et al., 2012).

iii. Simulate the likelihood function to avoid the path construction (Gilbride and Allenby, 2006).

The first two strategies work efficiently while the structure of the elimination process does not change among DMs. This happens when there is a limited number of choice sets to be faced by the DMs –like in SP data– and the thresholds are assumed constant population-wide. However, these strategies are not suitable for RP data where every individual may face a different choice set or if thresholds are individually analysed. In these cases, the solution does not exploit the recursive nature of the problem.

The third strategy simulates the decision process for each individual. Therefore, it does not have any problem with changing choice sets among individuals and could be time efficient if the choice set is extensive. However, it has several drawbacks such as not calculating the likelihood function precisely and restricting the use of some Bayesian techniques[6] such as Gibbs Sampling since simulated likelihood does not offer conditional probability densities.

Our objective here is to propose a mathematical implementation for calculating the probabilities of the EBA model when the choice set varies across individuals. Our approach makes use of the recursive structure of the problem to speed up the first strategy.

4.1.1.1. Analytical estimation of the EBA model

We show that for low and intermediate complexities, exploiting the EBA structure to obtain analytical solutions offers smaller running times and higher accuracy than the simulated likelihood strategy. To obtain the choice probabilities, the choice set is analysed recursively to exploit the problem's recursive structure.

Let A be the set of all available aspects and a an element of A; let A_i the set of all aspects of alternative i and a_i an element of it. The initial condition of the recursion is expressed by (4.2):

---

[6] Indeed, Gilbride and Allenby (2006) use Metropolis-Hastings which could also be time consuming.

$$A = \bigcup_{\forall i} A_i \tag{4.2}$$

Let $P_i(A)$ be the probability of choosing alternative i when A is the set of available aspects. In each step, an aspect is chosen and all alternatives without that aspect are eliminated. Let $P(a, A)$ be the probability of selecting aspect a from the set of available aspects A (2.10). Then, the probability of selecting an alternative is given by (4.3).

$$P_i(A) = \begin{cases} 0 & \{\forall a_i \in A_i\} \notin A \\ \sum_{\forall a \in A} P_i(A - \{a\})\, P(a, A) & otherwise \\ 1 & A \in A_i \end{cases} \tag{4.3}$$

Expression (4.3) states three conditions. First, if no aspect of alternative i is part of the set of available aspects A, then it cannot be chosen. Then, the last equation states that if every other alternative has been discarded and only aspects of i remain, then i is the chosen alternative. Finally, the second expression is the recursion, which expands for every combination of the tree of possible decision routes. Note that this considers that no alternative is identical in terms of the aspects.

When applying (4.3) parametrically –rather than numerically– we obtain the analytic expression of the probability of choosing each alternative –like (4.1). This approach finds the tree of paths that has as outcome each of the alternatives; this is more efficient that enumerating paths independently. Finally, these analytic expressions can be directly plugged in into the desired estimation algorithm.

Using either the analytical approach or the simulation approach has different advantages in terms of the number of calculations, the ability to scale with complexity, and the precision of the probabilities estimated.

Regarding the number of calculations, the analytical approach calculates the tree of paths that selects each alternative. The algorithm obtains the probability density function of each branch and to which alternative corresponds. Thus, the number of calculation depends on the number of branches.

The simulation approach requires to simulate several times the path taken by the individual. Thus, the estimation time depends on the number of steps to reach an alternative and the number of simulations. However, it is straightforward noting that the number of steps to select a single alternative is always smaller or equal to the number of branches. Also, the number of simulations should be enough to guarantee that in each iteration of the algorithm – in our case a maximum likelihood algorithm– the chosen alternative has strictly positive probability density of being chosen.

The relation of the estimation times depends on the relation between the number of branches analysed in the analytical approach and the steps taken and the number of simulations in the simulated approach; neither of them is *a priori* more efficient.

The second characteristic of the algorithms is how they scale as complexity increases. The analytical approach estimation time is related with the number of possible branches. The complexity of the tree of possible paths increases as either the number of alternatives increases or as the number of aspects increases. Because of the structure of the EBA heuristic, it is straightforward noting that the number of branches tends to increase exponentially. The simulation approach estimation time is related with the number of steps required to reach a single alternative. When complexity increases, the number of steps required to reach an alternative increase, but at a slower rate than the number of possible branches. Therefore, the analytical approach scales worst with complexity than the simulation approach.

Finally, the last element is related with the accuracy of the estimated likelihood. The analytical approach calculates exactly the likelihood value at each point. Conversely, the simulation approach obtains an approximate value of the likelihood function at each point. The accuracy of such approximation may be improved by increasing the number of simulations per individual with the drawback of increasing the simulation time at each point calculated.

The inaccuracy of the estimation of the likelihood can have different impacts depending on the method used: Bayesian or maximum likelihood estimation. In Bayesian estimation the inaccuracy of the estimated likelihood impacts the accuracy of the sampling process, but does not slow it since the sampling number is fixed. On the other hand, in maximum likelihood estimation, the deviations could prevent the optimisation algorithm from taking the optimal path, hence, increasing the estimation time.

In conclusion, the balance between the advantages and disadvantages of each estimation procedure indicates the fastest option. In the next experiment we provide evidence for a specific context about the efficiency of the different approaches.

4.1.1.2. Experiment for comparing simulation and analytical approaches

The structure of the decision trees in the EBA model implies that the analytical approach scales worse with complexity than the simulation strategy. Nevertheless, we show that for the type of dataset used in this thesis, the analytical approach is faster than the simulation strategy for every available sample size.

To compare the performance of the estimation techniques, we used the “Las Condes - CBD, San Miguel CBD” dataset. We generated datasets of 1,000 DMs and explored choice sets of three and eight alternatives. For each of the two meta-experiments, ten experiments were performed. In each experiment, the model was estimated through both the simulation and parametrical approaches.

To generate the data, we used the procedure described in section 3.1 to obtain several datasets with the desired number of alternatives. The threshold values were designed through the methodology presented in Appendix B. Then, the EBA model is simulated considering two thresholds for cost and one for each of the *times* with weights and thresholds. Table 4-2 presents these parameters and the alternative specific constant parameters (ASC).

Table 4-2 EBA weights for each aspect

Aspect	Thresholds	Log-Weights	Aspect	Log-Weights
Cost ($CLP)	40; 100	1.39; 1.39	ASC4	0.59
Vehicle time (min)	15	1.39	ASC5	0.53
Waiting time (min)	5	2.30	ASC6	0.47
Walking time (min)	3	2.08	ASC7	0.18
ASC1	-	0.41	ASC8	0.26
ASC2	-	0	ASC9	0.34
ASC3	-	0.10		

To estimate the models, through both approaches we used maximum likelihood. The implemented algorithm for both approaches does not differ much and uses the same number of computational cores. To maximise the likelihood we used the Maxlik package (Henningsen and Toomet, 2011) for the R software. The starting point considered the same weight (equal to one) for each aspect.

The estimation through simulation requires to repeat the process several times to obtain average probabilities. Not having the required sample size could imply that a zero probability may be assigned to a chosen alternative and, therefore, obtain a null likelihood and fail to estimate the model. In this specific dataset, first 100 simulations per individual were used, but, it usually failed due to encountering a null likelihood at some point of the simulation. When, 200 simulations per individuals where considered; we could estimate the model in 19 out of 20 estimations. In the only failed case, the estimation was repeated with a different random number generator seed successfully.

4.1.1.3. Results of the analytical and simulation approaches

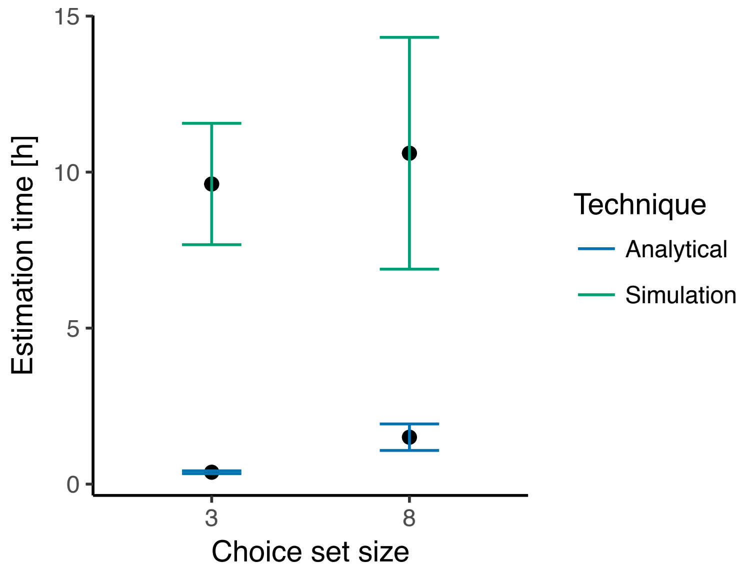
Table 4-3 presents the general results of the experiment. First, the estimation time for the analytical and simulation time is presented. Then, the ratio of estimation times is shown. Note that the analytical approach is always faster for this dataset. However, its advantage decreases as choice set complexity increases. Nevertheless, even in the most unfavourable case, the analytical approach is 7.1 times faster than the simulation approach.

Table 4-3 Estimation time of EBA model with the analytical and simulation approaches

Choice set size	Simulation approach estimation time [h]	Analytical approach estimation time [h]	Ratio of estimation times
3	9.62	0.38	25.2
8	10.60	1.50	7.1

Figure 4-1 presents the mean point and 95% confidence interval (two tailed t-student distribution with 9 degrees of freedom) of the estimation time for each choice set size and approach used. Note that again for our sample, the analytical approach outperforms the simulation approach, but the difference decreases as complexity increases. Also note that the simulation approach estimation time increases with sample size, but the difference is not statistically significant. Finally, note that the difference between the analytical and simulation approaches are statistically significant at the 95% confidence level.

Figure 4-1 Distribution of estimation times for each approach



In conclusion, for our sample size structure, the analytical approach outperforms significantly the simulation approach. Much complex choice sets are required for the simulation approach to be a better alternative than the analytical approach.

4.1.2. Estimation of thresholds for continuous attributes

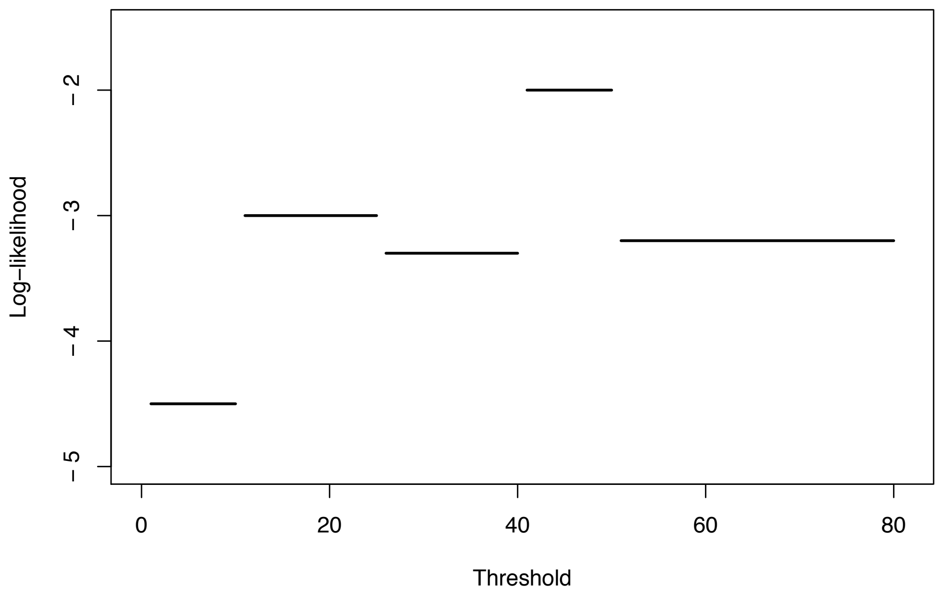
Subsections 2.2.2 and 4.1.1 show the typical structure of an EBA model. Working with it is straightforward if aspects are discrete; however, if attributes are continuous there is no readily way to model the problem. Typically, acceptability thresholds are imposed over continuous attributes, which impact the maximum likelihood obtained. If thresholds impact the maximum likelihood, then, they can be estimated using maximum likelihood. Therefore, thresholds and weights can be simultaneously estimated.

The EBA model does not have a generalizable equation; to estimate it, a tree of possible paths is built as shown in subsection 4.1.1. Thresholds determine the aspects belonging to each alternative and thus the structure of aspects in the alternatives. Hence, a change in the thresholds may change the aspect distribution and rearrange the tree of possible paths to choose an alternative.

Several approaches have been used to cope with the issue of the continuous attributes. Some authors have used elimination rules that compare alternatives (Hess et al., 2012), for example, by eliminating the alternative with the higher cost. Other authors have modified the EBA structure to accommodate alternatives with continuous attributes (Manrai and Sinha, 1989; Gensch and Ghose, 1992); however these approaches have had small impact. Finally, recently some other authors have included continuous attributes into the weight functions of the aspects (Kohli and Jedidi, 2015; 2017), bypassing the problem but not solving it.

Simultaneous estimation of thresholds and aspects entails two problems. First, as mentioned above, changing the thresholds may change the structure of the aspects. In this thesis, we do not tackle this issue. Therefore, when estimating thresholds, we recalculate the aspects' structure at each step. The second problem is its impact in the likelihood function. Changing thresholds only has an impact in the likelihood function if this changes the structure of the aspects. Therefore, the likelihood is flat across changes in the thresholds until at least one alternative of one individual changes the structure of its aspects (see Figure 4-2). When the aspect structure changes, the likelihood function changes, generating a discontinuity in it.

Figure 4-2 EBA's log-likelihood versus a fictitious threshold



As Figure 4-2 shows, there are two problems regarding the use of traditional optimisation tools for calculating the maximum likelihood. First, under changes in the thresholds, the concavity of the log-likelihood is not granted. For this reason, a robust optimisation tool must be used. Second, the likelihood function is flat under small changes in the thresholds; therefore, any method that is gradient based will fail. And finally, the likelihood function is non-continuous under changes in the thresholds, which could be problematic for some optimisation tools.

We could partially solve the three problems using the Simulated Annealing optimiser. Simulated Annealing can increase the likelihood of mis specified thresholds, but fails to find the maximum likelihood, even locally. To obtain a locally optimal solution, three optimisations were performed. First, to find an acceptable initial solution, we optimised the weights under fixed thresholds using a gradient optimiser like BHHH (Berndt et al., 1974). Then, Simulated Annealing was used to find a better threshold/weight combination by changing them simultaneously. Finally, a local optimum was obtained using the gradient based optimiser. An example of this procedure is described below.

4.1.2.1. Example of estimation of thresholds and weights

We simulated a population that chose alternatives following a EBA heuristic. We used 5,000 DMs which chose from random choice sets of the “Las Condes - CBD, San Miguel - CBD” dataset with three alternatives. The dataset was built following the methodology described in Section 3.1. The DMs chose using the weights and thresholds given in Table 4-2.

When estimating the EBA model, there are two time-consuming processes. The first is the formulation of the probability of choosing each alternative for each individual, which took 1.32 minutes in our experiment. The second time consuming process is the numerical calculation of the probability matrix for each individual, which took 7.18 seconds in our experiment. When optimising exclusively weights, the first calculation is done only once, since no threshold change occurs. Whereas in the threshold-weight optimisation, in each iteration of the algorithm, both processes are involved.

We started by optimising the initial weights subject to wrong thresholds as shown in Table 4-4. This provides a reasonable starting solution for the threshold-weight optimisation. The initial weights were equals to one and the optimiser used was BHHH.

Table 4-4 Initial thresholds for the weight-threshold estimation

Parameter	Cost	Vehicle time	Waiting time	Walking time
Thresholds	30; 80	25	10	8

After eight iterations of the algorithm[7] (equivalent to 2.6 hours), the optimiser found an optimal solution for the fixed thresholds EBA model (Table 4-5). This solution is the initial solution of the optimisation problem solved by using the Simulated Annealing algorithm.

Table 4-5 Optimal parameters for the fixed threshold EBA model

Aspect	Log-Weights	Aspect	Log-Weights
Cost ($CLP)	-84.6; 2.01	ASC4	1.62
Vehicle time (min)	-0.67	ASC5	1.77
Waiting time (min)	1.57	ASC6	1.76
Walking time (min)	1.22	ASC7	1.56
ASC1	1.67	ASC8	1.11
ASC2	2.01	ASC9 (fixed)	0.34
ASC3	1.22		

In each step, Simulated Annealing tests a combination of thresholds and weights. For each combination of thresholds, the formulation of the probability of choosing each alternative is analysed. After analysing 10,000 points in 223 hours, the algorithm produced the parameters

[7] In each iteration of the BHHH algorithm, the log likelihood is estimated at several points.

shown in Table 4-6. Finally, the thresholds obtained from the second optimisation were fixed and used to calculate final weights in a third optimisation. The results are shown in Table 4-7 and the likelihoods in Table 4-8.

Table 4-6 Estimation of thresholds and weights for the EBA model

Aspect	Thresholds	Weights	Aspect	Weights
Cost ($CLP)	32.1; 80.1	-82.3; 2.55	ASC4	1.73
Vehicle time (min)	24.1	-0.68	ASC5	1.65
Waiting time (min)	9.5	1.21	ASC6	1.58
Walking time (min)	5.0	2.88	ASC7	1.33
ASC1	-	1.54	ASC8	1.10
ASC2	-	1.38	ASC9	0.34
ASC3	-	0.89		

Table 4-7 Estimation of weights for the optimised thresholds of the EBA model

Aspect	Log-Weights	Aspect	Log-Weights
Cost ($CLP)	-82.3; 2.38	ASC4	1.83
Vehicle time (min)	-0.17	ASC5	1.6
Waiting time (min)	1.50	ASC6	1.59
Walking time (min)	2.98	ASC7	1.32
ASC1	1.55	ASC8	1.10
ASC2	1.44	ASC9 (fixed)	0.34
ASC3	0.96		

Finally, we calculated the likelihood of two models: that of the real underlying model and a model with the real thresholds but optimised weights (Table 4-8).

Table 4-8 Log-likelihood of successive estimations of the EBA model

Estimation	Log-likelihood
1a. Fixed thresholds	-5,188
1b. Thresholds and weights	-4,739
1c. Weights on optimised thresholds	-4,736
2. Real underlying model	-4.909
3. Optimised real underlying model	-4.565

Table 4-8 shows the likelihood of five estimations. First, note the important improvement in likelihood when considering different thresholds (1a vs 1c). Also note that although the thresholds of estimation 1b do not differ much from the initial case, the difference in structure of the tree of possible routes to choose an alternative is important, and this modifies substantially the likelihood. Also note that even though the Simulated Annealing algorithm did not find the optimal solution, the likelihood obtained is close to the likelihood obtained when optimising the weights with a continuous algorithm (1b vs 1c). Finally, note that the solution found using the proposed procedure (1c) is better than the real data generating process (2) but not better than an optimised case under real thresholds (3).

In conclusion, the proposed procedure can estimate simultaneously thresholds and weights for the EBA model. The proposed approach finds solutions that may be even better than the real underlying model; however, it does not outperform an optimisation of the weights under the real thresholds. Nevertheless, it provides a useful approach when thresholds cannot be fixed naturally.

The proposed approach has two main limitations. The first is the large computational time due to the difficulty in finding the tree of possible routes for choosing each alternative under EBA. The second is related to the nature of the solution found. As exposed in Figure 4-2, the likelihood is flat under small changes in the thresholds, therefore, the hessian matrix of the likelihood function is flat in the optimum. This issue implies that a covariance matrix cannot be calculated and no confidence interval may be obtained for the thresholds.

4.2. The Stochastic Satisficing (SS) Choice Model

Since psychologist first pointed out the potential impact of bounded rationality in decision making (Simon, 1955), there has been a growing consensus that people's limited processing faculties may affect the way they make decisions (Conlisk, 2014). This way, the concept of bounded rationality has permeated several disciplines, such as behavioural economics (McCain, 2015) and choice modelling (Araña *et al.*, 2008; Stüttgen et al., 2012).

Simon's work on Satisficing Theory (Simon, 1955, 1956), henceforth ST, provides the basis for the Satisficing choice heuristic. Even though Simon's work does not give a precise definition for this heuristic (Manski, 2017), it highlights what elements of 'rational' choice are highly implausible and what reasons could trigger a simpler behaviour by DMs.

Simon analysed three simplifying principles. First, he argued that any choice model requiring the inspection of all attributes and a comparison (or consideration) of all alternatives would be highly implausible in many practical applications; thus, simple pay-off functions should be expected[8]. Then, Simon argued that information gathering is costly due to cognitive and processing efforts, suggesting a reservation value or acceptance threshold. Finally, the third principle explicitly recognized is that DMs may have trouble combining attributes of a different nature (e.g. quality and cost) into a single figure of merit (e.g. utility). Thus, DMs actually consider only partial ordering pay-off functions.

Several of Simon's ideas have been applied into decision and search theory. For example, some studies implemented directly the cost of information (Gabaix et al., 2006). Other researchers have implemented indirectly the cost of information through sequential inspection of the choice set (Caplin *et al*., 2011; Manzini and Mariotti, 2014; Aguiar *et al.*, 2016) or by analysing sequential menus (Papi, 2012). Several models considered a reservation utility in accordance to ST (Gabaix *et al.*, 2006; Caplin and Dean, 2011; Papi, 2012). However, to the best of our knowledge, none of these studies have applied the third principle of ST (i.e. partial ordering pay-off functions), probably because it implies dismissing the concept of utility.

[8] Simple pay-off functions are, for example, distinguishing between acceptable and unacceptable alternatives. Even though Simon (1955) does not restrict the pay-off functions to be binary, to the best of our knowledge, only binary pay-off functions have been implemented when modelling Satisficing behaviour.

Following these theoretical considerations, discrete choice models have attempted to implement the principles in different ways; however, most models have not incorporated important cornerstones of the theory[9]. Whilst some applications of ST completely inspects all available alternatives –violating the second principle (Recker and Golob, 1979; Young *et al.,* 1983; Durbach, 2009), other applications mix attributes of a different nature into a single figure of merit, violating the third principle (e.g. Radner, 1975; Richardson, 1982; Araña et al., 2008). Only recently, ST has been thoroughly applied using eye-tracking technology (Stüttgen *et al.*, 2012), yet, this is not possible in most choice settings. Therefore, despite several attempts to implement ST in practice, a Satisficing choice model that can be used broadly with simple data[10] does not exist.

The main contribution of this chapter is the proposal of an econometric model, the Stochastic Satisficing model, that applies ST as rigorously as possible for a simple dataset[7]. To create this model, we start by describing a general Satisficing behaviour, which incorporates the three ST principles that could lead to several ST models. Then, simplifications are stated to adapt to the data structure and the econometric model is solved. As a result, the model considers that DMs choose the first alternative which is stochastically satisfactory on all dimensions of the pay-off vector. Thus, DMs are assumed to explore the choice set

---

[9] This is probably because Satisficing has been interpreted in different ways among researchers, without reaching a consensus (Manski, 2017).

[10] By simple data, we understand only alternative profiles and the chosen alternative.

sequentially, in a process based on alternatives rather than on attributes (Williams and Ortuzar, 1982a).

One of the key features of the Stochastic Satisficing model is the consideration of a multidimensional pay-off acceptability function. This approach, explicitly suggested by Simon (1955) and identified by us as the third principle, differentiates our model from previous work. By using this approach, we imprint further realism to the choice heuristic and do not restrict the model structure, which has scalar utility functions as a particular case. As the multiple dimensions of the pay-off function interact into a single stochastic acceptability, different substitution patterns are analytically obtained.

We test the proposed model's properties on synthetic and real data. The analysis on synthetic data suggests that the model could be unbiased and that consistency is reached with common sample sizes. The real data case provides an example where the model can adapt its behaviour when the evidence in the data suggests that constant compensation among attributes does actually exist.

The rest of the sections are organised as follows. In Subsection 4.2.1 we describe the ST principles and the reported evidence in the literature that motivates people using a Satisficing choice heuristic. In Subsection 4.2.2, we propose a general Satisficing behaviour theory which is later simplified into the Stochastic Satisficing model. We end Subsection 4.2.2 by analysing the analytical properties of the model. Then, Subsection 4.2.3 analyses the model in two contexts: synthetic and real data. Conclusions regarding our model is presented in

Subsection 4.2.4. Additional information is presented in Subsection 4.2.5 as well as the publication history of this theme.

4.2.1. On Simon's theory: principles and motivation

We first address the behavioural theory of rational choice proposed by Simon (1955, 1956) and discuss its main principles. Then, we analyse how context can induce a Satisficing choice heuristic.

4.2.1.1. Simon's theory principles

Most discrete choice models, such as RUM, EBA and RRM among others, require the evaluation of all alternatives, involving a large cognitive load for DMs. Furthermore, this burden is increased in RUM and RRM due to the consideration of all alternative attributes in compensatory trade-off terms of either utility or regret.

ST suggests several simplifications, or principles, that make the behavioural process more plausible for the human mind. We have categorized such simplifications into three main principles. The first, states that DMs may assume only a few evaluation outcomes per alternative (e.g. acceptable or not; desirable, neutral, or undesirable) instead of a continuous outcome (e.g. utility). In the Stochastic Satisficing model, we postulate that an alternative can be either acceptable or non-acceptable.

The second principle is based on the fact that information gathering is not costless. People may use information sequentially as they acquire it and use only a subset of the available information. For example, neither the need to visit an apartment before deciding if it is acceptable, looking at a shelf of a supermarket before settling for a bottle of wine, nor examining the attributes of alternatives presented in a stated choice survey are free of cost or burden. The higher the information cost is –probably relative to the importance of the choice decision– the simpler the cognitive process may become (e.g. not inspecting all alternatives or attributes). Simplifications can be attained by inspecting a subset of attributes of each alternative, as in the EBA model, or by inspecting a subset of alternatives as in the Satisficing heuristic. In the Stochastic Satisficing model, we assume that people truly choose the first "good enough" alternative.

Finally, the third principle is associated with the difficulty that DMs may have in mixing attributes of a different nature (e.g. quality and cost). Contrary to this principle, in random utility modelling for example, the analyst assumes that DMs are willing to compensate attributes at certain marginal rates of substitution. ST suggests that DMs may not analyse such attributes conjointly, but rather consider them independently and still infer if the alternative is acceptable or not.

4.2.1.2. How context can induce satisficing

Several reasons why someone could choose using a Satisficing heuristic have been reported in the literature. Simon (1956) suggests that when the choice is too complex, people could use simpler heuristics to cope with the high cognitive burden. We think that if complexity is related to the size of the choice set, then Satisficing behaviour or any other heuristic based on alternative discarding is highly plausible. Conversely, if complexity is associated with the number of attributes, then EBA or any other choice heuristic based on attribute discarding may be expected. If complexity is not an issue, then utility maximization might be a plausible choice heuristic.

A second argument related to the propensity of using a Satisficing heuristic is that search costs could prevent people from inspecting the complete choice set (Simon, 1955). Indeed, maximizing utility considering search costs leads to a class of satisficing behaviour (Richardson, 1982). Nevertheless, even if satisficing is optimal under utility maximisation with search costs, ST dismisses the concept of utility as being intractable for DMs (Manski, 2017). Finally, we postulate that costs may be interpreted not only as direct costs (e.g. monetary or time), but as indirect costs (e.g. effort) –a similar definition may be found in Chassang (2013). Under this interpretation, even the mere possibility of losing a quasi-unique good for not making the choice fast enough (e.g. in a dwelling or real-estate choice), could be a cost that triggers the Satisficing choice heuristic.

Finally, Simon (1956) also questions the idea that DMs could even try to optimize a decision or to maximize utility. We do not discuss this notion here, but rather concentrate on the formulation of a practical model under the assumption that a Satisficing heuristic is appropriate.

4.2.2. The Stochastic Satisficing model

We start by describing a general Satisficing behaviour in accordance to ST. Then, several simplifying assumptions are proposed to adapt the model to the type of data that we want to use. Then, upon that simplified behaviour, we propose and solve an econometric model. We end up this subsection by analysing the proposed model's analytical properties.

4.2.2.1. A general Satisficing behaviour

A choice heuristic describes how DMs choose one alternative from a choice set. It starts by analysing how an individual faces a choice set and ends by choosing an alternative. In this model, DMs face alternatives sequentially and choose the first satisfactory alternative. This simple choice heuristic is divided into four stages or components.

First, DMs start by analysing an alternative of the choice set. The starting alternative is chosen in accordance to a probability density function. Depending on the nature of the choice set, the way to approach each alternative may differ and, therefore, the probability of choosing such alternative first could vary. For example, a list of alternatives in a stated choice experiment may be read sequentially; while products in a shelf may be faced differently depending on their position. Thus, attributes related with the probability of inspecting a certain alternative first may need to be estimated (Stüttgen *et al.*, 2012).

The second component is a transition probability between alternatives. Once an alternative is inspected, the probability of inspecting another one could be identical or could vary in terms of its attributes. For example, a product located low in a shelf may be harder to reach than another 'better located', reducing the former probability of inspection.

The third component is an acceptability function. Typically, the acceptability of an alternative has been modelled by means of a utility function and a reservation utility. DMs choose an alternative if its utility surpasses the reservation utility (Richardson, 1982; Tyson, 2008; Caplin and Dean, 2011; Zhao and Huang, 2016). Despite common practice, we model the acceptability function as a partial pay-off function or a vectorial function of acceptability in accordance with ST (Simon, 1955)[11]. In our model, the acceptability of an alternative i (A_{iq}) evaluated by individual q is given by the acceptability of each component of the acceptability vector as stated in (4.4).

$$Pr\left(A_{iq} = 1\right) = \prod_{\forall\, k\, \epsilon\, K} Pr(a_{kiq} = 1) \tag{4.4}$$

Each element that impacts the acceptability of an alternative may be interpreted as an attribute or combination of attributes. Attributes being compared in a compensatory way are evaluated in the same acceptability function, whereas attributes non-compensated are evaluated in different functions; further comments are addressed in Subsection 4.2.2.4. Thus,

---

[11] Note that in ST, the acceptability function could be dynamic since the preferences are built in the choice process rather than being defined externally. However, we do not incorporate this element of the theory.

modelling a utility function is a restricted case where every attribute is compared in the same acceptability function.

Finally, the last stage concerns the behaviour of DMs once an acceptable alternative is found. Theoretically, we state that DMs can continue searching alternatives with certain probability, which could even be decreasing while the search continues longer after encountering the first acceptable alternative. Indeed, eye-tracking data (Stüttgen et al., 2012) suggests that once DMs find a satisfactory alternative they do not choose it immediately.

4.2.2.2. Simplifications to the general Satisficing behaviour

The general Satisficing behaviour applies rigorously ST; however, it requires rich datasets to be estimated. We adapt the general Satisficing behaviour to formulate a model that may be used with simple datasets. The data we want to work with contains just a full profile of each available alternative and knowledge of which was the chosen alternative. Thus, we intend to create a model based on a path dependent heuristic with unknown search paths. To accomplish it, several assumptions or simplifications need to be considered.

The first simplification is associated to the probability of starting with a particular alternative. Given that the search path is unknown, the starting alternative is also unknown. Then, it is not possible to model what factors affect the probability to start with a certain alternative. Thus, we assume that the first alternative is randomly chosen with equal probability.

The second simplification involves the transition probability between alternatives. It is not possible to estimate a probability function of transition since the inspected alternatives and search paths are unknown. For this reason, we consider an equal transition probability between all alternatives.

Finally, we simplify how long the search continues after finding the first acceptable alternative. Because the length of this search is unknown, it is not possible to estimate any stopping criteria. Therefore, we assume that DMs choose their first acceptable alternative.

With these three simplifications, which are summarised in Table 4-8, we intend to estimate a Satisficing behaviour model with simple data. Nevertheless, availability of richer data could allow us to relax some of these simplifications, and another model could be formulated.

Table 4-9 Satisficing choice elements, data limitations and simplifications

Element	Problem	Simplification
Probability of starting with a particular alternative	Initial alternative is unknown	Equal probability of starting for all alternatives
Transition probability	Choice path is unknown	Equal probability of inspecting each alternative that has not been inspected
Probability of choosing, conditional on having found an acceptable alternative	Number of alternatives after inspecting the acceptable one is unknown	The first satisfactory alternative is chosen immediately

4.2.2.3. The mathematical model

The mathematical model described in this subsection represents the simplified Satisficing behaviour proposed in subsection 4.2.2.2. In this model, DMs randomly inspect their choice sets and choose the first satisfactory alternative. We start by defining the criteria for satisfaction by linking the acceptability of an alternative to the acceptability of each attribute. Finally, our model ends by linking the acceptability of each alternative to the probability of choosing it.

Alternative acceptability

Following the simplified Satisficing behaviour, the DMs chooses the first acceptable alternative. According to ST, an alternative is acceptable if all attributes are satisfactory. Let A_{iq} be the probability that alternative i is acceptable for individual q. Then, A_{iq} is the joint probability that each attribute of i is acceptable. If we assume that the acceptance of each attribute is independent, then the joint probability is given by the product of the acceptability of the various attributes (4.4) as stated in Subsection 4.2.2.1.

Attribute acceptability

The acceptability of an alternative is based on the acceptability of its attributes. To define attribute acceptability, let K be the set of attributes and k an element of it. Each individual q has a set of acceptability thresholds; let F' be that set and f' an element of it. Each threshold f' is associated with a certain attribute k. Without loss of generality, we assume that more of each attribute is desirable. Furthermore, we could also assume that thresholds f' are a function for each individual. An attribute could be a single trait of the alternative or a combination of several characteristics. We refer to the acceptability of an attribute, a_{kiq}, in terms of its quantity or level x_{kiq}, as in (4.5).

$$a_{kiq} = \begin{cases} 1, & if\ x_{kiq} > f'_{kiq} \\ 0, & otherwise \end{cases} \tag{4.5}$$

The threshold (f') represents the aspirational level for the specific attribute; a similar definition can be found in the work of Radner (1975) and Stüttgen *et al.*, (2012). This function does not depend on the level of the attributes but could be influenced by socio-demographic characteristics and experimental conditions. For example, people with lower income could be very sensitive to cost; whereas the wealthy could almost ignore this attribute by having a higher cost threshold. We do not analyse the origin of this aspirational level, but follow Simon (1956): "it has no problem of maximization" involved. Further, we will assume, as in the random utility framework, that there are elements that the researcher can observe, which we will denote by f, and others that the researcher cannot observe and will be captured by random disturbances, ϵ_{kiq} . We will further assume that ϵ_{kiq} has a Logistic distribution with mean zero and variance σ^2_{kiq}. Then, the probability that an attribute is acceptable is given by (4.6):

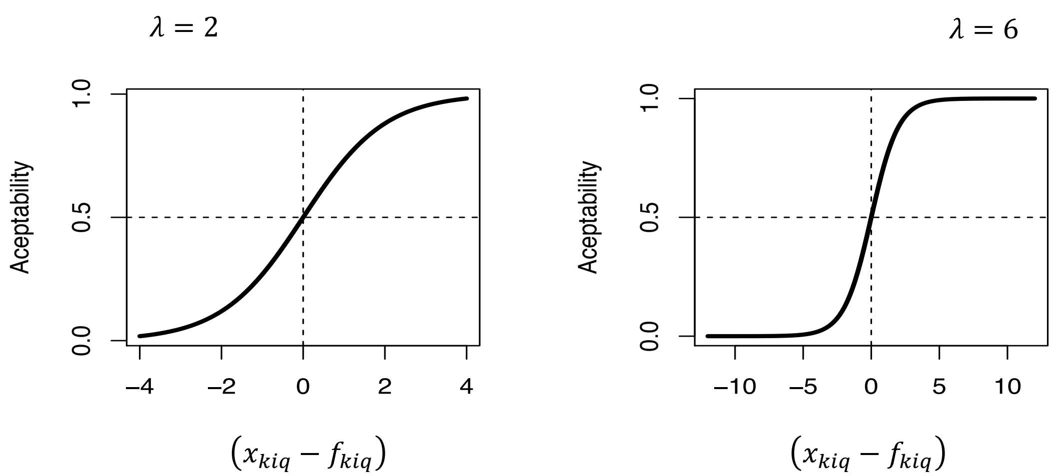
$$Pr\left(a_{kiq} = 1\right) = Pr\left(x_{kiq} > f_{kiq} + \epsilon_{kiq}\right) = Pr(x_{kiq} - f_{kiq} > \epsilon_{kiq}) \tag{4.6}$$

Expression (4.6) is transformed into (4.7) by using the expression for the cumulative probability function of the logistic distribution.

$$Pr(a_{kiq} = 1) = \frac{\exp\left(\lambda_{kiq}(x_{kiq} - f_{kiq})\right)}{1 + \exp\left(\lambda_{kiq}(x_{kiq} - f_{kiq})\right)} \text{ , } with\ \lambda = \frac{\pi}{\sigma\sqrt{3}} \tag{4.7}$$

Expression (4.7) is based on two terms, the scale factor and the threshold function; both are analysed in Figure 4-3. The scale factor (λ_{kiq}) represents the impact of an additional unit of x_{kiq} in the probability of accepting attribute k; thus, higher values imply a higher sensitivity to changes in the attribute. On Figure 4-3, the horizontal axis presents the difference between the attribute and its threshold.

Figure 4-3 Acceptability function versus different scale factors and attribute-threshold differences



As the level of the desirable[12] attribute increases, higher is the probability of acceptance. The interpretation of the threshold function is two-fold. On the one hand, it indicates the point

---

[12] An attribute is desirable if the scale function is positive; further details are provided in subsection 4.2.2.4.

where attribute acceptability is 50%. Then, an increase in the threshold implies an increase in the point where acceptability reaches 50%. On the other hand, higher values of the threshold are related with a decrease in the difference $x_{kiq} - f_{kiq}$ or a decrease in the quantity that surpasses the threshold; hence, increasing a threshold implies a decrease in the probability of accepting the attribute.

Probability of choosing an alternative

The link between acceptability of an alternative and its probability is developed using the assumptions made in subsection 4.2.2.2. Let t_{qi} determine the alternative chosen by the individual as in (4.8):

$$t_{qi} = \begin{cases} 1 & \text{if alternative i is chosen} \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

Because we only know the choices but not their search paths, we assume that DMs choose the first satisfactory alternative. To link this assumption to the probability that alternative i is chosen by the individual, every possible path that ends in choosing alternative i must be computed. To avoid this calculation, we suppose that DMs could have separated their complete choice sets into two: a set of acceptable alternatives and another one of unacceptable alternatives. Note that these separated choice sets are designed as an artifice to solve the mathematical problem; indeed, DMs never split their choice sets.

Let I' be the set of acceptable alternatives and $|I'|$ its cardinality. If the search path is completely random (i.e. every alternative has the same probability of being analysed), then, the probability of choosing alternative i is the probability of inspecting it before another alternative in I'. This probability is given by (4.9):

$$\Pr(t_{qi} = 1|I') = \frac{1}{|I'|} \tag{4.9}$$

Expression (4.9) is based on the DMs acceptable choice sets. The acceptable choice sets are based on the probability that each of its alternatives is acceptable. Let $\Pr(I')$ be the probability that only alternatives in I' are acceptable. Then, assuming that each alternative is analysed independently, the probability of choosing alternative i is given by the total probability over every conditional choice sets, shown in (4.10):

$$\Pr(t_{qi} = 1|I') = \frac{1}{|I'|} \tag{4.10}$$

Expression (4.10) establishes the link between the acceptability of an alternative, through the acceptable choice set, and the probability of choosing it. The only unknown element in (4.10) is the acceptable choice set probability, which is explained below.

Acceptable choice set probability

The probability that subset I' exists, $\Pr(I')$, is built from its alternatives' acceptability. For example, if there are m acceptable alternatives and n non-acceptable ones, the probability that only the m alternatives in I' are acceptable is given by (4.11):

$$\Pr(I') = \prod_{m \in I'} Pr(A_{mq} = 1) \prod_{n \notin I'} \Pr(A_{nq} = 0) \tag{4.11}$$

Opt-out alternative

The only case that has not been defined yet is the probability of choosing no alternative (i.e. opting-out). If an opt-out option is available, as in many stated choice experiments, its probability is straightforward and is given by (4.12):

$$\Pr(opt - out) = \prod_{\forall i \in I} Pr(A_{iq} = 0) \tag{4.12}$$

Likelihood function

Given that every outcome has been analysed, it is possible to define the likelihood function. If p_{*q} is the probability assigned by the model to the chosen alternative, then, when the opt-out alternative is available, the log-likelihood of the model is given by (4.13):

$$ll = \sum_{\forall q \in Q} \log(p_{*q}) \tag{4.13}$$

If there is no opt-out alternative (i.e. DMs are forced to choose) or if we have information only about the DMs that chose, the estimated probabilities must be adjusted to represent only this spectrum of DMs. We can estimate the conditional probability upon their choices as in (4.14):

$$\Pr(alt\ i) = \Pr(alt\ i|choosing\ any\ alt) * \Pr(choosing\ any\ alt)$$

$$\Pr(t_{qi} = 1) = \Pr\left(t_{qi} = 1|\sum_i t_{qi} = 1\right)\Pr\left(\sum_i t_{qi} = 1\right)$$

$$p'_{qi} = \Pr\left(t_{qi} = 1|\sum_i t_{qi} = 1\right) = \frac{\Pr(t_{qi} = 1)}{\sum_j \Pr(t_{qi} = 1)} \quad (4.14)$$

And the estimated log-likelihood without an opt-out alternative is given by (4.15).

$$ll = \sum_{\forall q \in Q} \log(p'_{*q}) \quad (4.15)$$

Finally, if no opt-out alternative is available, then estimating absolute acceptability is meaningless since the possibility of not choosing any alternative does not exist. In such case, the model estimates a relative acceptability, rather than an absolute alternative acceptability.

4.2.2.4. Model properties

We explore some of the model's analytical properties, relax some assumptions made in the process and explore its performance.

Identifiability

The identifiability of the acceptability functions depends on whether there is an opt-out alternative or not. If there is one, every element of the acceptability function is identifiable.

In that case, both the scale factor and threshold functions of each attribute plus an alternative specific constant can be estimated. If there is no opt-out option, the model can only estimate the relative difference in acceptability and one alternative specific constant must be fixed, as in classical discrete choice models.

For example, in the classic logit model the scale factor is unidentifiable from the utility function. In our model, the scale factor is identifiable in the two cases discussed above. The probability of an attribute being acceptable is given by (4.7); in it, the expression in the exponential function can be decomposed as (4.16):

$$\lambda_{kiq}\left(x_{kiq} - f_{kiq}\right) = \lambda_{kiq}x_{kiq} - \lambda_{kiq}f_{kiq} \tag{4.16}$$

Since by definition f_{kiq} does not depend on the attribute level, x_{kiq}, both f_{kiq} and the scale factor (λ_{kiq}) can be identified.

Working with non-desirable attributes

In Subsection 4.2.2.3 we assumed that attributes were desirable. If an attribute is undesirable (e.g. cost), the probability function of acceptance – (4.6) and (4.7) – switches to (4.17):

$$\begin{gathered} P\left(a_{kiq}\right) = P\left(x_{kiq} < f_{kiq} + \epsilon_{kiq}\right) = P(x_{kiq} - f_{kiq} < \epsilon_{kiq}) \\ P\left(a_{kiq}\right) = \frac{\exp\left(-\lambda_{kiq}\left(x_{kiq} - f_{kiq}\right)\right)}{1 + \exp\left(-\lambda_{kiq}\left(x_{kiq} - f_{kiq}\right)\right)} \end{gathered} \tag{4.17}$$

Note that the difference between (4.7) and (4.17) is that the scale factor –that is always positive– is preceded by a negative sign if the attribute is undesirable. As the scale factor sign can be estimated, there is no need to define *a priori* if an attribute is desirable or not. Indeed, if λ_{kiq} is freely estimated, a positive result would imply that an attribute is desirable; while if it is negative, the attribute is undesirable. As in random utility models, the only problem is when the scale factor is zero, in which case the parameters cannot be identified. This could only happen if the variance of the error term is infinite, and thus the model would not suit the problem.

Use of categorical variables

A third analysis refers to the use of dummies or categorical variables. When using this type of variables there is no continuous threshold implied, just the presence or absence of the variable. Then, we could define the probability of accepting the presence of the characteristic as in (4.18):

$$\Pr\left(a_{kiq} = 1\right) = \frac{exp\left(\lambda_{kiq} f_{kiq}\right)}{1 + exp\left(\lambda_{kiq} f_{kiq}\right)} \tag{4.18}$$

Given (4.18), it can be deduced that the scale factor is not identifiable from f_{kiq} –the same happens in RUM models– and must be interpreted simultaneously or normalised, as usual to unity, as in (4.19):

$$\Pr\left(a_{kiq} = 1\right) = \frac{\exp\left(f_{kiq}\right)}{1 + \exp\left(f_{kiq}\right)} \tag{4.19}$$

Search costs

Several authors have theoretically accounted for search costs (Richardson, 1982; Tyson, 2008; Caplin and Dean, 2011). In our framework, the total search cost cannot be incorporated explicitly since each attribute is treated independently. However, we can assume that there is a fraction of the costs that could be associated with each attribute (e.g. monetary costs associated with the cost variable); let c_k be that cost and γ_k its sensitivity. Following Richardson (1982), the search costs would imply that the probability of accepting an attribute is given by (20).

$$\Pr(a_{kiq}) = \Pr(x_{kiq} > f_{kiq} + \gamma_k c_k + \epsilon_{kiq}) = \Pr(x_{kiq} - f_{kiq} - \gamma_k c_k > \epsilon_{kiq}) \tag{4.20}$$

An increase in the search costs increases the probability of accepting an attribute, since continue searching induces additional costs. If we assume that the cost is constant at each step of the process and that the function f_{kiq} has an attribute specific constant, then the cost is not identifiable from the constant. Hence, if we use an acceptance function that has an attribute specific constant, it would incorporate implicitly the search cost. Therefore, the use of an attribute specific constant in each acceptability function is highly recommended.

Understanding the rate of substitution in the Stochastic Satisficing model

The Stochastic Satisficing model allows for two types of substitution patterns between attributes. The first type focuses on attributes modelled in the same acceptability function; whereas the second, targets attributes in different acceptability functions.

Note that in this model there is no utility function, so a marginal rate of substitution (MRS) over a utility function is not possible. However, a MRS over the acceptance function attends a similar purpose in the sense of allowing to study the trade-offs that enable the probabilities to be constant. Still, any substitution pattern should not be used in cost-benefit analysis, since there is only a vague relationship with individual's welfare. Yet, the MRS over the acceptability functions are valuable because they enable us to understand the behaviour of the model given a change in the attributes.

The first analysis considers attributes modelled in the same acceptability function. As anticipated in Subsection 4.2.2.1, considering attributes in the same acceptability function entails a direct compensation between them. Let C_k be the space of attributes to be compensated with attribute x_k and c an element of C_k. (4.21) states the structure attributes of an acceptability function where direct compensation is allowed. As a result, the term x_{kiq} in (4.7) is replaced by (4.21) leading to (4.22):

$$x_{kiq} + \sum_{\forall c \in C_k} \theta_c x_{ciq} \tag{4.21}$$

$$\Pr(a_{kiq} = 1) = \Pr\left(x_{kiq} + \sum_{\forall c \in C_k} \theta_c x_{ciq} - f_{kiq} > \epsilon_{kiq} \right) \tag{4.22}$$

From (4.21), it is straightforward to note that the rate of substitution between x_{ciq} and x_{kiq} in the whole data spectrum is θ_c. Subsections 4.2.3.1 and 4.2.3.2 presents two examples of applications of this type of substitution patterns.

The second analysis considers the relationship between attributes in different acceptability functions. We suggest that attributes should be modelled in different acceptability functions if the degree of compensation is limited. Even though no direct compensation is possible, in Equation (4.4) an increase in the acceptability of one attribute could substitute for the loss in another one. Then, the Stochastic Satisficing model, at least in an analytical way, allows for a (weak) substitution of one attribute for another in the case of attributes of a different nature.

Let v_{kiq} be the difference of the current attribute from its threshold ($v_{kiq} = x_{kiq} - f_{kiq}$). To obtain the MRS for alternative acceptability, first we need the derivative of the acceptability function (Equation 4.4). Let I^k be the space of all alternatives except k (i.e. $I^k = I - \{k\}$). Then the derivative of the acceptability function is given by (4.24):

$$\frac{\partial \Pr\left(A_{iq} = 1\right)}{\partial x_k} = \frac{\partial \Pr\left(a_{kiq} = 1\right) \prod_{\forall j \in I^k} \Pr\left(a_{jiq} = 1\right)}{\partial x_k}$$

(4.24)

$$\frac{\partial \Pr\left(A_{iq} = 1\right)}{\partial x_k} = \frac{\lambda_{kiq} \exp\left(\lambda_{kiq}\left(x_{kiq} - f_{kiq}\right)\right)}{\left[1 + \exp\left(\lambda_{kiq}\left(x_{kiq} - f_{kiq}\right)\right)\right]^2} \prod_{\forall j \in I^k} \Pr\left(a_{jiq} = 1\right)$$

If we replace the difference of the attribute and threshold for v_{kiq}, then the marginal rate of substitution is given by (4.25).

$$MRSx_1x_2 = \frac{\frac{\partial \Pr\left(A_{iq} = 1\right)}{\partial x_1}}{\frac{\partial \Pr\left(A_{iq} = 1\right)}{\partial x_2}} = \frac{\frac{\lambda_1 \exp(\lambda_1 v_1)}{(1 + \exp(\lambda_1 v_1))^2} \prod_{\forall\, j_1 \in I'^1} P\left(a_{j_1}\right)}{\frac{\lambda_2 \exp(\lambda_2 v_2)}{(1 + \exp(\lambda_2 v_2))^2} \prod_{\forall\, j_2 \in I'^2} P\left(a_{j_2}\right)}$$

$$MRSx_1x_2 = \frac{\frac{\lambda_1 \exp(\lambda_1 v_1)}{(1 + \exp(\lambda_1 v_1))^2} \Pr\left(a_{2iq} = 1\right)}{\frac{\lambda_2 \exp(\lambda_2 v_2)}{(1 + \exp(\lambda_2 v_2))^2} \Pr\left(a_{1iq} = 1\right)} \tag{4.25}$$

$$MRSx_1x_2 = \frac{\frac{\lambda_1 \exp(\lambda_1 v_1)}{(1 + \exp(\lambda_1 v_1))^2} \frac{\exp(\lambda_2 v_2)}{(1 + \exp(\lambda_2 v_2))}}{\frac{\lambda_2 \exp(\lambda_2 v_2)}{(1 + \exp(\lambda_2 v_2))^2} \frac{\exp(\lambda_1 v_1)}{(1 + \exp(\lambda_1 v_1))}} = \frac{\frac{\lambda_1}{1 + \exp(\lambda_1 v_1)}}{\frac{\lambda_2}{1 + \exp(\lambda_2 v_2)}}$$

From (4.25) we can see that the marginal rate of substitution (MRS) between any two attributes in different acceptability function is given by (4.26):

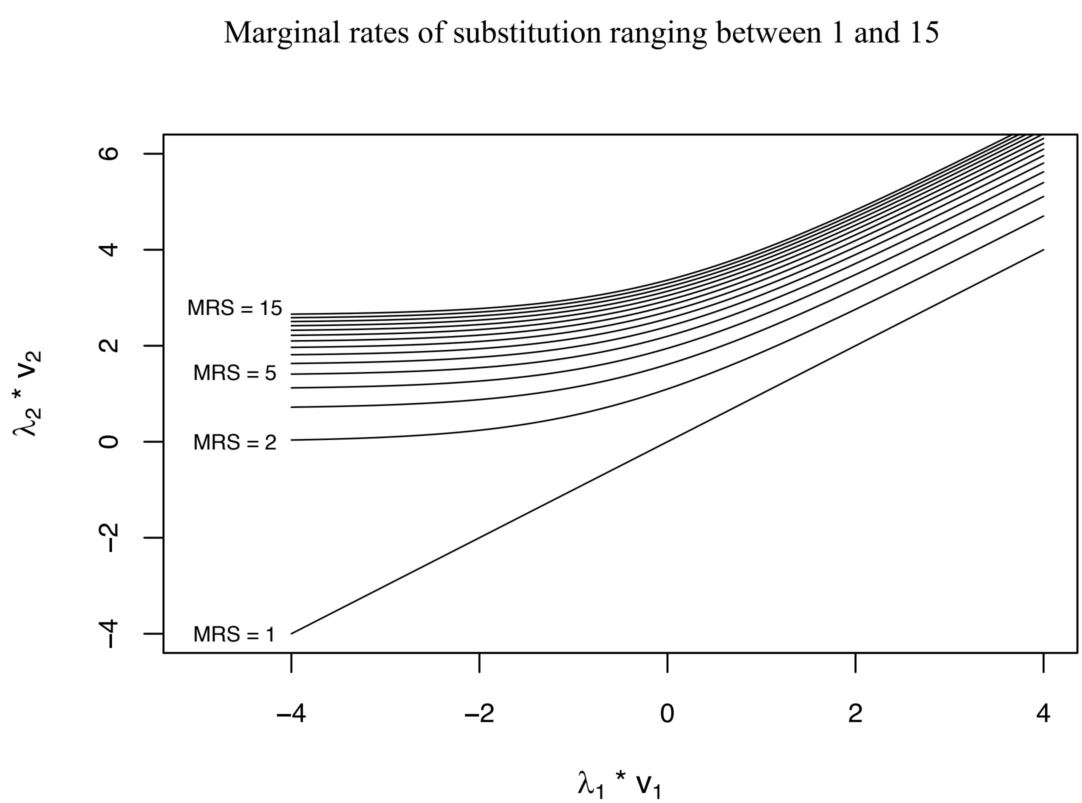
$$\mathrm{MRS}_{\mathrm{x1,x2}} = \frac{\lambda_1(1 + \exp(\lambda_2 v_2))}{\lambda_2(1 + \exp(\lambda_1 v_1))} \tag{4.26}$$

We are interested in finding if an acceptable attribute can compensate for a non-acceptable one; that is, if a positive $\lambda_2 v_2$ can substitute a negative $\lambda_1 v_1$. Figure 4-4 shows the indifference curves between two attributes of different acceptability functions; the MRS varies over the product of λ_i and v_i. Note that if $\lambda_i v_i = 0$ there is a 50% probability of accepting the attribute and for $\lambda_i v_i = -2$ there is a 12% acceptance probability.

Expression (4.26) suggests that the MRS is not constant over attribute levels. From Figure 4-2 we can deduce that when both attributes have similar satisfaction levels, a substitution is reasonable since MRSs are close to one. However, to trade an unacceptable attribute $(\lambda_1 v_1)$ for an acceptable one $(\lambda_2 v_2)$, the high MRS suggests that many units of

$\lambda_2 v_2$ must be traded for one unit of $\lambda_1 v_1$, which might be unfeasible. For example, 8.5 units of $\lambda_2 v_2$ at +2 must be traded to compensate the loss of one unit of $\lambda_1 v_1$ at -4; while the MRS of 15 is reached when $\lambda_2 v_2$ is +2.65. These results suggest that the DMs are willing to compensate when the satisfaction of attributes has similar levels; further, if one attribute is undesirable individuals may not be willing to trade-off.

Figure 4-4 Indifference curves of two attributes of different acceptability functions



Moreover, (4.26) could represent another feature of human behaviour: when an attribute is satisfactory, a person might prefer to increase another non-satisfactory attribute rather than obtaining an increase in the satisfactory one. So, this model allows to analyse two phenomena: first, why people could focus on second order needs only after high order needs are fulfilled, and second, the issue of attribute non-attendance. For example, when buying a public transport ticket, low income people could be highly sensitive to cost (cost is in the unacceptable spectrum) and thus, they would not accept to pay a higher price for additional comfort. Conversely, wealthier people feeling that cost is acceptable, would be prepared to pay for additional comfort since price is already in an acceptable level. Therefore, this model would allow to identify restrictions in the DM's decision process.

To sum up, the Stochastic Satisficing model enables to model attributes that can be compensated, weakly compensated or non-compensated at all. This feature could give the model the flexibility to interpret different contexts.

4.2.3. Application to data

In this subsection, we apply our model to test its properties with finite samples and in different behavioural conditions. First, we use synthetic data with the objective of analysing convergence to simulated parameters; later we apply the model to a real choice context. We use maximum likelihood estimation throughout.

In both experiments we use the Las Condes – CBD, San Miguel – CBD dataset (Chapter 3.1) In the application to synthetic data, we use choice set sizes ranging from two to nine alternatives and three samples sizes: 500, 1000 and 5000 observations. Whereas in the real data experiment, we use the whole sample.

It is well known that maximum likelihood estimates are consistent or asymptotically unbiased. However, we are interested in testing the model in finite samples sizes that are frequently reported in the literature. To test the approximate behaviour in finite samples we used synthetic data. The objective was to analyse the bias, the dispersion of the estimates, and possible identifiability issues. To test these elements, we analysed several sample sizes and different experiments within each sample size.

The simulated individuals searched for alternatives sequentially until they found a satisfactory one; then, this alternative was immediately chosen and the search finished. No opt-out alternative was considered, so only individuals who choose are present in the data. The synthetic population only considered cost, travel time and walking time. There was no compensation between cost and times, but there was compensation allowed between travel time and walking time. We also created a small preference for each alternative given by an alternative specific constant (ASC). The values of the parameters used to generate the simulated choices are shown in Table 4-10.

We analysed three sample sizes: 500, 1,000 and 5,000 observations; representing typical sample sizes that may be found in practice. To create the databank, we sampled random observations (i.e. sets of alternatives) from the real data set. For each sample size, 30 different and independent datasets were generated.

Table 4-10 Parameters used for simulation

Parameter	**Value**	**Parameter**	**Value**
Cost sensitivity	-0.1	Cost threshold	45
Time sensitivity	-0.2	Time threshold	37
MRS travel – walking time	3.0	ASC1	1.6
ASC2	1.4	ASC3	1.2
ASC4	1.0	ASC5	0.8
ASC6	0.6	ASC7	0.4
ASC8	0.2	ASC9 (fixed)	0

In Figures 4-5 to 4-10, the boxplot presents the 25% quantile, the median, and the 75% quantile of the estimated parameters. Alternative specific constants are analysed in Appendix B. Additionally, the vertical line through the box, shows the minimum and maximum estimates for each parameter. We divided the estimated parameters by their target values to obtain a relative statistic easier to visualize. Finally, we plotted the mean estimate with a dot and, as a matter of reference, the unit value with a dashed line. Since every value is divided by the target value, estimates near the dashed line show unbiased estimations. The difference between the mean point – mean of the 30 estimates – and the dashed line would show a systematic bias throughout the experiments.

Figure 4-5 Estimated parameters relative to targets in the 500 observations' sample

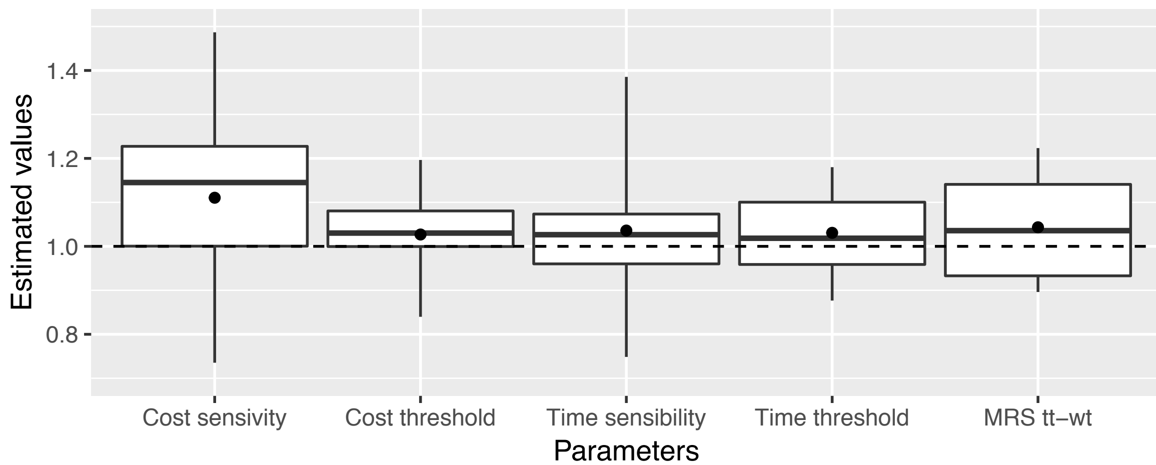


Figure 4-6 Estimated parameters relative to targets in the 1,000 observations' sample

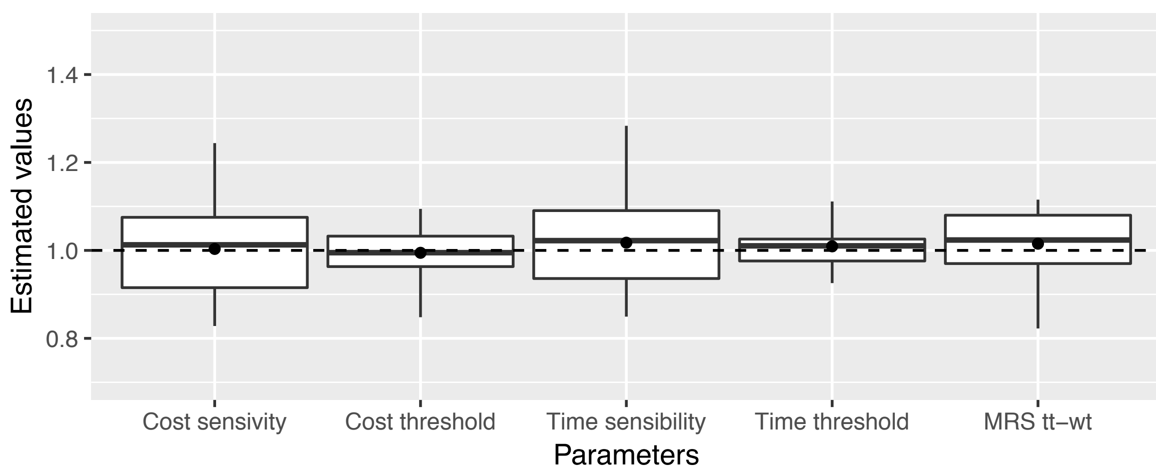
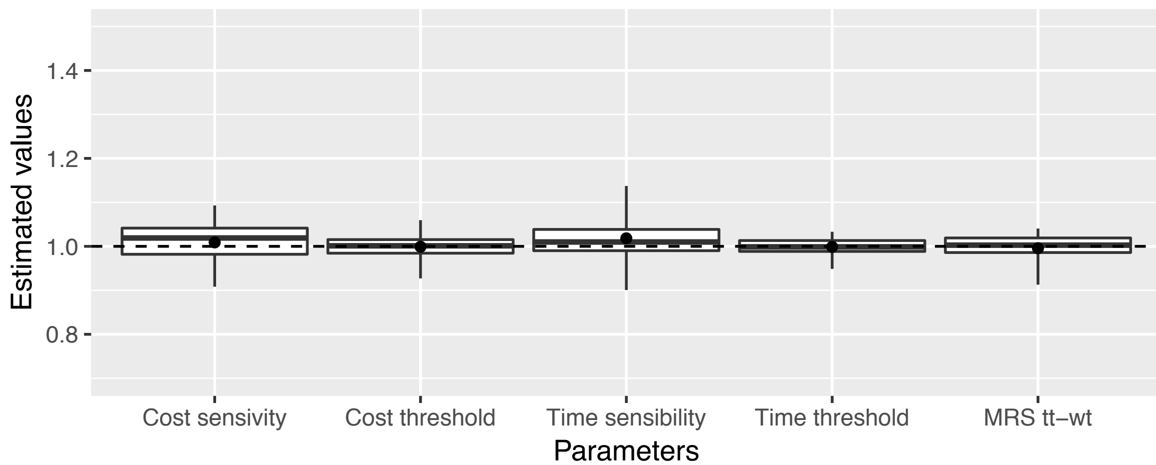


Figure 4-7 Estimated parameters relative to targets in the 5,000 observations' sample



The 1,000 and 5,000 observations' samples allow us to estimate the model with little variance throughout the 30 estimations. Interestingly, for every sample size the model tends to be unbiased in this dataset, since the points (mean estimates) are contiguous to the dashed line (target values). Moreover, model consistency has a desirable behaviour since the estimates tend toward their target values relatively fast as the sample grows.

The alternative specific constant has a higher variance than the sensitivities and is biased for the smaller sample sizes, overestimating the attribute as shown in Figures 4-8 and 4-9. For higher number of observations, the model is unbiased and consistent as shown in Figure 4-10.

Figure 4-8 Alternative specific constants relative to targets in the 500 observations' sample

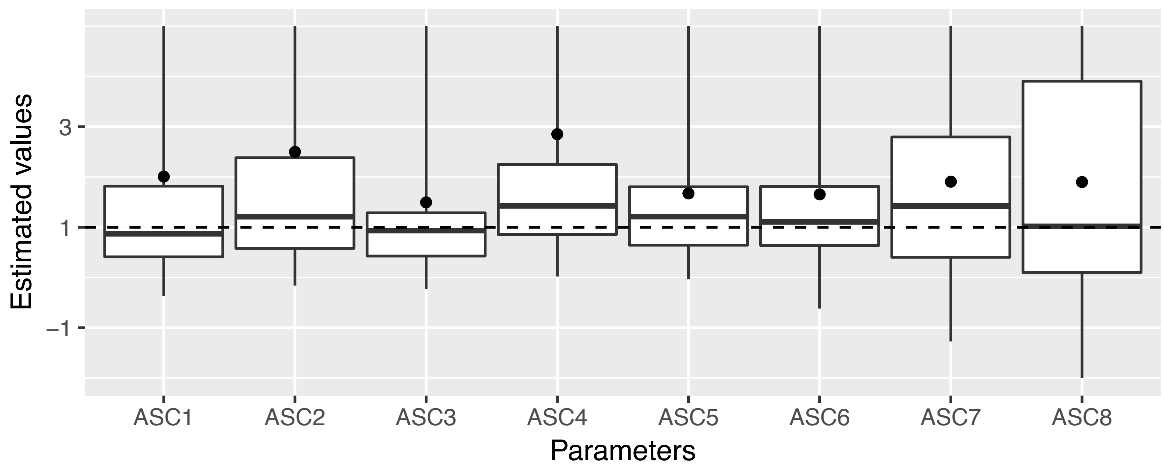


Figure 4-9 Alternative specific constants relative to targets in the 1,000 observations' sample

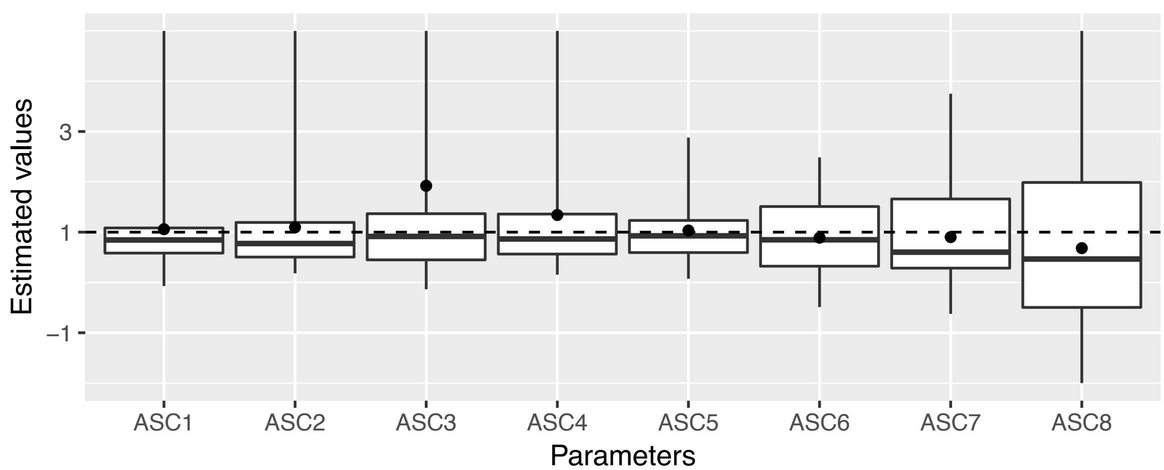
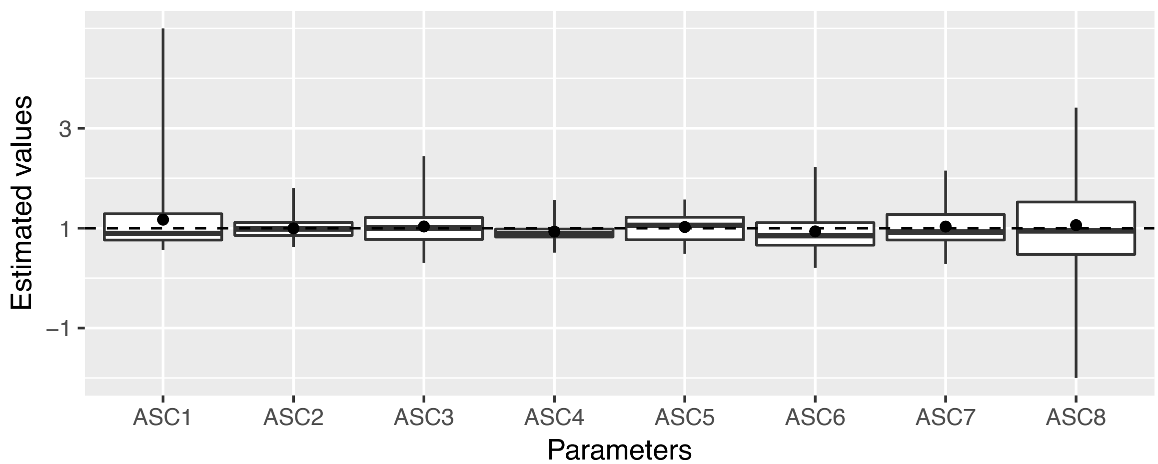


Figure 4-10 Alternative specific constants relative to targets in the 5,000 observations' sample



To compare the difference in performance with a traditional discrete choice model, we also estimated a MNL model for each experiment. The estimation results, shown in Table 4-11, indicate that ignoring the choice heuristic and simply using a RUM model would imply a statistically significant loss of likelihood.

As a conclusion of the synthetic experiment, testing the Stochastic Satisficing model with synthetic data indicates that it is an unbiased and consistent model when applied to frequently used sample sizes. When the nature of the data tends to be of a Satisficing nature, considering a RUM model could imply non-negligible loss of performance. Moreover, we expect higher differences when predicting with the RUM model if the attribute values change, as shown by Williams and Ortúzar (1982a) in their pioneering response analysis work.

Table 4-11 Satisficing and MNL performance in the simulated experiment

Mean log-likelihood

(standard deviation)

Sample size	Satisficing	MNL	Difference
500	-800 (17)	-845 (18)	45
1,000	-1,617 (23)	-1,702 (25)	85
5,000	-8,084 (51)	-8,503 (51)	418

4.2.3.1. Application to real data

We estimated the proposed model using the real choices of the dataset. We modelled DMs choosing their transport mode depending on cost, travel time, walking time and waiting time of each available alternative. All *times* were modelled using the same acceptability function, thus being able to compensate each other at a constant MRS –to be estimated– for the whole data spectrum. Additionally, we estimated alternative specific constants (ASC). Table 4-12 presents the results of the estimates of the Stochastic Satisficing model using this dataset[13].

---

[13] We provide standard deviations rather than t-tests because a standard t-test with respect to zero is meaningless for the threshold attributes.

Table 4-12 Satisficing model estimation results for real data

Parameter	Value (s. deviation)	Parameter	Value (s. deviation)
Cost sensitivity	-0.00474 (1.00)	Cost threshold	-1,850 (10.61)
Time sensitivity	-0.0846 (0.36)	Time threshold	-1,670 (5.77)
MRS travel – walking time	1.29 (0.10)	MRS travel–waiting time	4.05 (0.32)
ASC1	-3.32 (0.09)	ASC2	-5.04 (0.13)
ASC3	-4.63 (0.14)	ASC4 (fixed)	0
ASC5	-2.54 (0.07)	ASC6	-3.16 (0.11)
ASC7	-4.02 (0.16)	ASC8	-3.99 (0.13)
ASC9	-3.59 (0.12)		
Log-likelihood	-1,609		
Sample size	1,374		

The model results are reasonable. First, the signs of the sensitivities are both negative; meaning that the DMs do not like to pay more or to travel more. The marginal rates of substitution of travel time compared with walking time and waiting time have the expected signs and order of magnitude; furthermore, they are statistically different from unity at a 99% confidence level.

In this case, absolute acceptability cannot be obtained since an opt-out alternative is not available; thus, only relative acceptability can be analysed from this model. For example, the model indicates that if an alternative's cost is \$CLP 40 (approximately 0.25 USD) and its cost is raised by 20%, then acceptability decreases by 4%.

MRS are an interesting output in this model; not because there is a welfare implication, but rather because they help to interpret the behaviour of the model. The Stochastic Satisficing model can identify constant or flexible MRS. We modelled times and cost in different acceptability functions. Walking and waiting times are both more onerous than in-vehicle time by 1.29 and 4.05 times respectively at an -imposed– constant rate.

Even though we allowed for a flexible MRS of times and cost, in this sample the marginal rate of substitution was flat and around 18 CLP\$/min for the whole time and cost domains. This result provides evidence that, for this case, the MNL assumption of constant marginal rates of substitution is probably reasonable. To test this hypothesis, we estimated a MNL with a utility function involving the same variables: cost, travel time, walking time, and waiting time. The results in Table 4-13 imply that the logit model outcome is reasonable with almost the same likelihood than the Stochastic Satisficing model.

The marginal rate of substitution between travel time and walking time is 1.4 and between travel time and waiting time is 4.3, which are similar to the Satisficing model MRS. Similarities are also found in the case of the marginal rate of substitution between cost and time, being valued at 18 CLP\$/min.

Table 4-13 Random utility model estimation results for the real data

Parameter	Value (s. deviation)	Parameter	Value (s. deviation)
Cost (x100)	-0.48 (0.17)	Travel time	-0.08 (0.01)
Walking time	-0.11 (0.01)	Waiting time	-0.34 (0.05)
ASC1	-3.04 (0.26)	ASC2	-4.75 (0.25)
ASC3	-4.33 (0.23)	ASC4 (fixed)	0
ASC5	-2.30 (0.20)	ASC6	-2.89 (0.21)
ASC7	-3.73 (0.24)	ASC8	-3.69 (0.20)
ASC9	-3.30 (0.19)		
Log-likelihood	-1,607		
Sample size	1,374		

Note that although both models have similar flat MRS over the whole spectrum of data, this condition is imposed in the MNL rather than estimated as in the case of the Stochastic Satisficing model. Hence, the latter model is structurally more flexible.

4.2.4. Conclusions on the Stochastic Satisficing model

Starting from Satisficing Theory –ST– (Simon, 1955), we have analytically derived a behavioural choice model, where the probability that an alternative is accepted is equal to the joint probability of accepting each attribute. In the presence of an opt-out option, absolute acceptability is obtained, otherwise, only relative acceptability is possible.

Most model properties were also obtained analytically. We discussed identifiability issues and showed the link with the implicit search costs. An analysis of the MRS reveals that if attributes are analysed independently, they can only be compensated if they have a similar degree of acceptability. From this analysis, several features of human behaviour can be explained, such as attribute non-attendance because it is already highly acceptable or valuing more an improvement of an inadequate level attribute that an acceptable one (*akin* to attribute saturation).

We tested the model properties with synthetic data, showing that the model seems unbiased and consistent. The model can estimate flexible MRS. When modelling with real data, the estimated parameters have the correct sign and magnitude, thus giving reasonable predictions. In this case a constant MRS (that must be assumed in classical discrete choice models) was estimated for the whole data spectrum, demonstrating the flexibility of the Stochastic Satisficing model.

The main contribution of this model is the explicit characterization of non-compensatory behaviour as described in the literature. It explains why people could not be influenced by improving higher order attributes if basic ones are not fulfilled.

The model appears as an attractive alternative to traditional discrete choice compensatory models when decision makers find cognitive burden difficult to handle. It can capture extreme behaviour when one attribute is not compensated (e.g. cost for poor people or comfort for the well-off) and leave traditional models to work where they perform best, i.e. when individuals can compensate.

4.2.5. Publication history

This chapter was presented as a paper at the *International Choice Modelling Conference 2017* (Gonzalez-Valdes and Ortúzar, 2017) as well as in seminars in Chile and in the United Kingdom. It was finally published in the *Journal of Choice Modelling* (González-Valdés and Ortúzar, 2018).

5. THEORETICAL ANALYSIS OF IDENTIFIABILITY OF LATENT CLASS MULTIPLE HEURISTIC MODELS

In Section 2.4 we have shown that estimating multiple heuristic discrete choice model is not straightforward. Most studies that use multiple choice heuristics model the class membership function as a constant across the population. The two studies that have reported a more explanatory formulation have required latent variables (Hess and Stathopoulos, 2013) or normalisation between heuristics' sensitivities (Leong and Hensher, 2012b). Our objective is to study the identifiability provided the absence of latent variables and with no normalisations between the choice heuristics parameters.

We develop a theoretical framework that facilitates understanding the identifiability of multiple heuristics discrete choice models. We start by analysing a binary case, where the simple structure illuminates the underlying phenomena. Then, we generalise by analysing a multi-heuristics case. In each of the analyses, we establish the first-order optimality conditions on the likelihood function to understand if and when several coexisting choice heuristics can be identifiable. Finally, the hessian matrix of the likelihood function is analysed to relate the choice heuristics and their empirical identifiability.

5.1. Binary Case

5.1.1. The balance of choice heuristics

Suppose that two choice heuristics, denoted as a and b, are followed by DMs with probability π_a and $(1-\pi_a)$ respectively. Let $P_{hqi}(\theta)$ be the probability that individual q chooses alternative i following heuristic h using parameters θ. Then, $P_{qi}(\theta)$, the probability of choosing alternative i under a latent class model, is given by (5.1).

$$P_{qi}(\theta, \pi_a) = \pi_a P_{aqi}(\theta) + (1-\pi_a) P_{bqi}(\theta) \tag{5.1}$$

The log-likelihood function of this model with set (θ, π_a) is given by (5.2), where $P_{hq*}(\theta)$ represents the probability that q would have chosen the selected alternative under heuristic h:

$$l(\theta, \pi_a) = \sum_q \log\left(\pi_a\, P_{aq*}(\theta) + (1-\pi_a)\, P_{bq*}(\theta)\right) \tag{5.2}$$

The likelihood can be maximised either at a boundary or in an interior solution. In the case of a boundary solution, i.e. $\pi_a \in \{0,1\}$, it is optimal for the model to consist of a single heuristic. Whereas in the case of an interior solution, the two choice heuristics will coexist in the model.

The solution depends on the balance between the losses and gains in likelihood associated with including an additional heuristic and, therefore, reducing the proportion of the original one. There may be individual observations where the initial heuristic performs better than the additional one, whilst others where the latter performs better. Including the alternative heuristic should provide an improvement in likelihood for the observations where it performs better than the former. However, in the cases where the former heuristic performs better, there should be a loss of likelihood due to the reduction of its initial proportion in favour of including the additional one. The balance between these two performances determines the type of solution obtained (i.e. whether the solution is a boundary or interior one). Both cases are illustrated in Section 5.1.2.

In the more interesting case of an interior solution, the point of maximum likelihood is obtained when the likelihood function is stationary with respect to variations in the class membership probability π_a. This can be detected as an interior point at which the derivative of the log-likelihood function equals zero. Among the variables to analyse, an interesting one is π_a, since it determines the proportion of individuals following each choice heuristic and, therefore, connects them in the model. The first order condition regarding π_a is examined next.

We start by considering the case where the class membership function π_a is constant across the population (i.e. every individual chooses among heuristics with the same probability). This basic case is the most frequent formulation reported in the literature (Adamowicz & Swait, 2013; Araña et al., 2008; Balbontin et al., 2017; Hess et al., 2012; McNair et al., 2012). Under this specification the following theorem describes the coexistence of choice heuristics:

THEOREM 1*: Two choice heuristic coexist optimally in a discrete choice model with constant heuristic probabilities if the vector of estimated parameters satisfies the balance specified by (5.3):*

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta)} = \sum_q \frac{P_{bq*}(\theta)}{P_{q*}(\theta)} \tag{5.3}$$

where $P_{q*}(\theta) = \pi_a \, P_{aq*}(\theta) + (1 - \pi_a) \, P_{bq*}(\theta)$ represents the probability that individual q would have chosen the selected alternative according to the combined model

PROOF: For an interior solution, the first order condition for the maximisation problem is given by (5.4):

$$\frac{\partial l(\theta, \pi_a)}{\partial \pi_a} = \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{\pi_a \, P_{aq*}(\theta) + (1 - \pi_a) \, P_{bq*}(\theta)} = 0 \tag{5.4}$$

Manipulation of (5.4) leads to (5.5):

$$\sum_q \frac{P_{aq*}(\theta)}{\pi_a \, P_{aq*}(\theta) + (1 - \pi_a) \, P_{bq*}(\theta)} = \sum_q \frac{P_{bq*}(\theta)}{\pi_a \, P_{aq*}(\theta) + (1 - \pi_a) \, P_{bq*}(\theta)} \tag{5.5}$$

Using the definition of $P_{q*}(\theta)$, this is equivalent to (5.3).

Expressions (5.3) and (5.5) indicate that when both choice heuristics are present in the model, there is a balance between them. When present, this balance indicates that the gain in likelihood due to the inclusion of an alternative heuristic surpasses the losses in likelihood due to the decrease in the former heuristic until balance is obtained.

The optimal proportion of choice heuristics (i.e. the proportion that achieves maximum likelihood) is given by their relative performance. The higher the performance of a heuristic, higher is its optimal proportion. If the performance difference is too important, then, no balance could be optimal.

A second force that plays a role in this balance is the proportion in which the heuristics are present in the data generating process. If this process is aligned with a higher proportion of one choice heuristic, then its performance in the overall sample will be better and its optimal proportion will increase. Different proportions may favour the balance if the less effective heuristic dominates the sample or may deter it if it promotes the more effective heuristic. Therefore, if the most effective heuristic also dominates the sample, it may be sub-optimal to include an alternative heuristic and no balance will be achieved.

Under the conditions of a constant class membership function, the balance of (5.5) has a known value described by Theorem 2:

THEOREM 2*: Two choice heuristic coexist optimally in a discrete choice model with constant heuristic probabilities if the balance quantity in (5.3) is equal to the sample size Q.*

PROOF: Expanding the left-hand side of (5.3) gives (5.6):

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{\pi_a\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)}$$

$$= \sum_q \frac{\pi_a \, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} + \sum_q \frac{(1-\pi_a)\, P_{bq*}(\theta) - (1-\pi_a)\, P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)}$$

$$\Rightarrow \sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{\pi_a \, P_{aq*}(\theta) + (1-\pi_a)\, P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} + (1-\pi_a) \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} \tag{5.6}$$

In the first summation of the right-hand side of (5.6), every term is identically equal to one, therefore the summation adds to Q. The second summation is equal to zero because of stationarity (5.4). Because of (5.3) and in light of the symmetry between both choice heuristics, the condition corresponding to heuristic a applies equally to heuristic b. Then, (5.7) describes the balance in a model with two choice heuristics and constant heuristic probabilities:

$$\sum_q \frac{P_{aq*}(\theta)}{P_{q*}(\theta,\pi_a)} = \sum_q \frac{P_{bq*}(\theta)}{P_{q*}(\theta,\pi_a)} = Q \tag{5.7}$$

Examples of this balance are given in Section 5.1.2. The balance is broken (i.e. the optimal model contains only one choice heuristic), when it is optimal to not include any amount of the other heuristic, as previously explained. A diagnostic condition for this is presented in (5.8) and (5,9) for the case of a model that includes heuristic b alone:

$$\pi_a^* = 0 \Leftrightarrow \left. \frac{\partial l(\theta,\pi_a)}{\partial \pi_a} \right|_{\pi_a = 0} = \sum_q \frac{P_{aq*}(\theta) - P_{bq*}(\theta)}{\pi_a \, P_{aq*}(\theta) + (1-\pi_a)\, P_{bq*}(\theta)} < 0 \tag{5.7}$$

$$\pi_a^* = 0 \Leftrightarrow \sum_q \frac{P_{aq*}(\theta)}{P_{bq*}(\theta)} < Q \tag{5.9}$$

The optimality of a single choice heuristic can be identified using (5.9) or its counterpart for heuristic a alone. For this to occur, the prevalent heuristic must perform well, even when DMs actually choose using the other heuristic; otherwise, the loss for not considering an alternative heuristic would be too high. Conversely, a plausible way for a balanced interior combination of heuristics to be optimal is that (5.10a) holds for some observations and (5.10b) for others:

$$\frac{P_{aq*}(\theta)}{P_{bq*}(\theta)} \gg 1 \tag{5.10a}$$

$$\frac{P_{bq*}(\theta)}{P_{aq*}(\theta)} \gg 1 \tag{5.10b}$$

If the class membership function π_a is not constant but some function $\pi_a(\theta)$, the balance is stated in the following theorem:

THEOREM 3: *Two choice heuristic coexist optimally in a discrete choice model if the vector of estimated parameters satisfies the balance specified by (5.11):*

$$\sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{aq*}(\theta) + \frac{\partial P_{aq*}(\theta)}{\partial \theta} \pi_a(\theta)}{P_{q*}(\theta)} = \sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{bq*}(\theta) + \frac{P_{bq*}(\theta)}{\partial \theta} (\pi_a(\theta) - 1)}{P_{q*}(\theta)} \tag{5.11}$$

PROOF: (5.12) states the stationarity condition required for optimality.

$$0 = \frac{\partial l(\theta)}{\partial \theta}$$
$$= \sum_q \frac{\frac{\partial \pi_a(\theta)}{\partial \theta} P_{aq*}(\theta) + \pi_a(\theta) \frac{\partial P_{aq*}(\theta)}{\partial \theta} - \frac{\partial \pi_a(\theta)}{\partial \theta} P_{bq*}(\theta) + \left(1 - \pi_a(\theta)\right) \frac{\partial P_{bq*}(\theta)}{\partial \theta}}{\pi_a(\theta) P_{aq*}(\theta) + \left(1 - \pi_a(\theta)\right) P_{bq*}(\theta)} \tag{5.12}$$

Expression (5.11) is a direct rearrangement of (5.12), where the balance between choice heuristics is stated.

Suppose that the set of parameters β of the class membership function is disjoint from the set θ affecting the choice heuristics. Then, the following corollary follows from Theorem 3:

COROLLARY: *If the class membership function is independent from the choice heuristics, the balance is given by (5.13):*

$$\sum_q \frac{\frac{\partial \pi_a(\beta)}{\partial \beta} P_{aq*}(\theta)}{P_{q*}(\theta, \beta)} = \sum_q \frac{\frac{\partial \pi_a(\beta)}{\partial \beta} P_{bq*}(\theta)}{P_{q*}(\theta, \beta)} \tag{5.13}$$

This analysis identifies when it is optimal for the model to include more than one choice heuristic. Nevertheless, the coexistence of choice heuristics does not guarantee that the model is identifiable; it only guarantees that the optimal point contains multiple heuristics. However, this optimal point might not be unique nor empirically identifiable. This empirical identifiability issue is addressed next.

5.1.2. Examples of balance of heuristics

We present three small examples of the balance of choice heuristics. In each example the DMs choose six times; three times with heuristic a and three times with heuristic b. In all examples individuals choose the same alternatives as in Table 5-1.

Table 5-1 Chosen heuristic and alternatives in the balance examples

Chosen heuristic	Chosen alternative
1	1
1	2
1	2
2	1
2	1
2	2

For this example and for simplicity, we assume that the choice heuristics are correctly identified but the heuristics used by the DMs are unknown. Therefore, we estimate a multiple heuristic model with only one unknown parameter, the class membership π_a as in (5.14).

$$P_{qi}(\pi_a) = \pi_a P_{aqi} + (1 - \pi_a) P_{bqi} \tag{5.14}$$

The multiple heuristic model is estimated via maximum likelihood to obtain the population split that maximises the model's likelihood.

The first four columns in Tables 5-2, 5-3 and 5-4 show the probabilities of choosing each alternative when following each choice heuristic. By changing the probabilities of heuristic b (third and fourth columns), we manipulate the point of maximum likelihood shown in the fifth column. The sixth column shows the probability of the multiple heuristic model which takes as input the probability π_a and the probabilities of choosing each alternative conditional on the choice heuristic. Finally, the last column shows the ratio of the probability that each heuristic assigns to the chosen alternative and the probability that the multiple heuristic model assigns to the chosen alternative.

Table 5-2 Multiple heuristic model example with strong balance

Heuristic a		Heuristic b		Prob. of following heuristic a	Probability assigned to the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}
0.50	0.50	0.35	0.65	0.31	0.40	1.26	0.88
0.50	0.50	0.60	0.40	0.31	0.43	1.16	0.93
0.50	0.50	0.70	0.30	0.31	0.36	1.38	0.83
0.50	0.50	0.80	0.20	0.31	0.71	0.71	1.13
0.50	0.50	0.80	0.20	0.31	0.71	0.71	1.13
0.50	0.50	0.30	0.70	0.31	0.64	0.78	1.10
					Sum	6	6

Table 5-2 shows an example where a balance of choice heuristic exists. The optimal class membership function indicates that the probability of following heuristic a is 0.31. Therefore, there is a strong balance across the two choice heuristics. Also note the balance given by the sum of the ratios of the heuristic and the model: as stated in Theorem 2, its value equals the sample size.

In the second example, shown in Table 5-3, one of the probabilities –which is underlined– is changed, improving the performance of heuristic b. In this example, the balance still exists but the model's estimated probability of following heuristic a decreases. Because the balance still exists, Theorem 2 holds, showing that the sum of the ratios of the choice heuristic and the models equals the sample size.

Table 5-3 Multiple heuristic model example with weak balance

Heuristic a		Heuristic b		Prob. of following heuristic a	Probability assigned to the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}
0.50	0.50	0.35	0.65	0.04	0.36	1.40	0.98
0.50	0.50	0.60	0.40	0.04	0.40	1.24	0.99
0.50	0.50	<u>0.64</u>	<u>0.36</u>	0.04	0.37	1.37	0.98
0.50	0.50	0.80	0.20	0.04	0.79	0.63	1.02
0.50	0.50	0.80	0.20	0.04	0.79	0.63	1.02
0.50	0.50	0.30	0.70	0.04	0.69	0.72	1.01
					Sum	<u>6</u>	<u>6</u>

Finally, the third example (Table 5-4) presents a model for which the optimal point is a single choice heuristic. Even though heuristic a performs better than heuristic b when predicting choices made by following the former, the loss of performance of the last three choices due to the inclusion of heuristic a outweighs the benefit of its inclusion. Therefore, even though a different heuristic is present in the underlying choice mechanism, it is optimal not to include it in the choice model. Finally, note that in this case the balance is broken the probability of the total model is identical to the probability of the identified heuristic, therefore only the identified heuristic ratio sums to the sample size.

Table 5-4 Multiple heuristic model example with no balance

Heuristic a		Heuristic b		Prob. of following heuristic a	Probability assigned to the chosen alternative		
Alt 1	Alt 2	Alt 1	Alt 2	π_a	P_{q*}	P_{aq*}/P_{q*}	P_{bq*}
0.50	0.50	0.35	0.65	0	0.35	1.43	1
0.50	0.50	0.60	0.40	0	0.40	1.25	1
0.50	0.50	0.60	0.40	0	0.40	1.25	1
0.50	0.50	0.80	0.20	0	0.80	0.63	1
0.50	0.50	0.80	0.20	0	0.80	0.63	1
0.50	0.50	0.30	0.70	0	0.70	0.71	1
					Sum	5.89	6

These results indicate that the balance can be fragile; however, we expect less fragility as the sample grows. Nonetheless, these results exemplify that even though the underlying process may contain several choice heuristics, a balance among them might not be achieved in estimation.

5.1.3. Behavioural diversity of choice heuristics and identifiability

To study the identifiability of a multiple heuristics model, that is, when the estimator can be identified uniquely without any parameter set being observationally equivalent (Hsiao, 1983; Matzkin, 2007), we assume that the model has an interior solution. If the model had a boundary solution (i.e. only one heuristic is estimated), then the empirical identifiability analysis of multiple choice heuristics would not be relevant.

For a parametric model to be identifiable, the information matrix (5.15) must be non-singular (Rothenberg, 1971). Moreover, it is desirable that the covariance matrix is reasonably small, which we refer as *strong identifiability*. The covariance matrix is related to the model via the Fisher information matrix (5.16). Thus, to obtain strong identifiability, the information matrix should have a large determinant so that the covariance matrix exhibits small values.

$$F = -\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial\theta_x \partial\theta_y}\right) \tag{5.15}$$

$$\Sigma = F^{-1} \tag{5.16}$$

Similar to the analysis of the first order condition for the two-heuristic case, we analyse the information matrix at the point determined by π_a. The diagonal element of the information matrix corresponding to π_a is given by the derivative of (5.4) with respect to π_a, as in (5.17):

$$\frac{\partial^2 l(\theta)}{\partial \pi_a^2} = -\sum_q \frac{\left(P_{aq*} - P_{bq*}\right)^2}{P_{q*}^2} \tag{5.17}$$

For F to have a large determinant, and thus for the standard errors of the estimator to be small, it is necessary for the expression given by (5.17) to be large. Two elements play an important role here, the sample size and the numerator. As the sample size increases the magnitude of the summation also increases, which promotes identifiability. Regarding the numerator, note that the maximum likelihood estimates are obtained when the probability P_{q*}^2 is maximum; then, identifiability is determined by the numerator of (5.17). Thus, expression $\left(P_{aq*} - P_{bq*}\right)^2$ is an important element in the identification of choice heuristics. High values of this expression are obtained when the choice heuristics exhibit disparate behaviour. Thus, that absence of substantial behavioural diversity between the two heuristics may cause identifiability problems. Therefore, this behavioural diversity requires not only different functional forms but should also be reflected in the data.

5.2. Multiple Heuristics Case

Now consider the general case where several choice heuristics are used by DMs. We start by analysing the first order conditions to generalise the balance obtained in Section 5.1. Then, the analysis of identifiability is extended to this multiple heuristic case.

Extending the notation of Section 5.1, let π_h be the probability that DMs behave according to heuristic $h \in H$ so that $\sum_{h \in H} \pi_h = 1$. Then, the log-likelihood function $l\ (\pi,\ \theta)$ of the model is given by (5.18):

$$l(\pi,\ \theta) = \sum_{q} \log\left(\sum_{h \in H} \pi_h P_{hq*}(\theta) \right) \tag{5.18}$$

To maximise the likelihood of the model subject to the sum constraint on the population probabilities π_h, we seek stationary points of the Lagrangean (5.19):

$$L = -l(\pi, \theta) - \lambda \left(1 - \sum_{h \in H} \pi_h \right) \tag{5.19}$$

Differentiating the Lagrangean with respect to π_a and equating it to 0 for stationarity gives the necessary condition for recovery of the optimal probabilities π_a (5.20):

$$\begin{gathered} \frac{\partial L}{\partial \pi_a} = 0 \Leftrightarrow \sum_{q} \frac{P_{aq*}}{\sum_{h \in H} \pi_h P_{hq*}(\theta)} = \lambda \\ \Rightarrow \sum_{q} \frac{P_{aq*}}{P_{q*}} = \lambda \quad \forall a \in H \end{gathered} \tag{5.20}$$

According to (5.20), stationarity is achieved when each choice heuristic $h \in H$ contributes the same aggregated ratio P_{hq*}/P_{q*} for the alternatives chosen. This result extends the balance exposed in section 5.1 to multiple heuristics.

Expression (5.20) shows the balance condition for the optimal point, but again does not guarantee the identifiability of the choice heuristics. For the vector $\boldsymbol{\pi}$ of different choice mechanism probabilities to be identifiable, the information matrix should be non-singular and, therefore, the hessian matrix of the Lagrangean should be positive definite. This requires that all principal submatrices of the hessian (that correspond to the second derivatives with respect to the proportions) should have positive determinants. The mixed second partial derivatives of the Lagrangean are stated in (5.21).

$$\frac{\partial^2 L}{\partial \pi_a \partial \pi_b} = \sum_q \frac{P_{aq*}P_{bq*}}{\left(\sum_{h\in H} \pi_h P_{hq*}(\theta)\right)^2} = \sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2} \tag{5.21}$$

Therefore, each 2×2 submatrix of this kind has the structure shown in (5.22).

$$\begin{bmatrix} \sum_q \frac{P_{aq*}^2}{P_{q*}^2} & \sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2} \\ \sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2} & \sum_q \frac{P_{bq*}^2}{P_{q*}^2} \end{bmatrix} \tag{5.22}$$

Because the elements on the principal diagonal are positive, the submatrix is positive definite if the determinant exceeds zero. Moreover, the determinant *D* given by (5.23) needs to be large so that the covariance matrix of the estimators is small.

$$D = \sum_{q\in Q} \frac{P_{aq*}^2}{P_{q*}^2} \sum_{r\in Q} \frac{P_{br*}^2}{P_{r*}^2} - \left(\sum_q \frac{P_{aq*}P_{bq*}}{P_{q*}^2}\right)^2 \tag{5.23}$$

Before analysing (5.23) to determine when D will be positive, note that this analysis is useful in the case that a balance exists between choice heuristics. In that case, we cannot have $P_{aq*} = P_{bq*}\,\forall q$. Therefore, there will be a proportion of outcomes where heuristic a outperforms the aggregate model and another proportion where its performance will be worse. The expressions $P_{hq*}^2/P_{q*}^2 \;\; h \in H$ tend to amplify the difference when one model outperforms the other. Provided that each heuristic outperforms simultaneously the aggregate model and every other heuristic for some observations, then every determinant D of the form given by (5.23) will be positive.

For a convenient analysis of (5.23), we introduce some notation for the moments of the estimated conditional probabilities $P_{hq*}/P_{q*} \;\; h \epsilon H$. Thus, let the first and second moments be respectively:

$$\mu_h = \mathbb{E}\left(\frac{P_{hq*}}{P_{q*}}\right), h \in H$$

$$\sigma_h^2 = Var\left(\frac{P_{hq*}}{P_{q*}}\right) \;\; h \in H \;\; and \;\; \sigma_{ab} = Cov\left(\frac{P_{aq*}}{P_{q*}}, \frac{P_{bq*}}{P_{q*}}\right) \quad a, b \in H.$$

With this notation, the expectation of elements involved in (5.23) can be written as:

$$\mathbb{E}\left(\sum_{q \in Q} \frac{P_{hq*}^2}{P_{q*}^2}\right) \approx n(\mu_h^2 + \sigma_h^2) \;\text{ and }\; \mathbb{E}\left(\sum_{q \in Q} \frac{P_{aq*}P_{bq*}}{P_{q*}^2}\right) \approx n(\mu_a\mu_b + \sigma_{ab})$$

Therefore, the expectation of (5.23) can be rearranged to express a sample estimate of the population quantity as (5.24) where the approximation arises from finite sample estimation of the moments.

$$\frac{1}{n^2}\mathbb{E}(D) \approx \mu_a^2\mu_b^2\left(\frac{\sigma_a^2}{\mu_a^2} - 2\frac{\sigma_{ab}}{\mu_a\mu_b} + \frac{\sigma_b^2}{\mu_b^2}\right) + \sigma_a^2\sigma_b^2\left(1 - \frac{\sigma_{ab}^2}{\sigma_a^2\sigma_b^2}\right) \tag{5.24}$$

Recall that from condition (5.20) for both heuristics a and b to be present in the model, $\mu_a = \mu_b$. If choice probabilities are perfectly correlated, $\sigma_{ab}^2 = \sigma_a^2\sigma_b^2$, the right-hand side of (5.24) will be identically zero showing that the hessian matrix would be singular in expectation. The expectation of the partial derivative of D with respect to the correlation σ_{ab} in (5.24) is negative, so the expectation of the determinant increases as the correlation decreases. In particular,

$$\mathbb{E}\left(\frac{dD}{d\sigma_{ab}}\right) \approx -2n^2(\mu_a\mu_b + \sigma_{ab}) = -2n^2\mathbb{E}\left(\frac{P_{aq*}P_{bq*}}{P_{q*}^2}\right) \leq 0\,. \tag{5.25}$$

Thus, estimation of the mixed model is better conditioned (as indicated by larger values of D) when correlation σ_{ab} is reduced and as sample size n increases.

The requirement for positive determinants of the principal submatrices of the hessian, therefore, generalises the requirement for the binary heuristic case presented in Section 5.1. To be identifiable, the behaviour of each heuristic should differ from that of all other heuristics; the greater the behavioural difference is, the larger the determinant of (5.23) and hence the smaller the covariance matrix of the estimators. In conclusion, a natural requirement for a mixture of choice heuristics to be identifiable is that each of them performs best for some of the cases in the data used for estimation.

5.3. Conclusions on the Theoretical Analysis of Identifiability

To analyse the identifiability of multiple heuristic models, we developed a theoretical framework to analyse when a mixed model of this kind is identifiable. We established two analytical conditions for this: first, a balance must exist between choice heuristics and second, the behaviour of the heuristics must differ sufficiently so that they can be identified with an acceptable accuracy of parameters.

The balance may not exist even if the underlying data generating process reveals a combination of two processes. The proportion of each process and the performance of each heuristic interpreting each of the data generating processes determines the existence (or not) of a balance. Higher proportion heuristic are more plausible to dominate the balance. Indeed, the dominant choice heuristic must perform poorly on some choices aligned with the other heuristic so that the model is able to incorporate two choice heuristics.

Finally, our analysis concludes that a necessary condition for the estimation to show a reasonable small covariance matrix is that the choice heuristics differ in their behaviour. Therefore, to identify several choice heuristics, the context tested must exhibit choice sets that let the heuristics perform disparately in some observations.

5.4. Publication History

Initial analysis of this chapter was presented at the *6th Symposium of the European Association for European Research in Transportation* (Gonzalez-Valdes and Raveau, 2017a). Further versions were presented in seminars in the United Kingdom and in the *18th Chilean Transport Engineering Conference* (Gonzalez-Valdes et al., 2017). Finally, a full developed paper is under revision in *Transportation Research Part B: Methodological* (Gonzalez-Valdes et al., 2018).

6. EMPIRICAL IDENTIFICATION OF CHOICE HEURISTICS AT SAMPLE LEVEL

In Section 2.4 we showed that estimating multiple heuristic discrete choice model is not straightforward. Most of the previous studies on this subject model the class membership function as a constant across the population. Moreover some studies use extremely simple heuristics (Araña et al., 2008), which probably capture heterogeneity in tastes rather than heterogeneity in rules. In the same spirit, some authors include attribute non-attendance as a heuristic *per se* (Leong, 2014; Balbontin et al., 2017), which captures different sensitivities in the RUM framework. Therefore, only few studies identify choice heuristics by their different behaviour exclusively.

In Section 2.4 and Chapter 5 we argue that no study has been able to identify non-constant population wide class membership functions without the need of normalisations or latent variables (Leong and Hensher, 2012b; Hess and Stathopoulos, 2013). On Chapter 5 we showed that the multiple choice heuristic model may exhibit a balance of two choice heuristic coexisting even when the class membership function is not a constant. For this, the context must be able to provide a scenario where the choice heuristics perform disparately.

In this chapter we study several elements that affect the identifiability of choice heuristics in the context of a pseudo-real mode choice context. First, we study the identifiability of pairs of choice heuristics changing: the type of heuristic, degree of correlation, proportion of each choice heuristic, sample size, and alternatives per choice set. Then, we examine the identifiability in the case of three heuristics. In all cases, we consider the same attributes affecting the different heuristics so that identifiability is obtained exclusively due to their different behaviour.

6.1. Empirical Identifiability of Two Heuristics Models

To guarantee the presence of different choice heuristic and to exert control over the choice parameters, a synthetic population was generated. We studied four dimensions affecting the choice process: the type of choice heuristic, the proportion of each choice heuristic in the synthetic sample, the correlation between the parameters of the probability of using each choice heuristic and the sensitivities for different attributes of the alternatives, and the number of alternatives per choice set. Finally, for each of these three dimensions, ten experiments were performed.

Among the elements we considered, first we analysed the degree of identifiability of each heuristic in the models. Then, we selected some interesting estimations cases and exposed the bias of the model's estimated parameters. We ended up by concluding how the popular heuristics tested differed in their degree of identifiability.

6.1.1. Experimental design

Among the dimensions considered, the first one is the type of choice heuristic. The analysis of Chapter 5 indicated that the difference between choice heuristics is key to their identification. Three different choice heuristics were tested against the most widely used RUM heuristic to investigate whether they could be identified in a practical context, namely: Elimination By Aspects –EBA (Tversky, 1972a; 1972b), Stochastic Satisficing –SS (González-Valdés and Ortúzar, 2018) and Random Regret Minimization –RRM (Chorus et al., 2008).

The second dimension considered was the proportion of each choice heuristic in the sample. The analysis in Chapter 5 indicated that the greater the proportion of a choice heuristic, the greater the number of observations for which it will outperform the other heuristics, thus increasing its identifiability. Two choice proportions were tested: approximated 70% of the sample chooses according to RUM and 30% according to the other heuristic, and *vice versa*, that is $\pi_a \in \{0.3, 0.7\}$. These proportions were designed so that there are cases in which one heuristic dominates the sample but without monopolizing the importance in the model.

The third dimension considered was the correlation between the choice parameters and the probability of selecting a heuristic. The purpose here was to analyse how any such correlation would bring complexities to the identification task. This correlation is added by introducing a personal trait that affects both the probability of using a choice heuristic and the RUM preferences.

Finally, the fourth dimension tested was the number of alternatives in the choice sets. We discuss the degree of identifiability from RUM under different number of heuristics. To adapt the number of alternatives we used the methodology exposed in Subsection 3.1.2.

Because the focus of this study is to work with models that have identifiability issues, we preferred Bayesian estimation over other maximum likelihood estimation to avoid being captured at local optima. The Bayesian estimation was performed using Gibbs sampling with the JAGS package (Plummer, 2016) for the R software system (R Core Team, 2016) .

Previous tests suggest that this estimation procedure usually requires numerous iterations to achieve stationarity. Five thousand burn-in samples were discarded before sampling from the Markov chain. Moreover, a large number of samples were required to sample the posterior distribution of the parameters broadly. Therefore, for each parameter, ten thousand useful samples were obtained after burn-in. Regarding the prior distributions, low precision zero-centred Normal priors were used.

We used a simulated dataset to investigate whether it is possible to capture different choice heuristics in a practical transport context. For estimation we required two components: a set of choice alternatives available to each individual and the individuals' choices. The choice sets for each individual were extracted from the Las Condes – CBD, San Miguel – CBD database explained in Section 3.1.

6.1.1.1. The choice heuristics

To incorporate sociodemographic characteristics and having control of such characteristic, a binary variable z was generated (simply named *trait*) with probability of 0.70. This probability was designed to represent an interesting trait to be studied in the population. Each simulated individual was also assigned, independently, to use one of two available choice heuristics: RUM and the contrasting one (i.e., EBA, RRM or SS). In each case, the probability of using RUM was given by the inverse logit function (6.1) with parameters as shown in Table 6-1. These were calculated to give a probability of 0.71 (as indicated in the previous section) for choosing RUM in one experiment and the same probability for adopting the contrasting heuristic in another.

$$\pi_{RUM} = \frac{\exp(\theta_0 + \theta_1 z)}{1 + \exp\,(\theta_0 + \theta_1 \mathrm{z})} \tag{6.1}$$

Table 6-1 Synthetic population latent class parameters

Parameter	Value
θ_0	0
θ_1	+/- 1.39

Once the individual chooses the heuristic, he will select the alternative using the parameters given in Table 6-2. The EBA heuristic employs the same formulation used in Section 4.1 and further detailed in Appendix B. We considered different thresholds for travel time (15 min), waiting time (5 min), and walking time (3 min). For cost, we considered two thresholds set at USD 0.25 and USD 0.65; these two thresholds split the costs into three aspects levels, where two of them are desirable. Finally, an alternative specific aspect was created for each alternative. The SS heuristic considers three acceptability functions one for cost, one for all the time components, and one for the alternative specific constant. The model also estimates the constant marginal rate of substitution between travel, waiting, and walking time.

Table 6-2 Choice heuristic simulation parameters

Parameter	EBA	RRM	SS	RUM
Cost sensitivity	1.39; 1.39	0.375	-6.25	-0.31; +0.09
SS cost threshold	-	-	0.28	-
Vehicle time sensitivity	1.39	2	-12	-5
Waiting time sensitivity	2.08	10	4	-20
Walking time sensitivity	2.30	4	1.5	-6.5
SS time threshold	-	-	0.60	-
μ	-	0.2	-	-
ASC1	0.41	0.1	-0.84	0.5
ASC2	0	0	0	0
ASC3	0.10	0.02	-0.96	0.1
ASC4	0.59	0.16	-0.77	0.8
ASC5	0.53	0.14	-0.80	0.7
ASC6	0.47	0.12	-0.82	0.6
ASC7	0.18	0.04	-0.93	0.2
ASC8	0.26	0.06	-0.90	0.3
ASC9	0.34	0.08	-0.87	0.4

With the objective of increasing the difference between RRM and RUM, we selected the $\mu - RRM$ version to increase the profundity of regret compared to the simplest version (both detailed in subsection 2.2.3). The μ parameter was fixed at 0.2 so that the regret was highly increased.

The RUM utility function was considered linear and additive in parameters. As detailed before, in some of the experiments, the cost attribute was modified based on the individual's sociodemographic *trait*. If DMs had it (indicated by $z = 1$), the sensitivity to cost was modified; we call this attribute "cost difference" of sensitivity.

6.1.2. Analysis of results

Typically, a model is non-identifiable if the information matrix is singular; this is equivalent to having an infinite element within the covariance matrix. In our context with Bayesian estimation, no matrix inversion is required; nonetheless, model non-identifiability can be detected when the standard deviations of the parameters are extreme with associated instability of the Markov chain.

Even though we have described what identifiability is, we detected different degrees of non-identifiability. Therefore, to better describe these, we developed three further descriptions:

- *Strong identifiability*: all parameters of the model are estimated with acceptable standard deviations. Both choice heuristics are identified, thus there is a balance between them.
- *Weak identifiability*: a small proportion of the model parameters are estimated with extreme standard deviations. Nevertheless, the model can clearly identify the presence of two choice heuristics and identify the variables governing their selection.
- *Non-identifiability*: most parameters are estimated with extreme standard deviation or no balance can be found between choice heuristics.

In Chapter 5 we analysed how behavioural differences impacted the identifiability of different choice heuristics. Figure 6-1 shows the distribution of behavioural differences between RUM and the other choice heuristics among the alternatives in the dataset. This difference is quantified by the absolute difference between the probabilities given by two choice heuristics. For example, if two heuristics a and b estimate probabilities P_{ai} and P_{bi} of choosing alternative i , then the difference is calculated as $|P_{ai} - P_{bi}|$.

Figure 6-1 Behavioural difference of RUM and RRM, SS, and EBA

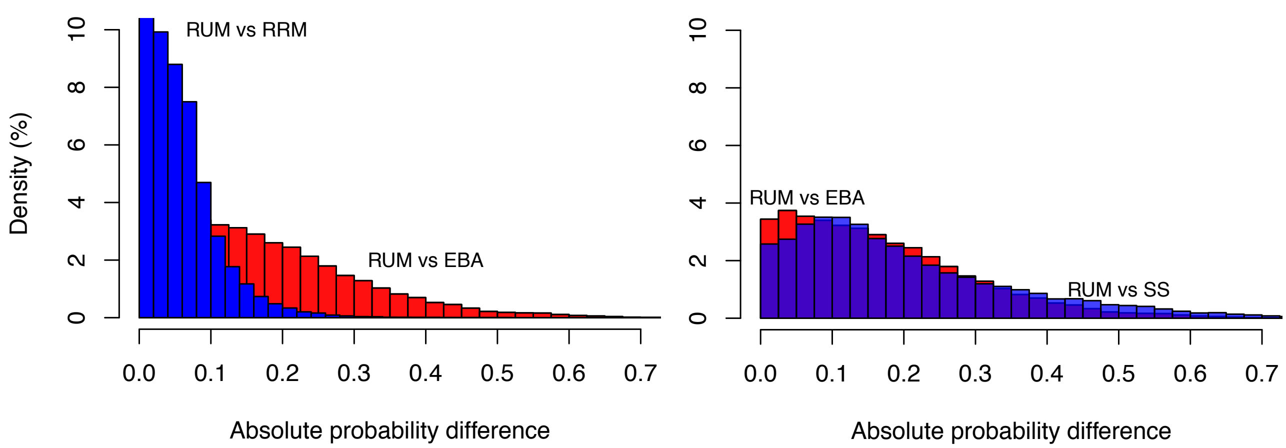


Figure 6-1 shows that among the choice heuristics, the RRM choice mechanism differs least from RUM. Thus, we expect that RRM to be the choice heuristic with the least probability of balance with RUM. Conversely, SS and EBA present important behavioural differences from RUM. Note however, that because this analyses only one dimension of the information matrix, it is useful for generating hypotheses but cannot guarantee them.

We analysed each pair of choice heuristics separately according to the three degrees of identifiability. We also analysed separately the two proportions of choice heuristics and each of the correlation cases below.

6.1.2.1. Analysis of the RUM and RRM case

Analysis of identifiability

Table 6-3 shows the results of the identifiability analysis for each of the 40 estimations among the four categories of correlation and proportions of choice heuristics in this case. When RUM dominates the sample, in most cases, RUM was the only choice heuristic identified with no balance between RUM and RRM in the models estimated. However, there were three cases where RUM was the main heuristic but RRM alone was identified. When positive correlation between the class membership function and the RUM's sensitivities was introduced, surprisingly, the number of cases where RUM was identified increased from 8 to 9. Therefore, in the overall no balance between RUM and RRM was detected.

In the case that RRM dominated the sample, its identifiability increased although the RRM model was less identifiable than the RUM model when the latter dominated. When no correlation was present between the class membership function and the parameters of the RUM heuristic, RRM was identified in seven of the ten cases. Nevertheless, in a high proportion of the cases where RRM was identified, it was identified weakly, with a few parameters having extreme variance. When the correlation between the class membership function and the RUM heuristic was greater, the identifiability of the RRM increased; this may be as being due to increased difficulty in identifying the RUM class.

Table 6-3 Identifiability results of RUM and RRM models (number of cases)

Correlation	RUM dominates (70%)	RRM dominates (70%)
No correlation	$\frac{8}{10}$ identifies RUM only $\frac{2}{10}$ identifies RRM only **No balance detected**	$\frac{3}{10}$ identifies RUM only $\frac{4}{10}$ identifies RRM only $\frac{3}{10}$ identifies weakly RRM only **No balance detected**
Positive correlation	$\frac{9}{10}$ identifies RUM only $\frac{1}{10}$ identifies RRM only **No balance detected**	$\frac{3}{10}$ identifies RUM only $\frac{6}{10}$ identifies RRM only $\frac{1}{10}$ identifies RRM weakly only **No balance detected**

Finally, we tested increasing the sample size to see if there was a point where both heuristics could be identified; particularly we tested 20 and 40 thousand samples in both proportions with no correlation. We were not able to detect any degree of coexistence of RUM and RRM in any of the 40 cases analysed.

The results from the cases where either RUM dominates or RRM dominates are consistent: the balance or coexistence of RUM and RRM in the choice model is improbable in this dataset. However, the RUM mechanism seems to be more robust by being able to accommodate RRM individuals better than the RRM can accommodate RUM individuals, thus, imposing itself in the balance more frequently than RRM. Moreover, note that although we used the μ-RRM with the intention to increase the probability of detecting coexistence by increasing the behavioural difference of RRM and RUM, this was not effective. Then, for this kind of real discrete choice data, a model that includes both RRM and RUM is not suitable and should be avoided.

Analysis of parameters

The RRM-RUM models provide evidence of the degree of bias obtained when it is not possible to identify both choice heuristics present in the sample. The RRM-RUM model analysed here corresponds to the case where RUM dominates the sample without correlation between the probability of choosing a heuristic and the RUM's parameters. This was found to be the most favourable pairing for the RUM heuristic as it is identified in most cases. Table 6-4 presents the mean of the estimators and the t-test statistic of the mean against the target value of the identified heuristic: the RUM column summarises the eight cases where the RUM heuristic was identified; whereas the RRM column summarises the two cases where the RRM was identified. No class membership function parameter is reported because no balance between the two choice heuristics was found in any of these cases.

The results of Table 6-4 show that both estimated heuristics experienced a systematic bias in their parameters. Because we analysed the case where the RUM model dominates, the bias of the RUM heuristic is lower than that of the RRM heuristic –as expected. The bias in the RRM is shown by high t-test statistics for the estimates against their target values. Even though the biases of the RUM parameters are smaller, they remain statistically significant and substantial.

Table 6-4 Mean and t-test statistic against target value for RUM and RRM models

Estimator	RUM (t-test)	RRM (t-test)	Estimator	RUM (t-test)	RRM (t-test)
μ	-	0.76 (3.72)	Cost	-0.38 (-1.46)	0.07 (19.72)
Vehicle Time	-4.53 (1.61)	0.76 (7.11)	Wait Time	-17.98 (1.74)	3.18 (9.12)
Walk Time	-6.37 (0.45)	1.11 (11.51)	ASC1	0.48 (-0.23)	0.10 (0.08)
ASC2	0 (fixed)	0 (fixed)	ASC3	0.10 (-0.04)	0.00 (-1.59)
ASC4	0.81 (0.10)	0.14 (-1.01)	ASC5	0.73 (0.48)	0.13 (-0.79)
ASC6	0.61 (0.20)	0.11 (-0.86)	ASC7	0.21 (0.09)	0.03 (-0.60)
ASC8	0.29 (-0.09)	0.05 (-0.75)	ASC9	0.42 (0.20)	0.06 (-1.07)

Finally, we consider the μ estimator in the RRM model, which controls the profundity of regret in the model. Smaller values represent greater "regret" behaviour, which departs more from the RUM behaviour. The estimate of the μ parameter is 0.76, which is substantially greater than the value of 0.2 used in the simulation ($t = 3.72$). Therefore, the μ parameter adapts the RRM to represent an intermediate behaviour, which still corresponds to greater regret than the traditional RRM, which has an implicit μ parameter of 1.

6.1.2.2. Analysis of the RUM and EBA case

Analysis of identifiability

Table 6-5 shows the results of estimating the latent class model with RUM and EBA as choice heuristics. The results suggest that a degree of balance was present in each of the 40 experiments. However, when RUM dominated the sample, the balance was weaker than when EBA dominated. Moreover, the introduction of correlation into the experiments decreased the identifiability of the balance both when RUM or EBA dominated. As expected, the model identified more readily the choice heuristic with greater proportion in the sample.

Table 6-5 Identifiability results of RUM and EBA models

Correlation	RUM dominates (70%)	EBA dominates (70%)
No correlation	$\frac{9}{10}$ identifies RUM and EBA $\frac{1}{10}$ identifies RUM and weakly EBA **Balance detected**	$\frac{10}{10}$ identifies RUM and EBA **Balance detected**
Positive correlation	$\frac{7}{10}$ identifies RUM and EBA $\frac{3}{10}$ identifies RUM and weakly EBA **Balance detected**	$\frac{9}{10}$ identifies RUM and EBA $\frac{1}{10}$ identifies EBA and weakly RUM **Balance detected**

These results suggest that when RUM and EBA are present in the data, they could be identified simultaneously with high identifiability in most cases. This indicates that neither RUM nor EBA can represent the behaviour of the other choice heuristic. Therefore, if it is possible to detect (e.g. with a qualitative method) the presence of an EBA kind of behaviour, then it is advisable to adopt the multiple heuristic model as RUM is not able to accommodate the EBA behaviour effectively.

Analysis of parameters

The EBA-RUM case presents the better indices of identifiability since in most cases it is strongly identifiable. Below, we analyse the less favourable case for the EBA heuristic. The EBA-RUM case analysed is the one where the RUM heuristic dominates with positive correlation between the RUM's sensitivities and the class membership function. This case provides evidence of the worst performance of the EBA heuristic. Table 6-6 presents the mean of the estimators and the t-test statistic of the mean against the target values across the experiments. The table analyses the seven identified models (of the ten tested).

Table 6-6 Mean of RUM parameter estimates and EBA log-weights together with t-test against target values

Estimator	RUM (t-test)	EBA (t-test)	Estimator	RUM (t-test)	EBA (t-test)
Cost 1	-0.34 (-0.24)	2.18 (1.10)	Cost difference/ Cost 2	0.14 (0.07)	1.92 (0.74)
Vehicle time	-4.62 (0.83)	1.99 (0.79)	Wait time	-19.58 (0.25)	2.65 (0.85)
Walk time	-6.38 (0.28)	3.06 (1.11)	ASC1	0.48 (-0.20)	0.72 (0.57)
ASC2	0 (fixed)	0 (fixed)	ASC 3	0.09 (-0.15)	0.16 (0.41)
ASC 4	0.75 (-0.36)	0.79 (0.31)	ASC 5	0.70 (-0.06)	0.63 (0.19)
ASC 6	0.57 (-0.28)	0.78 (0.47)	ASC 7	0.21 (0.11)	0.23 (0.12)
ASC 8	0.28 (-0.17)	0.58 (0.71)	ASC 9	0.40 (0.01)	0.59 (0.48)
θ_0	0.22 (0.94)		θ_1	1.30 (-0.50)	

All RUM parameters are unbiased with the greatest t statistic against the target value being 0.83. Thus, the RUM estimated parameters are not statistically different from their targets. The EBA parameters are also unbiased with the greater t statistic being 1.11. Note however, that this is the worst case for EBA in the EBA-RUM models.

6.1.2.3. Analysis of the RUM and SS case

Analysis of identifiability

Table 6-7 shows the results of estimating SS and RUM jointly. When RUM dominated the sample, our procedure was always able to identify a RUM class. Moreover, the model exhibited a certain degree of balance where the SS class was also identified, albeit weakly, since some parameters had extreme variance. The introduction of higher correlation did not influence identifiability.

When SS dominated the sample, a proper balance was detected. The model could identify, with reasonable variance, the estimators of the RUM heuristic, the SS heuristic, and the class membership function. When correlation was introduced, the degree of identifiability decreased, since one of the cases was no longer completely identifiable.

The results of the SS and RUM case suggest that a balance is plausible; however, it depends on the proportion of the population that use each choice heuristic. When the greater proportion of DMs follows the RUM heuristic, incorporating the SS heuristic did not compensate for the loss of likelihood of the RUM individuals. Conversely, when the proportion of SS dominated, the better performance of the RUM individuals compensated for the decrease in likelihood of the SS individuals. Hence, a balance may be achieved when

SS individuals are more numerous than the RUM ones. In either case, the RUM appears as the most robust heuristic as it can be identified even in cases of low proportion in the sample.

Table 6-7 Identifiability results of RUM and SS models (10,000 DMs)

Correlation	RUM dominates (70%)	SS dominates (70%)
No correlation	$\frac{9}{10}$ identifies RUM and weakly SS $\frac{1}{10}$ identifies weakly RUM and SS **Weak balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance detected**
Positive correlation	$\frac{9}{10}$ identifies RUM and weakly SS $\frac{1}{10}$ identifies weakly RUM only **Weak balance detected**	$\frac{9}{10}$ identifies RUM and SS $\frac{1}{10}$ identifies RUM and weakly SS **Balance detected**

Note that these results are specific for the present dataset; hence, if a dataset provides choice situation where SS differs more from RUM –and people behave following such heuristics– achieving balance would seem highly plausible.

Given the degree of identifiability detected in this experiment, we explored two larger sample sizes: 20 and 40 thousand DMs; for each of them, we studied the same two proportions and correlation structure. We summarise the results in Table 6-8.

Table 6-8 Identifiability results of RUM and SS models (20 and 40 thousand DMs)

Correlation	RUM dominates (70%)	SS dominates (70%)
	20,000 decision makers	
No correlation	$\frac{2}{10}$ identifies RUM and SS $\frac{8}{10}$ identifies RUM and weakly SS **Weak balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance detected**
Positive correlation	$\frac{10}{10}$ identifies RUM and weakly SS **Weak balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance detected**
	40,000 decision makers	
No correlation	$\frac{9}{10}$ identifies RUM and SS $\frac{1}{10}$ identifies RUM and weakly SS **Balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance detected**
Positive correlation	$\frac{10}{10}$ identifies RUM and SS **Balance detected**	$\frac{10}{10}$ identifies RUM and SS **Balance detected**

The results for the 20 thousand sample experiment achieve higher identifiability compared to the 10 thousand experiments. In both cases that RUM dominates the sample, the model can identify strongly the RUM and weakly SS; whereas in the case where SS dominates the sample, the model always identifies both heuristics. The difference between the 20 and 10 thousand experiments is that, for the 20 thousand experiments, in the RUM dominating case two models strongly identified both heuristics, whereas, in the SS dominating case no weakly identification was obtained.

The 40 thousand experiment results strongly improved identifiability when compared to the 20 and 10 thousand experiments. Even though the SS dominating case improved its results in terms of identifiability (as the standard deviation of the parameters tended to decrease), the main identifiability improvement was obtained in the RUM dominating case. The 40 thousand sample was enough to achieve strong identifiability in almost all cases where RUM dominates the sample. Therefore, there is evidence to confirm the theoretical findings of Chapter 5, plausible increases in sample size increases identifiability; however, a non-dismissible increase might be needed.

Finally, we tested whether the number of alternatives increases the behavioural difference between the two heuristics and, therefore, improves identifiability. We tested the SS and RUM in both heuristic proportions, with and without correlation, and ten thousand observations. The results are summarised in Table 6-9 and suggest that increasing the number of alternatives in the RUM-SS model decreases the identifiability of the heuristics. Even though the decrease is small, it is noticeable. A possible reason for this is that since the probability mass is distributed in more alternatives, the absolute difference of RUM and SS decreases; therefore, identifiability also decreases.

Table 6-9 Identifiability results of RUM and SS models with 10,000 DMs and 7 alternatives per choice sets

Correlation	RUM dominates (70%)	SS dominates (70%)
No correlation	$\frac{10}{10}$ identifies RUM and weakly SS **Weak balance detected**	$\frac{1}{10}$ identifies RUM and SS $\frac{9}{10}$ identifies RUM and weakly SS **Weak balance detected**
Positive correlation	$\frac{8}{10}$ identifies RUM and weakly SS $\frac{2}{10}$ identifies RUM only **Weak balance detected**	$\frac{10}{10}$ identifies RUM and weakly SS **Balance detected**

Analysis of parameters

The SS-RUM case enables us to analyse how well the RUM can be estimated even when it is the lesser used heuristic. The SS-RUM data analysed here corresponds to the case where the SS dominates the sample and there is positive correlation between the class membership function and the RUM sensitivities. Among the SS-RUM cases, this is the least unfavourable case for RUM. It reveals the RUM's performance in the worst case where it is identified. Table 6-10 presents the mean of the estimators and the mean of the t-test against the target values of the RUM parameters, SS parameters and class membership function parameters for the cases analysed. The table summarises the nine cases where both heuristics were identified.

Table 6-10 Mean and t-test against target values of RUM or SS estimation (10,000 DMs)

Estimator	RUM (t-test)	SS (t-test)	Estimator	RUM (t-test)	SS (t-test)
Cost sensitivity	-0.29 (0.08)	-6.31 (0.13)	Cost difference/ Cost threshold	0.09 (-0.32)	0.28 (0.15)
Vehicle time sensitivity	-4.69 (0.29)	-11.70 (0.47)	In-vehicle time threshold	-	0.61 (0.25)
Wait time / MRS veh-wait	-22.29 (-0.51)	1.64 (0.50)	Walk time / MRS veh-walk	-6.65 (-0.20)	4.11 (0.36)
ASC1	0.42 (-0.41)	-0.83 (-0.08)	ASC2	0 (fixed)	0 (fixed)
ASC 3	0.05 (-0.32)	-0.81 (0.49)	ASC 4	0.91 (0.30)	-0.63 (0.48)
ASC 5	0.69 (-0.12)	-0.75 (0.23)	ASC 6	0.70 (0.35)	-0.76 (0.19)
ASC 7	0.13 (-0.21)	-0.87 (0.25)	ASC 8	0.41 (0.37)	-0.78 (0.40)
ASC 9	0.44 (0.09)	-0.77 (0.39)			
θ_0	-0.02 (-0.12)		θ_1	-1.33 (0.46)	

With this pairing of heuristics, all model parameters were estimated with small and not statistically significant bias. Indeed, t-test against the target values of the model parameters were low, with the greatest value being 0.51. Even though the RUM heuristic was used in smaller proportion, the model could estimate the parameters accurately, including the difference between the cost sensitivity of DMs with and without the *trait*. As expected due to its high proportion in the synthetic sample, we could identify all parameters of the SS model. Even though the SS proportion is higher than that of the RUM (70% versus 30%), the bias of the parameters is comparable to that of the RUM. Finally, the class membership function is also estimated accurately.

When comparing the identifiability of parameters with the EBA-RUM case, the results suggest that the parameters of the SS-RUM case are more accurately estimated. Indeed, the RUM parameters, EBA versus SS parameters, and class membership function parameters where better estimated in the SS-RUM case.

6.2. Empirical Identifiability of Three Heuristics Model

Chapter 5 provided evidence about the phenomenon behind the identifiability of multiple heuristic models. The findings for two heuristics is readily interpretable, yet, they also apply to the multiple heuristic case, as shown in Chapter 5.2. Then, Chapter 6.1 provided us with a deep understanding of the dynamics of different heuristics and the factors affecting the identifiability of choice heuristics. In this section, we show an application to a three heuristic case.

6.2.1. Experimental design

The objective here was showing a case where identifiability is successfully obtained in a three heuristics case. To achieve it, we selected the heuristics that exhibited higher degree of identifiability in conjunction with RUM: SS and EBA.

Following the same structure as in the two heuristics case, we simulated a synthetic population choosing from the dataset of the Las Condes – CBD, San Miguel – CBD experiment. All choice sets considered three alternatives and sample sizes of 20,000 and 40,000 observations.

We chose only one proportion of heuristics in the sample. As SS had the worst identifiability among the three heuristics, the experiment was designed to have a higher proportion of it. Thus, the synthetic population was designed to choose SS with a 40% probability and with a probability of 30% for both RUM and EBA.

With the objective of generating a non-constant class membership function, we introduced a socio-demographic characteristic that affected the probability of choosing each heuristic. DMs had this trait ($z = 1$) with a probability of 70%. The probability of using each choice heuristic conditional on the trait was given by the inverse Logit function[14] (6.3) which has as input the function shown in (6.2). The parameters are given in Table 6-11.

$$W_h = \theta_{0,h} + \theta_{1,h} \cdot z \tag{6.2}$$

$$\pi_h = \frac{\exp(W_h)}{\sum_{j \in H} \exp(W_j)} \tag{6.3}$$

---

[14] Even though the inverse Logit function is equivalent to a MNL model, we explicitly avoid referring to the class membership function as a logit model, because we do not interpret it under maximum utility theory. We avoid employing maximum utility theory due to the presence of non-utility maximiser classes.

Table 6-11 Three heuristic class membership function

	RUM	EBA	SS
θ_0	0	0.581	-0.674
θ_1	0	-1.000	1.200

The conditional probability of choosing each heuristic is giving in Table 6-12. The results of it are obtained by applying (6.3) with parameters given in Table 6-11.

Table 6-12 Conditional probability of choosing each heuristic in the three heuristics experiment

	RUM	EBA	SS
With trait	29.9%	19.6%	50.5%
Without trait	30.3%	54.2%	15.5%

Regarding the sensitivities for the attributes of each heuristic, we used the same values used for the two heuristics' experiment in Section 6.1 (Table 6-2). Nonetheless, we simulated no correlation between the probability of choosing a heuristic and the individual sensitivity for attributes, since the degree of correlation simulated was not shown as a crucial element.

6.2.2. Analysis of results

Table 6-13 presents the identifiability of the three heuristics case for the 20 estimations. Although the results are consistent with the two heuristics experiment, they are not directly comparable. This is because in the two heuristics case, in one scenario the SS heuristic was clearly dominant, whilst in the other it was in a lower proportion; but in this case, the SS heuristic is slightly dominant in the sample. Results suggests that the sample of 20 thousand DMs has a size for which the model is transitioning from non-identifiable to identifiable. In 40% of the cases, the model was completely identified and in one case only SS was identified alone. Regarding the RUM and EBA heuristics, they were successfully identified in most cases. The four parameters that determine the class membership function were also identified in nine of the ten cases.

Table 6-13 Identifiability of the three heuristics case

Sample size	RUM identifiability	EBA identifiability	SS identifiability
	$\frac{9}{10}$ identified	$\frac{9}{10}$ identified	$\frac{5}{10}$ identified
20,000			$\frac{5}{10}$ weakly identified
	$\frac{1}{10}$ no identification	$\frac{1}{10}$ no identification	
		Weak balance detected	
40,000	$\frac{10}{10}$ identified	$\frac{10}{10}$ identified	$\frac{10}{10}$ identified
		Balance detected	

The 40 thousand sample is consistent with the two heuristics case: all heuristics are identified. This case shows that even though the model consumes numerous degrees of freedom (39 in total), it can identify all parameters. Moreover, note that the model could identify the different heuristics sensitivities based only in the different way individuals integrate the attributes in their decision heuristics. This demonstrates that it is possible to identify heuristics without confounding attribute sensitivity with attribute processing.

We present the results of several estimations of the RUM, EBA, and SS heuristics respectively. We show the mean estimator and t-test statistic of the mean against the target value. First, we show cases where weak or no identifiability was achieved for the 20 thousand estimations. Then, we show results with strong identification for the 20 and 40 thousand samples respectively. In some cases, only a single model falls in the category of non-identified; whereas in most cases, several models fall in their respective category.

In the case where more than one experiment had available results (with a maximum of ten), we calculated the mean of the estimators and the standard deviation of such mean from which the t-test is calculated. Therefore, the interpretation of the t-tests is different: in the former case with one estimation, we analyse if the confidence interval of the parameter contains the target estimate; whilst in the latter with several estimations, we analyse if the confidence interval of the mean of the parameter contains the target estimate.

Table 6-14 shows the RUM estimation in each of the twenty models. The single case where RUM was not identified is shown first. This estimation is characterized by large standard deviation[15], justifying relatively small t-tests (up to 3.80) compared to the magnitude of the bias and serves as an example of the magnitude of bias obtained in an unidentified heuristic.

Table 6-14 Estimation results for RUM in the RUM-EBA-SS model

	20 thousand observations		40 thousand observations
	Non identified cases	Identified cases	All cases identified
Cost	5.11 (1.51)	-0.42 (-0.45)	-0.49 (-2.71)
In-vehicle time	-31.13 (-2.35)	-5.33 (-0.37)	-5.10 (-0.15)
Waiting time	-141.38 (-3.03)	-16.08 (1.00)	-18.71 (0.51)
Walk time	6.05 (0.70)	-7.00 (-0.49)	-6.71 (-0.35)
ASC1	6.27 (2.37)	0.18 (-1.62)	0.43 (-0.40)
ASC2	0	0	0
ASC3	0.63 (0.24)	-0.08 (-0.60)	0.07 (-0.32)
ASC4	-1.11 (-0.36)	0.44 (-0.91)	0.70 (-0.40)
ASC5	4.89 (2.02)	0.55 (-0.94)	0.68 (-0.12)
ASC6	1.67 (0.22)	0.28 (-1.00)	0.59 (-0.05)
ASC7	-1.58 (-0.34)	-0.04 (-1.00)	0.13 (-0.50)
ASC8	2.67 (0.86)	0.07 (-0.68)	0.26 (-0.20)
ASC9	7.32 (3.80)	0.05 (-1.32)	0.42 (-0.10)

[15] Recall that in a t-test, the higher the standard deviation, the smaller the t-test value. Therefore, even if the bias of a parameter is large, it may be statistically non-different from the target due to even larger variance.

In the identified cases, either in the 20 thousand or 40 thousand estimations, most of the means of the estimates are within the 95% confidence interval (the critical value of a t-student with 9 degrees of freedom is 2.26). In the case of the 40 thousand estimations only the cost parameter presents a high bias, which is coincident with the bias in the SS heuristic presented later; we will look at this in more depth in the SS heuristic analysis.

In the case of the EBA heuristic, presented in Table 6-15, the non-identified heuristic presents heavy biases as expected. In the nineteen identified cases, the model correctly identifies the target parameter. Indeed, no t-test statistic is even close to unity, which shows the accuracy of the estimation.

The case of the SS heuristic, presented in Table 6.16, is consistent with the two heuristics case presented before. The non-identified cases show high bias as expected, while most of the parameters in the identified cases show small and non-significant bias. In this case, the only parameters with non-dismissible bias in the identified estimation are associated with the cost attribute. This is consistent with the bias in the RUM model; we interpret that there could be a confounding, where these models might be capturing variations in cost tastes (or cost sensitivity) rather than only a different heuristic.

Table 6-15 Estimation results for EBA in the RUM-EBA-SS model

	20 thousand observations		40 thousand observations
	Non identified cases	Identified cases	All cases identified
Cost	0.42 (-4.53)	1.59 (0.48)	1.28 (-0.45)
Cost 2	0.45 (-3.84)	1.56 (0.30)	1.38 (-0.02)
In-vehicle time	0.53 (-4.04)	1.61 (0.43)	1.33 (-0.26)
Waiting time	1.38 (-3.48)	2.36 (0.49)	2.09 (0.05)
Waiting time	1.32 (-5.55)	2.49 (0.41)	2.23 (-0.34)
ASC1	0.53 (-2.57)	0.61 (0.59)	0.44 (0.17)
ASC2	0	0	0
ASC3	0.04 (-0.30)	0.02 (-0.17)	0.06 (-0.19)
ASC4	0.30 (-1.53)	0.82 (0.56)	0.61 (0.14)
ASC5	0.08 (-2.99)	0.66 (0.43)	0.52 (-0.07)
ASC6	0.09 (-1.97)	0.63 (0.45)	0.38 (-0.53)
ASC7	-0.42 (-3.30)	0.33 (0.54)	0.25 (0.38)
ASC8	-0.16 (-2.44)	0.27 (0.02)	0.18 (-0.50)
ASC9	-0.14 (-2.94)	0.47 (0.51)	0.28 (-0.34)

Table 6-16 Estimation results for SS in the RUM-EBA-SS model

	20 thousand observations		40 thousand observations
	Non identified cases	Identified cases	All cases identified
Cost	-10.96 (-4.53)	-9.57 (-2.49)	-10.13 (-6.27)
Cost threshold	0.39 (2.33)	0.43 (16.6)	0.44 (9.25)
In-vehicle time	-12.72 (-0.78)	-11.5 (0.45)	-12.29 (-0.42)
MRS waiting time	4.32 (0.81)	3.91 (-0.43)	4.02 (0.17)
MRS walking time	2.02 (1.63)	1.59 (0.59)	1.58 (0.39)
Time threshold	0.61 (0.39)	0.57 (-0.93)	0.61 (0.45)
ASC1	1.18 (1.27)	-0.64 (0.69)	-0.67 (0.45)
ASC2 (fixed)	0	0	0
ASC3	2.22 (1.27)	-0.77 (1.04)	-0.87 (0.42)
ASC4	1.53 (1.80)	-0.59 (1.23)	-0.71 (0.36)
ASC5	-0.46 (0.86)	-0.53 (1.69)	-0.77 (0.14)
ASC6	1.59 (1.42)	-0.32 (1.25)	-0.74 (0.34)
ASC7	-0.25 (1.19)	-0.68 (3.67)	-0.88 (0.25)
ASC8	3.02 (1.59)	-0.55 (0.96)	-0.80 (0.35)
ASC9	2.66 (1.74)	-0.41 (1.25)	-0.81 (0.26)

6.3. Conclusions of the Analysis of Empirical Identifiability

At the start of this Chapter, the link between the theoretical and empirical approaches was a graphical analysis exposing the difference in choice heuristics (Figure 6-1). Indeed, in our mode choice context, Random Utility Maximisation (RUM) did not differ much from Random Regret Minimization (RRM). However, RUM exhibited an important behavioural difference from both the Stochastic Satisficing (SS) and Elimination by Aspects (EBA) heuristics. This analysis was confirmed by our model estimation using synthetic data, confirming the theoretical findings in Chapter 5 and validating this simple diagnosis test. Therefore, a worthwhile strategy could be to analyse the heuristics before estimating a combined model, and this can be undertaken using the straightforward diagnostic analysis presented here. This way, with some testing parameters, the modeller can examine if the choice sets are sufficiently rich in their choice behaviour to estimate the desired heuristics.

When tested using realistic mode choice scenarios, the RUM heuristic was found to be the most robust. RUM was identified, even when its proportion in the population was smaller than that of the other heuristics. Indeed, the RUM heuristic frequently outperformed the RRM heuristic without generating a balance even when RRM was dominant. Therefore, in this specific context, RUM performs well when interpreting RRM results. When the RUM heuristic was tested together with either the SS or EBA, the most frequent output was found to be a balance between the two heuristics; with greater frequency of coexistence when either the SS or the EBA were dominant. This suggests that RUM does not interpret adequately SS or EBA behaviour, but it is only appropriate to include them as second heuristics when their proportions are substantial.

We also tested a scenario with three choices. Our findings indicate that it is possible to identify highly explanatory class membership functions even in the case of three heuristics. However, identifiability is not straightforward, even in the case tested where the choice heuristics propitiated identifiability. Therefore, if another less propitious context is faced, a sample size much larger than forty thousand observations might be needed to identify heuristics only based on their different behaviour when interpreting the attributes.

Among the meta-dimensions tested in the experiment, the most important ones were the type of heuristic, the sample size, and the proportion associated with each heuristic in the sample. Conversely, the degree of correlation tested and the size of the choice set had a smaller impact.

The experiments presented here show that, in principle, it is possible to estimate sophisticated class membership functions and choice heuristics simultaneously. The model remains identifiable even if some variables affect both the class membership function and heuristic sensitivities. Nevertheless, it is necessary to formulate both expressions appropriately, which is not always straightforward.

In this analysis several possible influences on the identifiability of different choice heuristics were considered. First, the choice heuristics simulated considered the same variables influencing the decision processes. Accordingly, the identifiability detected was due to different underlying choice behaviour rather than different influential variables. If the real generation process includes different heuristics and different variables within the heuristics, then we expect the combination of choice heuristics to be more readily identifiable. The second consideration includes the parameters associated with the influential variables of the experiment. We chose parameters so that the model would be readily identifiable, noting that greater values of alternative specific constants would decrease the identifiability of the heuristics.

Finally, recall that the empirical analysis was made for a specific context and with limited number of alternatives. Our analysis of a mode choice example in transport suggests that the data does not contain sufficient richness to identify RUM and RRM simultaneously, but it is rich enough to identify RUM simultaneously with either SS or EBA. However, other contexts may exploit the behavioural difference between RUM and RRM. Therefore, further analysis should be performed to explore the relationship between choice heuristics more fully.

6.4. Publication History

A first version of this chapter was presented at a seminar in the United Kingdom. Later, it was presented in the *18th Chilean Transport Engineering Conference* (Gonzalez-Valdes et al., 2017). Most of this chapter is currently under review for publication in the journal *Transportation Research Part B: Methodological* (Gonzalez-Valdes et al., 2018). The seven-choice set case for SS, the 20 and 40 thousand estimation results and analysis for SS and RRM, and the three heuristics case remain unpublished.

7. IDENTIFYING THE PRESENCE OF HETEROGENEOUS DISCRETE CHOICE HEURISTICS AT AN INDIVIDUAL LEVEL

Through Chapters 5 and 6 we have used latent classes (LC) throughout. The LC model has two decision levels: in the first, the class membership level, the choice heuristic followed by the individuals is modelled. In the second, the choice level, the preferences for alternatives are modelled conditional on the choice heuristic followed. Even though we have addressed the complete model in Chapters 5 and 6, the focus has been in the choice level. In Chapter 5 we have shown the conditions that enable identifying LC multiple heuristics models. In Chapter 6, we studied the conditions that enhance the possibility of identifying the several heuristics involved. In this chapter we focus on the class membership level and how its modelling at an individual level gives further insights into the population behaviour, allowing to solve several problems.

Still, there are several challenges that must be addressed when using a LC framework to model multiple choice heuristics: i) class membership is hard to model, ii) both LC levels must be modelled simultaneously, and iii) there is no certainty that the considered choice heuristics are actually present within the population.

Class membership is usually modelled using the inverse Logit function. As addressed in Section 2.4, the most frequent formulations assume that all individuals in the population have the same probability of following a specific choice heuristic. This simple formulation is unfortunately necessary, since more general and informative formulations have been found to be unidentifiable in practice. To the best of our knowledge, only two studies have been reported where a non-constant population-wide class membership function has been estimated (Leong and Hensher, 2012b; Hess and Stathopoulos, 2013). Leong and Hensher (2012b) work with a class of RUM and another of RRM individuals. In their experiment, some parameters of the RUM and the RRM class had to be fixed (as one equal to minus the other) for the model to be identifiable. Hess and Stathopoulos (2013) also worked with RUM and RRM; in their experiment, they had to use latent variables to explain the class membership function and choices to estimate the heuristics' parameters, otherwise, identifiability could not be obtained just from the choices of the individuals. Therefore, identifying heuristics and class membership function simultaneously is not straightforward.

Modelling the class membership level and the choice level simultaneously has further complications apart from identifiability. If modelling DMs choices following a single choice heuristic is challenging (as it has been a research topic for decades), modelling their choices through multiple choice heuristics is even harder, since each of them must be modelled. Moreover, as additionally a class membership level needs to be modelled, the complexity of the model increases even more. This issue is further complicated by the interaction between the two levels, since modifying one of them affects the other.

Selecting the choice heuristics to be considered is also a significant challenge. Even though in practice RUM models are the most popular alternative, there are numerous alternative choice heuristics to model individuals' behaviour. Ideally, a large number of choice heuristics could be considered, and a LC model would indicate if a given heuristic is present or not in the population by assigning a low probability to it. However, in practice, LC models have identifiability problems even when simply modelling two classes. Therefore, a limited number of choice heuristics must be properly selected beforehand, or different combinations of heuristics can be tested.

It is important to consider that these three challenges interact with each other. The LC model could be used to test the presence or absence of a particular choice heuristic within a population. However, if the model results indicate the absence of a heuristic, this could be due to an erroneous class membership function, to a misspecification of the choice heuristic itself in the choice level, or to its real absence.

We propose a model that addresses the three challenges of the LC approach. The Mixed Heuristic Model (MHM), can help identify the presence of choice heuristics by giving higher flexibility to the class membership function. This flexibility is given by a random variable that allows to capture heterogeneity in decision rules. This heterogeneity is confirmed by analysing the class membership probability at an individual level.

7.1. The Mixed Heuristics Model

When applying LC models to the context of multiple choice heuristics, each class represents an independent heuristic. This way, the probability of individual $q \in Q$ choosing alternative i, P_{qi} has two components. The first is the probability of individual q following choice heuristic $h \in H$, defined by the class membership function $\pi_{q,h}$. The second is the probability of individual q choosing alternative i conditional on following choice heuristic h, $P_{q,i}(h)$. Then, the probability of individual q choosing alternative i ($P_{q,i}$) is given by the *total probability* presented in (7.1).

$$P_{q,i} = \sum_{h \in H} P_{q,i}(h) \cdot \pi_{q,h} \tag{7.1}$$

In (7.1), the functional form of $P_{q,i}(h)$ depends on the specific choice heuristic h. For example, if the classic RUM model is one of the considered heuristics, then its probability $P_{q,i}(h)$ is given by the well-known ratio of the exponential of the utility functions.

The probability of individual q following choice heuristic h, $\pi_{q,h}$, is typically given by the inverse Logit function (7.2). To explain the fact that different individuals may follow different choice heuristics, this probability takes as an input a $\gamma_{q,h}$ function that varies across heuristics and individuals.

$$\pi_{q,h} = \frac{\exp(\gamma_{q,h})}{\sum_{m \in H} \gamma_{q,h}} \tag{7.2}$$

The main difference between the current approach to model heterogeneous discrete choice heuristics and the proposed MHM lies in the specification of the $\gamma_{q,h}$ function. This new functional form handles the three challenges discussed previously and presented below.

7.1.1. The current approach

The $\gamma_{q,h}$ function may include variables such as socio-demographic characteristics, experimental conditions, or latent traits. Let r be the set of such variables and θ_r the DMs' sensitivities to such variables. The $\gamma_{q,h}$ function could be a linear expression of the variables and their sensitivities, as shown in (7.3). However, the most traditional structure includes only the constant θ_m terms (Araña et al., 2008; Hess et al., 2012; Adamowicz and Swait, 2013).

$$\gamma_{q,h} = \theta_h + \sum_r \theta_r \cdot x_{q,r} \tag{7.3}$$

7.1.2. The proposed approach and its advantages

The MHM gives complete flexibility to the $\gamma_{q,h}$ function by treating class membership as a random variable (7.4):

$$\gamma_{q,h} \sim N(\mu_{q,h}, \sigma) \tag{7.4}$$

Like any mixed model, from the MHM we may compute the density functions of the class membership probabilities at an individual level by conditioning the population distribution on the DMs' choices. By exploiting this feature, this flexible γ function can be adapted to represent different phenomena. The simpler case is when the same individual follows the same choice heuristic across different choice situations. This case is obtained by simultaneously conditioning class membership on all the choices of the individual. The function could be also adapted so that the same individual can follow different choice heuristics in each choice situation. This case is obtained by independently conditioning the class membership on each of the choices of the individual.

In principle, any probability distribution could be used and in this research we consider a Normal distribution. The model estimates the mean of the distribution and its variance. The mean explains the overall proportions of the choice heuristics in the population and the variance adjusts the heuristic selection to the heterogeneity across individuals.

The MHM structure addresses the three challenges previously mentioned as follows:

- The variance of the random $\gamma_{q,h}$ function handles the modelling of the class membership level, without the need to propose a specification of explanatory variables such as (7.3).
- By giving higher flexibility to the class membership level, the MHM focuses on modelling the choice level. This avoids the challenge of modelling the class membership level and the choice level simultaneously.
- By conditioning the population class membership distribution on the DMs' choices, we compute their probabilities of using each heuristic. This way, it is possible to analyse whether a choice heuristic is followed by an individual with a non-negligible probability.

7.2. Testing the Mixed Heuristics Model with Synthetic Data

In this section we apply the MHM to a synthetic database to test the accuracy of the MHM in terms of identifying the individual probabilities of following a given choice heuristic. For the different scenarios considered we analyse the impact of: i) the sample size and ii) the number of observations per individual on the accuracy of the MHM.

The analysis is as follows. First, simulated choice sets are obtained from a real dataset. Then, choices of fictitious individuals are simulated according to one of the two choice heuristics. Later, the MHM is estimated. Finally, we analyse the probability that the model assigns, at an individual level, to the choice heuristic actually followed.

7.2.1. Experimental design

We use the Las Condes – CBD, San Miguel – CBD dataset detailed in Section 3.1; however, we do not restrict the choice set size, as it was not shown in Chapter 6 as a key element of identifiability. We considered total sample sizes of 1,000 and 5,000 observations.

Regarding the synthetic population, we considered that DMs face 1, 10, and 25 choice scenarios; each choice scenario is decided with the same heuristic. To avoid the issue of identifiability, we consider that DMs choose under the two heuristics with higher identifiability shown in Chapter 6: RUM and EBA.

Once more, we introduced a socio-demographic characteristic *z* for each simulated individual, simply named *trait*. This variable has, again, 70% probability of being present in the population. Conditional on *trait*, DMs' probabilities of following RUM or EBA is obtained accordingly to Table 7-1. This results in an overall 51% probability of following RUM.

Table 7-1 Heuristic proportions in the simulated experiment

	RUM	EBA
With trait	60%	40%
Without trait	30%	70%

Once the choice heuristic is defined, the individual chooses accordingly. The target parameters for both heuristics are presented in Table 7-2. Note that the alternative specific constants (ASC) are relatively low –in contrast to the attributes– to highlight the heuristic's behaviour rather than the preferences toward the alternatives themselves. In fact, when the ASC importance dominates other attributes, the different choice heuristics might not be identified since any of them could be able to recover the observed market shares.

The RUM heuristic has a linear utility function on the monetary and temporal attributes. The EBA choice heuristic is defined by weights (which are estimated) and thresholds (which are imposed). Two thresholds for the cost (60 and 100 CLP) and one for the travel time (15 min), walking time (8 min) and waiting time (5 min) are considered.

Table 7-2 Simulation parameters for the synthetic population

Parameter	EBA weight	RUM value	Parameter	EBA weight	RUM value
Cost 1	2.08	-0.50	ASC 3	0.10	0.10
Cost 2	3.00	-	ASC 4	0.59	0.80
In-vehicle time	2.08	-5.00	ASC 5	0.53	0.70
Waiting time	2.77	-20.00	ASC 6	0.47	0.60
Walking time	3.33	-6.5	ASC 7	0.18	0.20
ASC1	0.41	-0.50	ASC 8	0.26	0.30
ASC2 (fixed)	0	0	ASC 9	0.34	0.40

We estimate a MHM and a LC model using Bayesian estimation with the JAGS package for R software. To sample the joint posterior distribution of the parameters, we used Gibbs sampling. In our study, 5,000 burn-in samples were needed to reach the stationary state of the Markov Chain. Finally, additional to the burn-in samples, 10,000 samples were analysed with a thinning parameter of two. The specification of the model was divided into the two aforementioned levels: class membership level and choice level. For the former, the MHM considers the random specification of (7.4), whilst, the LC considers a constant population-wide. Therefore –and because it is hard to model–, the *real* specification with the trait is not estimated in either case. For the choice level, in both cases the *real* underlying choice heuristics is estimated. The γ's mean and the choice heuristics parameters prior follow a flat Normal distribution with mean zero and standard deviation 100. The γ standard deviation has a uniform prior between zero and three.

7.2.2. Analysis of results of the synthetic data experiment

Results are analysed in two ways. First, we show how the model can identify the choice heuristic that each individual follows *a posteriori* and compare this with the *a posteriori* performance of the LC approach. Then, we present how the MHM recovers the target parameters of each choice heuristic.

To analyse the identification of choice heuristics at an individual level, two dimensions were considered: the total sample size (either 1,000 or 5,000 total observations) and the number of observations per individual (either 1, 10 or 25). As we know which heuristic RUM or EBA– was followed by each simulated individual, we can compute the posterior probability of each individual following their actual choice heuristic. The posterior probability for both the LC and MHM models was calculated using (7.5). In the case of the MHM model, the population distribution of the heuristic is conditioned in the choices of the individual. In the case of the LC model, the constant probability of choosing a heuristic is the prior of the individual probability of using a specific heuristic. In both cases, once the heuristic is chosen (which is constant for all choices of the same individual) the individual's choices are independent from each other. Therefore, the calculation of $\Pr(choices|heuristic)$ is straightforward; under independence, the probability of all choices is just the product of the probabilities of each choice.

$$\Pr(heuristic|choice) = \frac{\Pr(choices|heuristic)\Pr(heuristic)}{\sum_{h\in H}\Pr(choices|h)\Pr(h)} \tag{7.5}$$

Table 7-3 presents the average choice heuristic success probabilities (i.e. the average probabilities of choosing the correct heuristic) and their standard deviation for the traditional LC approach and the proposed MHM. For the MHM the quintiles of these probabilities are also provided.

Table 7-3 LC and HMH success in identifying the chosen heuristic

Choices per individual	Heuristic	LC Success Prob% (sd)	MHM Success Prob% (sd)	MHM Success Probability Quantiles
		Sample size = 1,000 responses		
25	RUM	28.8 (42.7)	88.0 (9.1)	[81.5, 88.6, 92.0, 94.3]
	EBA	82.1 (33.0)	73.7 (17.4)	[62.3, 71.5, 80.9, 86.5]
10	RUM	44.8 (49.3)	75.7 (9.7)	[71.5, 75.9, 78.9, 81.5]
	EBA	59.9 (49.6)	53.3 (17.2)	[36.9, 47.8, 58.0, 68.2]
1	RUM	44.5 (38.1)	41.0 (7.1)	[35.1, 39.4, 43.6, 47.0]
	EBA	63.6 (32.5)	62.2 (6.7)	[56.1, 61.0, 63.7, 68.7]
		Sample size = 5,000 responses		
25	RUM	85.9 (25.5)	90.5 (10.5)	[87.4, 92.9, 95.3, 96.7]
	EBA	74.0 (21.4)	87.9 (13.4)	[81.7, 91.0, 94.3, 96.2]
10	RUM	83.9 (4.8)	83.4 (18.3)	[75.7, 88.2, 92.7, 95.0]
	EBA	82.4 (4.4)	76.0 (22.6)	[59.1, 80.4, 89.3, 93.3]
1	RUM	57.4 (5.9)	51.0 (9.0)	[42.5, 47.3, 53.3, 60.2]
	EBA	56.5 (7.3)	55.1 (7.7)	[48.8, 54.3, 57.9, 61.6]

For the MHM, as expected, as the sample size increases the predicted choice heuristic success probability also increases. As the number of observations per individual increases, the predicted choice heuristics success increases significantly. This higher success is obtained because the MHM has more information, as in the simulation each individual follows the same heuristic across choices. Regarding the success of each heuristic, EBA has a higher success than RUM when there is only one response per individual, but it is outperformed by RUM as the number of observations per individual increase.

For the traditional LC approach, there is no clear tendency for the performance as either the sample size or the number of choices per individual increases. At an individual level, the estimation results indicate that in most cases only one of the heuristics is present *a posteriori*. The favoured heuristic is not the same across the different simulations; therefore, the success probability exhibits high standard deviation (similar to a Bernoulli process). The unbalance of the heuristics –recall Chapter 6– implies that only one is identified. Then, the model always guesses for the individuals choosing the identified heuristic; whereas for the others, the LC model never identifies the chosen heuristic. These results suggest that the LC approach, at least when the class membership function is constant, does not achieve the objective of identifying the presence of choice heuristics consistently. In contrast, the MHM can identify the presence of different choice heuristics and infer, with high probability when the individual chooses several times, the heuristic actually followed by the individual. Moreover, the overall variance of the success, considering the aggregate standard deviation of the success probability of RUM and EBA, is smaller in most cases for the MHM compared with the LC model.

The best average case is achieved with total sample size of 5,000 and 25 responses per individual, where the MHM can identify the heuristic used by each individual with an accuracy around 90% for both heuristics (while the traditional LC approach has nine points less of accuracy and double the standard deviation). Furthermore, the quintiles show an accuracy of over 90% for at least four of the five quintiles. For this best case, the model can recover the RUM parameters with high accuracy and the EBA parameters with only slight bias. Model parameters and the t-value against target values are presented in Table 7-4. Even though these results are satisfactory, they must be considered with prudence, since an individual being consistent through 25 choice scenarios might not be realistic. Nevertheless, these results validate the capabilities of the MHM in terms of identifying both the considered choice heuristics and the parameters that define the decision making process.

Table 7-4 Average estimation bias for the best MHM case: 5,000 total sample with 25 observations per individual

Parameter	EBA weight	RUM value	Parameter	EBA weight	RUM value
Cost 1	2.36 (1.25)	- 0.43 (0.53)	ASC 3	0.14 (0.19)	0.15 (0.41)
Cost 2	4.00 (1.93)	-	ASC 4	0.61 (0.00)	0.80 (0.02)
In-vehicle time	2.43 (1.81)	-5.22 (0.35)	ASC 5	0.62 (0.57)	0.69 (0.10)
Waiting time	3.12 (0.33)	-22.42 (0.94)	ASC 6	0.55 (0.45)	0.66 (0.40)
Walking time	3.78 (2.29)	-6.06 (0.76)	ASC 7	0.30 (0.68)	0.19 (0.03)
ASC1	0.23 (1.03)	0.53 (0.23)	ASC 8	0.38 (0.62)	0.36 (0.35)
ASC2 (fixed)	0 (fixed)	0 (fixed)	ASC 9	0.42 (0.39)	0.50 (0.57)

7.3. Searching for Heterogeneous Choice Heuristics in an Air Travel Context

In real choice experiments, the underlying choice heuristics are unknown and therefore identifying (and then modelling) them is not always straightforward. We use the MHM to analyse if a non-RUM choice heuristic can be detected on the Singapore air travel survey.

Air travel is an attractive market to test the presence of non-RUM choice heuristics since, to the best of our knowledge, only RUM have been reported in the literature (Chin, 2002; Adler et al., 2005; Bekhor and Freund-Feinstein, 2006; Theis et al., 2006; van Eggermond, 2007; Carrier, 2008; Wen and Lai, 2010; Rezaei et al., 2011; Drabas and Wu, 2013). Another attractive feature of air travel is how consumers generally access to a set of alternatives from which to choose: using search engines or inspecting directly on the airlines' website. In this section, we check whether people using search engines may follow different choice heuristic.

We start by describing the experiment and briefly recalling the database. Then, we analyse several possible choice heuristics that could be present in the sample. Finally, we discuss the conclusions of this empirical experiment.

7.3.1. Experiment design and estimation

We analyse DMs choosing from the Singapore air travel dataset detailed in Section 3.2. The two approaches for modelling heterogeneous choice heuristics, described in Section 7.1, are tested for this database. First, the traditional LC approach is used to analyse the behaviour at a population level. Then, the MHM is estimated to characterise DMs individually. Finally, we compare the models' results and extract the conclusions.

Five choice heuristics were individually tested as alternatives to RUM: EBA, two-stage EBA-RUM (detailed in Subsection 2.2.7), RRM, and SS. As this section does not have the objective of characterizing the population deeply, but rather analysing the possible presence of alternative choice heuristic, only simple and robust specifications of each heuristic are considered.

The RUM choice heuristic has a linear and additive structure and considers fare, flying time, and stop time. Even though individuals were given the airline name, given the number of airlines, no alternative specific constants were estimated for them.

The EBA choice heuristic considers the number of stops, fare, and if the carrier is regular or low-cost. As in EBA models all aspects are desirable, an alternative has the first aspect if the flight has no stops. The alternative has the second aspect if the fare is not much higher than the cheapest alternative (two thresholds were set at 13% and 65% price difference with the cheapest alternative). Finally, the alternative has the last aspect if it is a regular carrier.

In the two-stage EBA-RUM, two types of models were estimated; the first screens alternatives with high fare (EBA-RUM fare), while the other screens alternatives with stops (EBA-RUM stops). The specification of the subsequent RUM choice is the same as in the pure RUM model.

The RRM choice heuristic is linear and additive considering fare, flying time and stop time. The formulation implemented is the one proposed by Chorus (2010), also known as RRlog (Jang et al., 2017) which we detail in Subsection 2.2.3. With the objective of increasing the profundity of regret (van Cranenburgh et al., 2015), the fare attribute was scaled by 0.1.

Finally, the SS choice heuristic considers two attribute acceptability functions. The first acceptability function analyses cost and the other analyses flying time and stop time; in the latter, the marginal rate of substitution (i.e. the valuation of one hour at a layover in terms of one hour flying) is estimated.

Each of the non-RUM choice heuristics is tested individually as an alternative to the RUM choice heuristic. As the objective of this study is to identify the presence of heterogeneous choice heuristics, the analysis focuses on the class membership level. The detailed results of the choice level for each combination of heuristics are presented in Appendix C.

7.3.2. Analysis of results

Table 7-5 presents the non-RUM probabilities according to the traditional LC approach and the MHM, for all the considered combinations of heuristics. We also present the Deviance Information Criterion (DIC) which is a measure of fit that penalises for additional parameters. We present the traditional LC constant class membership probability for the entire population. For the MHM we present the mean and maximum values of the class membership probability within the population.

Table 7-5 Non RUM probability and estimation fit of the LC and MHM models

	Non-RUM probability (%)			DIC	
Second heuristic	LC	MHM Mean	MHM Max	LC	MHM
EBA	6.2	6.4	11.5	2829	2829
EBA-RUM stops	9.5	14.5	44.8	2831	2829
EBA-RUM fare	4.3	6.8	12.4	2830	2831
RRM	2.3	2.5	3.3	2828	2827
SS	15.8	29.7	72.2	2820	2819
None	-	-	-	2830	

The traditional LC results suggest that the probability of following a non-RUM choice heuristic is low, except for SS. Over the five non-RUM choice behaviours presented, the SS seems to be the most likely alternative behaviour, with a population-wide probability of 15.8%. For all combinations of heuristics, the traditional LC approach systematically results in a lower mean non-RUM probability than the MHM, possibly underestimating their presence. This is an interesting result, which we believe might occur due to the simplistic specification of the LC class membership function (i.e. only a population constant is used).

The LC and MHM models tend to coincide when the non-RUM choice heuristic is negligible, since both models assign a very low probability to it. As the behavioural heterogeneity –in terms of choice heuristics followed– rises, the traditional LC approach and the MHM differ, probably due to the incapacity of the traditional LC model to capture the heterogeneity across individuals in the sample.

In terms of model fit, the MHM model tends to slightly outperform the LC model by one DIC point. Note that even though the MHM has one additional parameter, the DIC incorporates a penalisation for the *effective* number of parameters. Results also suggests that incorporating a second heuristic could improve the overall fit when compared to the single heuristic RUM model.

Analysing whether the alternative choice heuristic is actually present or not is not straightforward. First, the non-RUM probability –or any probability– does not distribute Normal. This way, testing a null hypothesis for the probability being zero (generally by using a t-test), which is the boundary of the feasible values, is not possible. If the model were estimated via maximum likelihood, a likelihood ratio test could be used; in our case a DIC comparison could serve the same purpose. However, this type of test does not deny the presence of an alternative choice heuristic; it only indicates that it is not worth considering it in the model.

The traditional LC approach presents another problem if the objective is analysing whether a choice heuristic is present or not. The results of Chapter 7.2 suggest that the LC model presents an unstable behaviour regarding the precision of the results. For this, the proposed MHM is used to determine the plausibility of having any of the alternative heuristics in the sample.

To analyse the individual behaviour within our population, we test the MHM approach and compute the posterior probability of choosing the non-RUM choice heuristic. Following the results obtained in Section 7.3, if the non-RUM probability is high for a non-negligible proportion of the population, then the presence of a non-RUM choice heuristic is highly plausible.

As in the traditional LC approach, the most plausible alternative choice heuristic is the SS model. Any other choice heuristic presents a posterior maximum probability that does not surpass the RUM probability. Therefore, the MHM suggests that the SS might be present in the sample.

Several reasons could explain the absence of other non-RUM choice heuristics apart from the SS. First, RRM behave too similar to RUM as demonstrated in Chapter 6. It is also likely that the search engines eliminate the alternatives that are not attractive, so individuals might not need to choose by EBA. Indeed, the search engine could be the *eliminator by aspects* itself.

7.4. Conclusions on the Mixed Heuristics Model

Latent classes are the traditional approach for modelling multiple choice heuristics within a population. When applying this approach to a given context, there are three main challenges: i) class membership is hard to model, ii) the class membership level and the choice level must be modelled simultaneously, and iii) there is no certainty that the choice heuristics considered are being followed in the population. To tackle these challenges, we propose the Mixed Heuristic Model (MHM). The MHM can identify the presence of a choice heuristic, focusing only on the specification of the choice level, without modelling the class membership level.

We tested the properties of the MHM with synthetic data, showing how it can properly identify the choice heuristic followed by individuals. As the sample size and/or number of observations per individual increases, the MHM accuracy increases. Nevertheless, the model needs to be tested furthermore, by applying it to different contexts, to guarantee that this desirable behaviour holds across different samples.

The MHM was applied to analyse a real air travel SP survey. Five heuristics where tested as a complement to RUM. The MHM indicates that the presence of four of them in the sample is unlikely. Only the Stochastic Satisficing heuristic seems to be a plausible complement to RUM in this context. The MHM is more explicative than traditional LC models when indicating the potential presence of a choice heuristic: the MHM provides the distribution of the probabilities of the alternative choice heuristics. Additionally, unlike the traditional latent class approach, the MHM does not need to formulate a class membership function (one might argue that such a function does not exist on a real application), and as a result indicates that the alternative non-RUM heuristics are more plausible.

Finally, the MHM addresses the three stated challenges for considering multiple choice heuristics through a latent class approach but does not relate the individual characteristics to the probability of following a specific heuristic. Thus, the MHM does not replace the LC model –since it does not describe why heuristics are followed–, but it is an important complement for traditional approaches that may be used beforehand to select the most appropriate choice heuristics to be considered.

7.5. Publication History

This chapter was presented as a paper at the *International Choice Modelling Conference 2017* (Gonzalez-Valdes and Raveau, 2017b). Later, further versions were presented at the *Euro Working Group on Transportation Conference* (Gonzalez-Valdes and Raveau, 2017c) and at the *6th Symposium of the European Association for European Research in Transportation* (Gonzalez-Valdes and Raveau, 2017a). Finally, it was accepted for publication in the *Journal of Choice Modelling* (Gonzalez-Valdes and Raveau, 2018).

8. IDENTIFICATION VERSUS FORECASTING: COMPARING THE PERFORMANCE OF ALTERNATIVE MULTIPLE HEURISTICS MODELS UNDER WEAK AND STRONG IDENTIFIABILITY

Through the different chapters of this thesis, we have discussed the possibility of capturing the behaviour of individuals in a discrete choice model. First, in Chapters 5 and 6 we discussed the possibility of estimating the multiple heuristics model. Then, in Chapter 7 we studied the heuristic followed by DMs at an individual level. Indeed, Chapter 7 provided a methodology to identify the presence of different heuristics. However, we have not discussed, yet, how to select a model among several alternative models when the difference between them lies in the choice heuristic.

Several objectives might be pursued when estimating a model; for example, depending on the objective (explaining or forecasting) the metric used to choose the model may vary. In this chapter, we study the different techniques for model selection addressed in Section 2.3. These techniques evaluates the Bias-Variance trade-off (Hastie et al., 2001; McElreath, 2012), which indicates that more complex models might forecast worse if they present higher variance in their estimates. Our objective is to assess the usefulness of these techniques when several multiple heuristics models are available. Specifically, we analyse the case of the RUM and RRM heuristics.

In Chapter 6 we showed that the RUM and RRM behaved similarly. Therefore, it is not hard to find contexts in which any of those heuristics may perform with equally high likelihood. If there is a second heuristic available in the model, it might be even easier for the model to fill the gap that RUM or RRM fails to fill when interpreting the other heuristic. In this chapter, we analyse what statistical tools are required to identify the real underlying model in finite samples and how to evaluate its forecasting performance in the context of multiple choice heuristics. We analyse two multiple heuristics models, one containing RUM and SS, and the other RRM and SS. We chose SS because it is readily estimable unlike EBA.

In our experiment we test several in sample and out of sample statistical techniques to test the performance of the models in a simulated sample. The meta-experiment varies several dimensions to try and obtain conclusions as general as possible.

8.1. Experimental Design

We use the findings of the previous chapters to design an experiment where identifiability of the common class, SS, is not an issue. We fix several dimensions that promote the identifiability of the model either in relation to the individuals' behaviour and the choice scenario. Fixing these dimensions allows us to increase the dimensionality of the meta-experiment in the dimensions of the econometric tools; this way, we study in more depth these techniques to test their performance for different sample sizes and degrees of identifiability.

The behaviour simulated was designed to maximise the identifiability of the SS class, whilst also keeping interesting scenarios. The first meta-dimension is the heuristics simulated: we used either RRM-SS or RUM-SS. The simulation parameters are those used in Chapter 6 and detailed in Table 6-2.

The second meta-dimension was sample size. We tested three different sample sizes of ten, twenty, and forty thousand observations. For each combination of the meta-experiment, ten experiments were performed.

In all cases we maintained the same proportion between SS and the alternative heuristic; this was designed to slightly promote SS over the other heuristic. Since identifiability of SS requires larger sample sizes than RUM, we considered that the DMs used SS with 60% overall probability and either RUM or RUM with 40%. No correlation among the levels of the LC model was simulated.

As in previous experiments, we also considered a sociodemographic characteristic named *trait* that allowed variation of the probability of selecting a heuristic. The proportion of the characteristic in the sample was kept identical as in our previous experiments (i.e. fixed at 70%). The presence of *trait* implied a 64.3% propensity of selecting SS and of 50% if it was not present. These probabilities were designed to match the total probability of selecting SS with the desired 60% overall proportion explained above.

In all experiments we considered choice sets of size three coming from the "Las Condes – CBD, San Miguel – CBD" dataset. We also tested three in-sample and two out of sample techniques. In the former group we analysed the selection of the model based in the estimation likelihood and two information criteria: DIC and BIC. In the latter group of techniques, we tested out of sample validation and response analysis. Finally, regarding response analysis, nine different choice scenarios were analysed.

In estimation, we contrasted the performance of the RUM-SS and the RRM-SS models. We employed Bayesian estimation using the Gibbs sampling algorithm with 5,000 burn-in iterations and 10,000 samples per parameter in each of the two Markov chains employed.

8.2. Analysis of Results

Our objective was to test the performance of models that compete in terms of goodness of fit measures. Through the analysis of the model performance when using several measures, numerous conclusions may be extracted. First, we analysed in sample techniques; with these, we study how they can inform about the real underlying model in the case of weakly identified and strongly identified models. Then, we analysed out of sample techniques. These allow to elaborate conclusions regarding the forecasting performance of the models.

8.2.1. Model identifiability

Most of the results we will present and the conclusions regarding the performance of the models under the different testing scenarios find its origin in the quality of the estimation in each case. In a series of tables below we present model estimation results for each sample size: the mean estimator and the mean standard deviation of the parameters; no t-tests against the target parameters are presented, because in half of the cases the estimated heuristic does not correspond to the underlying heuristic, so no target parameters exist.

The results are separated by degree of identifiability. The RUM-SS cases present only strong identifiability. The RRM-SS model present two degrees of identifiability: weak and strong;

we present each identifiability case separately. Moreover, in the weak identifiability case we recognize two subdivisions: very weak and weak identifiability. In both cases the model parameters of the RRM class are non-identifiable. However, in the former group the posterior distribution of the parameters covers the whole prior distribution, whereas, in the latter (weak identifiability), the posterior distribution of the parameters also presents extreme variance (thus unidentifiable) but do not cover the whole posterior feasible range.

Table 8-1 presents the results of the RRM-SS and the RUM-SS models estimated for the 10,000 sample size experiment, with DMs actually choosing using the RRM-SS heuristic. In the case of the RRM-SS estimations, only extremely weak identifiable models were obtained. Conversely, the RUM-SS models always present strong identifiability. When contrasting both estimations for the SS class, both models exhibit similar estimates which are unbiased in comparison with the real underlying model (Table 6-2).

Recall that in the μ-RRM model, all parameters are multiplied by the reciprocal of μ. Therefore, the parameter value and the already important variance in the parameters is further increased on average 12.5 times due to the low value of the μ parameter.

Table 8-2 presents the results of the RRM-SS and RUM-SS estimated on the 10,000 sample size experiments with RUM-SS as the underlying behaviour. Surprisingly, the results of the RRM-SS model exhibit less identifiability issues than in the case where the underlying heuristic was RRM-SS.

Table 8-1 Parameters for the 10,000 sample experiment with RRM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters (standard deviation)	Very weak identifiability	Strong identifiability
SS class		
Cost sensitivity	-9.56 (0.89)	-10.18 (1.03)
Cost threshold	0.43 (0.03)	0.45 (0.03)
In-vehicle time sensitivity	-11.65 (1.00)	-12.22 (1.14)
Time threshold	0.59 (0.03)	0.6 (0.04)
MRS Vehicle - Waiting time	4.00 (0.28)	4.01 (0.29)
MRS Vehicle – Walking	1.68 (0.31)	1.67 (0.32)
Non-SS class		
μ	0.08 (0.05)	-
Cost sensitivity	-0.80 (0.53)	-1.09 (0.20)
In-vehicle Time sensitivity	-2.82 (1.77)	-3.51 (0.72)
Waiting time sensitivity	-13.27 (8.79)	-12.84 (2.77)
Walking time sensitivity	-5.22 (3.27)	-6.24 (0.96)
Class membership function		
SS base desirability	0.08 (0.18)	-0.05 (0.20)
Trait desirability for SS	0.59 (0.12)	0.61 (0.12)
Cases	10/10	10/10

Most cases present weak identifiability, whereas one case was non-identifiable at all. Like the previous case, the SS heuristic in both models presents similar parameter values. In the case of the RUM-SS the model is always strongly identified.

Table 8-2 Parameters for the 10,000 sample experiment with RUM-SS underlying heuristic

	RRM-SS		RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability
SS class			
Cost sensitivity	-9.01 (0.82)	-9.16 (0.81)	-10.18 (1.03)
Cost threshold	0.46 (0.03)	0.42 (0.04)	0.45 (0.03)
In-vehicle time sensitivity	-10.84 (0.81)	-11.29 (0.94)	-12.22 (1.14)
Time threshold	0.57 (0.03)	0.6 (0.04)	0.6 (0.04)
MRS Vehicle - Waiting time	3.86 (0.26)	4.06 (0.29)	4.01 (0.29)
MRS Vehicle – Walking	1.45 (0.25)	1.79 (0.33)	1.67 (0.32)
Non-SS class			
μ	0.20 (0.26)	0.54 (0.23)	-
Cost sensitivity	-0.56 (0.63)	-0.28 (0.37)	-1.09 (0.2)
In-vehicle Time sensitivity	-9.36 (6.63)	-4.51 (4.41)	-3.51 (0.72)
Waiting time sensitivity	-45 (32.87)	-20.42 (20.82)	-12.84 (2.77)
Walking time sensitivity	-14.05 (10.33)	-5.8 (5.64)	-6.24 (0.96)
Class membership function			
SS base desirability	0.21 (0.17)	0.12 (0.18)	-0.05 (0.2)
Trait desirability for SS	0.68 (0.12)	0.58 (0.12)	0.61 (0.12)
Cases	2/10	7/10	10/10

Tables 8-3 and 8-4 present the results of the twenty thousand sample experiment. These results mirror the ones obtained in the ten thousand sample experiment, but with a higher balance for less extreme variances. In the case of the RRM-SS, results suggest the second kind of weak balance is more prevalent in the sample, showing increased identifiability. Strong identifiability is achieved in only one of the estimations.

Table 8-3 Parameters for the 20,000 sample experiment with RRM-SS underlying heuristic

	RRM-SS			RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability	Strong identifiability
SS class				
Cost sensitivity	-9.68 (0.61)	-9.89 (0.68)	-9.88 (0.67)	-10.18 (1.03)
Cost threshold	0.43 (0.02)	0.46 (0.02)	0.44 (0.02)	0.45 (0.03)
In-vehicle time sensitivity	-11.68 (0.68)	-11.11 (0.72)	-12.2 (0.75)	-12.22 (1.14)
Time threshold	0.6 (0.02)	0.59 (0.03)	0.6 (0.02)	0.6 (0.04)
MRS Vehicle - Waiting time	4.07 (0.2)	4.02 (0.21)	3.94 (0.2)	4.01 (0.29)
MRS Vehicle – Walking	1.62 (0.21)	1.71 (0.23)	1.67 (0.22)	1.67 (0.32)
Non-SS class				
μ	0.05 (0.03)	0.15 (0.1)	0.21 (0.09)	-
Cost sensitivity	-0.64 (0.27)	-1.08 (0.89)	-0.68 (0.35)	-1.09 (0.2)
In-vehicle Time sensitivity	-2.51 (1.02)	-3.36 (2.88)	-2.13 (1.1)	-3.51 (0.72)
Waiting time sensitivity	-11.08 (4.74)	-13.61 (11.78)	-11.77 (6.58)	-12.84 (2.77)
Walking time sensitivity	-4.68 (1.93)	-6.55 (5.54)	-4.63 (2.4)	-6.24 (0.96)
Class membership function				
SS base desirability	0.02 (0.12)	0.04 (0.13)	0.01 (0.13)	-0.05 (0.2)
Trait desirability for SS	0.61 (0.08)	0.6 (0.09)	0.54 (0.08)	0.61 (0.12)
Cases	7/10	2/10	1/10	10/10

In these results, note that the μ parameter is higher in the case when the underlying heuristic is RUM-SS (Table 8-4) instead of RRM-SS (Table 8-3), this is consistent with the behaviour of the model. Recall that higher μ parameters are related to less “regretted” behaviour. Also, note that in the case of the RRM-SS underlying heuristic, the trend is that as the model is more identifiable, the μ value increases until it reaches the real underlying value of 0.2. Finally, mirroring the results of the ten thousand sample size estimations, the parameters of the SS class are similar in all models and close to the real underlying parameters. Therefore, we could infer that in these models the identified class is not biased due to the non-identification of the RRM class.

Tables 8-5 and 8-6 present the results of the forty thousand sample estimation experiment. This experiment is the first one that exhibits an important number of strong identifiable RRM-SS models. Indeed, six of the estimations were strongly identifiable. In the case of the RUM-SS model, the results show high identifiability with low standard deviations.

8.2.1. In sample techniques

We analyse three metrics that evaluate the goodness of fit of the estimated models. We study the likelihood in the estimation sample, the deviance information criterion, and the Bayesian information criterion. The latter two criteria penalise the likelihood depending on the number of estimated parameters. The RRM-SS model has one more parameter than the RUM-SS model due to the extra μ parameter in the RRM class.

Table 8-4 Parameters for the 20,000 sample experiment with RUM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters	Weak identifiability	Strong identifiability
SS class		
Cost sensitivity	-9.86 (0.63)	-10.05 (0.65)
Cost threshold	0.44 (0.02)	0.44 (0.02)
In-vehicle time sensitivity	-11.69 (0.66)	-11.87 (0.68)
Time threshold	0.6 (0.02)	0.6 (0.02)
MRS Vehicle - Waiting time	4.07 (0.2)	4.07 (0.2)
MRS Vehicle – Walking	1.6 (0.22)	1.59 (0.21)
Non-SS class		
μ	0.74 (0.74)	-
Cost sensitivity	-0.33 (-0.33)	-0.52 (0.13)
In-vehicle Time sensitivity	-3.38 (-3.38)	-5.07 (0.53)
Waiting time sensitivity	-13.49 (-13.49)	-18.66 (2.15)
Walking time sensitivity	-4.31 (-4.31)	-6.45 (0.58)
Class membership function		
SS base desirability	0.02 (0.11)	-0.02 (0.11)
Trait desirability for SS	0.58 (0.08)	0.58 (0.08)
Cases	10/10	10/10

Table 8-5 Parameters for the 40,000 sample experiment with RRM-SS underlying heuristic

	RRM-SS			RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability	Strong identifiability
SS class				
Cost sensitivity	-10 (0.42)	-9.65 (0.46)	-9.86 (0.47)	-10.1 (0.49)
Cost threshold	0.44 (0.01)	0.45 (0.02)	0.45 (0.02)	0.46 (0.02)
In-vehicle time sensitivity	-11.99 (0.46)	-11.45 (0.5)	-12.16 (0.51)	-12.25 (0.53)
Time threshold	0.57 (0.02)	0.61 (0.02)	0.61 (0.02)	0.61 (0.02)
MRS Vehicle - Waiting time	3.84 (0.13)	4.08 (0.14)	4.01 (0.14)	3.98 (0.14)
MRS Vehicle – Walking	1.61 (0.14)	1.67 (0.15)	1.56 (0.14)	1.62 (0.14)
Non-SS class				
μ	0.04 (0.02)	0.1 (0.06)	0.21 (0.08)	-
Cost sensitivity	-0.03 (0.01)	-0.93 (0.68)	-0.71 (0.27)	-1.13 (0.1)
In-vehicle Time sensitivity	-0.09 (0.04)	-3.2 (2.38)	-2.37 (0.92)	-3.44 (0.34)
Waiting time sensitivity	-0.45 (0.18)	-14.77 (11.35)	-10.94 (4.44)	-13.24 (1.28)
Walking time sensitivity	-0.21 (0.08)	-5.88 (4.35)	-4.45 (1.68)	-6.25 (0.44)
Class membership function				
SS base desirability	0.07 (0.08)	0.1 (0.09)	-0.04 (0.09)	-0.06 (0.09)
Trait desirability for SS	0.59 (0.06)	0.54 (0.06)	0.61 (0.06)	0.6 (0.06)
Cases	1/10	3/10	6/10	10/10

Table 8-6 Parameters for the 40,000 sample experiment with RUM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters	Strong identifiability	Strong identifiability
SS class		
Cost sensitivity	-9.88 (0.44)	-10.05 (0.65)
Cost threshold	0.45 (0.01)	0.44 (0.02)
In-vehicle time sensitivity	-11.85 (0.46)	-11.87 (0.68)
Time threshold	0.6 (0.02)	0.60 (0.02)
MRS Vehicle - Waiting time	4.04 (0.13)	4.07 (0.2)
MRS Vehicle – Walking	1.55 (0.14)	1.59 (0.21)
Non-SS class		
μ	0.88 (0.09)	-0.52 (0.13)
Cost sensitivity	-0.31 (0.07)	-5.07 (0.53)
In-vehicle Time sensitivity	-3.32 (0.47)	-18.66 (2.15)
Waiting time sensitivity	-13.34 (1.98)	-6.45 (0.58)
Walking time sensitivity	-4.07 (0.55)	-0.52 (0.13)
Class membership function		
SS base desirability	0.03 (0.08)	-0.02 (0.11)
Trait desirability for SS	0.57 (0.06)	0.58 (0.08)
Cases	10/10	10/10

Table 8-7 shows two pieces of information. In the first column, the number of cases where each metric chose the correct underlying heuristics. In the last column, the difference of each metric in terms of its mean and standard deviation. In the last column, we include the most frequent category of each RRM-SS model. In the case of underlying RRM-SS heuristics, for each sample size the categories analysed were the very weakly identified, very weakly identified, and strongly identified respectively. In the cases of underlying RRM-SS, we analysed for each sample size the cases were weakly identifiability, weakly identifiability, and strongly identifiability was obtained respectively.

Table 8-7 shows the difference of every metric between the RRM-SS and the RUM-SS. Therefore, as higher likelihood is desirable, positive differences in the likelihood metric promotes the RRM-SS model. Conversely, as lower DIC and BIC is desirable, negative differences in the DIC and BIC metric promotes the RRM-SS model. Results suggests that even though we compare very weakly identifiable models against a strongly identifiable one, the likelihood and DIC tend to promote the real underlying heuristic. BIC penalises much harder the additional parameter of the RRM-SS which only selects the real underlying heuristic at larger sample sizes.

We further explore the mean and standard deviation of the difference of the different metrics in Table 8-8. For this, we assume normal distributed differences of the different metrics and indicate the probability of choosing the wrong underlying model under each. We assume a null hypothesis of choosing a RRM-SS model. Therefore, the type I error indicates the probability of choosing RUM-SS when the underlying model is RRM-SS. The type II error indicates the probability of choosing RRM-SS model when the underlying one is RUM-SS.

Table 8-7 Goodness of fit for the ten, twenty and forty thousand sample size experiments

	Cases actual heuristic is chosen			Mean (SD) of the difference of		
Sample size	Likelihood	DIC	BIC	Likelihood	DIC	BIC
Underlying heuristic: RRM and SS						
10,000	$\frac{7}{10}$	$\frac{6}{10}$	$\frac{3}{10}$	2.39 (5.78)	-4.11 (12.5)	4.43 (11.5)
20,000	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{4}{10}$	7.97 (4.50)	-15.6 (8.98)	-6.04 (9.09)
40,000	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	12.07 (3.87)	-22.7 (7.81)	-13.6 (7.74)
Underlying heuristic: RUM and SS						
10,000	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	-4.74 (2.68)	10.6 (6.36)	18.7 (5.37)
20,000	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	-3.33 (1.25)	6.83 (2.37)	18.7 (5.4)
40,000	$\frac{10}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	-7.32 (2.44)	14.8 (4.94)	25.2 (4.89)

The degrees of freedom of the t-distribution vary depending on the number of experiments falling in the analysed category. Finally, to complement the analysis, the identifiability of the RRM-SS is presented (RUM-SS is always strongly identified).

Consistent with the results of Table 8-7, the error distribution of Table 8-8 indicates that the likelihood and DIC metrics report a similar error. These metrics support the real underlying heuristic even if it is very weakly identified; whereas, BIC only supports the RRM-SS for important differences in the likelihood. As expected, the error of all metrics decreases as the sample size increases. From these results, we conclude that likelihood and DIC could be useful in identifying the real underlying heuristic when two different models are available.

Table 8-8 Errors type I and II of the in sample metric studied

	Type I error (False negative)			Type II error (False positive)		
Metric \ Sample size	10,000	20,000	40,000	10,000	20,000	40,000
Likelihood	34.4%	6.4%	1.4%	5.5%	1.2%	0.7%
DIC	37.5%	6.6%	1.6%	6.5%	1.0%	0.8%
BIC	64.5%	26.3%	7.0%	0.4%	0.1%	0.01%
t distribution's degrees of freedom	9	6	5	6	9	9
Identifiability of RRM-SS	Very weak	Very weak	Strong	Weak	Weak	Strong

8.2.2. Out of sample techniques

In the previous subsection we analysed how in sample techniques may identify the real underlying heuristic. However, even though estimations may present higher likelihood, weakly and very weakly identifiable models present higher variances that may cause the model to underperform in forecasting. In this section we assess the forecasting performance of the models either in the same context estimated or in context variations through cross validation and response analysis techniques.

Both cross validation and response analysis were used with validation samples of size of 10,000 observations. As shown in section 2.3, large validation samples (as the one used here) reduce the variance of the estimation of the assessed metric. For testing cross validation and response analysis, the same estimations as those used in Table 8-8 are analysed. These are the most frequent outcome in each sample size. Therefore, in the case of cross validation up to ten different datasets were used for each sample size used in estimation; for each of them, the RRM-SS and RUM-SS models were tested. In the case of response analysis, for each combination of sample size used in estimation and response scenario, up to ten datasets were used. In these experiments, the metric calculated was the likelihood of the validation sample.

Table 8-9 presents the results of the cross validation. First, we present the average likelihood across the ten datasets for each of the models and, then, the likelihood difference.

Table 8-9 Cross validation results

Estimation Sample size	RRM-SS Likelihood (standard deviation)	RUM-SS Likelihood (standard deviation)	Likelihood difference (standard deviation)
Underlying heuristic: RRM and SS			
10,000	-8,522 (346)	-8,408 (85.3)	-114 (335)
20,000	-8,408 (104)	-8,394 (103)	-13.9 (18.3)
40,000	8,317 (86.0)	8,318 (84.8)	1.39 (3.27)
Underlying heuristic: RUM and SS			
10,000	-8,446 (154)	-8,363 (92.0)	-83.0 (114)
20,000	-8,343 (82.2)	-8,339 (80.7)	-4.09 (3.10)
40,000	-8,298 (86.1)	-8,297 (85.5)	-0.94 (1.11)

The results of Table 8-9 suggest that only the strongly identified RRM-SS outperforms the RUM-SS model when the underlying heuristic is, indeed, RRM-SS. When contrasted against Table 8-8, we note that even though very weakly identified results present higher likelihood when they coincide with the underlying heuristics, their forecasting performance may be worse. As forecasting performance may not be a good measure to identify the real underlying heuristic, the selected model may vary depending on the objective of the modeller.

To further analyse the forecasting performance of the models, we used the response analysis technique (Williams and Ortúzar, 1982a). Other studies have also used response analysis with the "Las Condes – CBD, San Miguel – CBD" dataset (Munizaga, 1997; Cantillo, 2004). We performed similar policy variations as these previous studies as detailed in Table 8-10.

Table 8-10 Policies for response analysis

Scenario	Cost private alternatives	Time private alternatives	Cost semi- private alternatives	Time semi -private alternatives
1	+25%	-	+12.5%	-
2	+50%	-	+25%	-
3	+100%	-	+50%	-
4	+25%	-15%	+12.5%	-7.5%
5	+50%	-15%	+25%	-7.5%
6	+100%	-15%	+50%	-7.5%
7	+25%	-30%	+12.5%	-15%
8	+50%	-30%	+25%	-15%
9	+100%	-30%	+50%	-15%

The objective of these policies is to identify how the model reacts against contextual variations through increases and decreases of attributes. Recall that the “Las Condes – CBD, San Miguel – CBD” datasets involve three types of modes: pure private, private and public combinations, and pure public modes. We tested policies (Table 8-10) that impact the cost and travel time of alternatives involving private modes. The cost policy increased the cost of private modes, whereas the time policies reduced the time of private modes. Mixed - modes were affected 50% less than pure private modes.

Table 8-11 presents the results of the response analysis for each policy. The base scenario with no policy shows the results of the cross validation; whereas, the rest shows the result of the respective policy. Each result considers between six to ten estimations according to the number of estimations that falls in the analysed category.

The results of the response analysis further exploits the findings of the cross validation. In the cases where the RRM-SS is the underlying heuristic and the sample size is either ten thousand or twenty thousand, the RRM-SS analysed is very weakly identified and the RUM-SS is strongly identified. The likelihood in the estimation sample promoted RRM-SS (which is the real underlying heuristic) over RUM-SS. Conversely, cross validation indicates that, despite RUM-SS not being the underlying heuristic, RUM-SS performs better when forecasting than RRM-SS. The reason behind this behaviour is the large variances obtained in the estimation of this last model. Therefore, we may conclude that the RUM-SS is more robust than the RRM-SS.

Table 8-11 Likelihood difference of the cross validation and response analysis

		Likelihood difference (standard deviation)					
Scenario \ Underlying heuristic		RRM-SS	RUM-SS	RRM-SS	RUM-SS	RRM-SS	RUM-SS
Cost	Time	10,000		20,000		40,000	
-	-	-114 (335)	-83.0 (114)	-13.9 (18.3)	-4.09 (3.10)	1.39 (3.27)	-0.94 (1.11)
+25	-	-126 (359)	-85.3 (109)	-13.4 (18.7)	-4.56 (3.73)	5.04 (4.72)	-0.34 (1.23)
+50	-	-131 (382)	-87.0 (109)	-13.4 (16.4)	-5.06 (4.01)	4.21 (2.46)	-0.27 (1.27)
+100	-	-139 (417)	-93.5 (113)	-9.69 (13.6)	-5.52 (3.55)	-0.32 (3.62)	-0.87 (1.70)
+25	-15	-124 (350)	-86.8 (108)	-14.8 (19.7)	-4.78 (3.24)	5.51 (4.29)	-0.51 (1.19)
+50	-15	-130 (374)	-85.7 (101)	-13.2 (17.8)	-4.43 (3.83)	4.00 (2.15)	-0.74 (1.26)
+100	-15	-137 (414)	-94.5 (109)	-10.2 (13.5)	-5.54 (3.67)	-0.81 (2.64)	-0.86 (1.56)
+25	-30	-129 (363)	-79.7 (92.4)	-14.4 (21.6)	-4.45 (2.92)	5.34 (5.08)	-0.77 (0.82)
+50	-30	-133 (379)	-83.4 (92.8)	-13.4 (20.0)	-4.20 (3.62)	5.33 (2.67)	-0.69 (1.42)
+100	-30	-134 (406)	-91.4 (101)	-8.47 (13.6)	-5.25 (3.52)	0.68 (2.08)	-1.25 (1.42)

For larger sample sizes, i.e. forty thousand samples, the strongly identified RRMSS competes against the strongly identified RUM-SS. In the case of the RRM-SS underlying heuristic, the model that estimates the real underlying heuristic presents higher likelihood, however larger deviations in the estimates (see Table 8-5) than the RUM-SS model. This may justify that as the policies present more extreme variations, the RRMSS performance decreases.

We further analysed the difference in likelihood in Table 8-12. Assuming that the likelihood difference distributes Normal, we examine the probability of Type I and Type II errors for each policy. We also show the degrees of freedom of the t distribution.

Table 8-12 Type I and Type II error probability for each response policy

		Error probability					
Scenario		Type I	Type II	Type I	Type II	Type I	Type II
Cost	Time	10,000		20,000		40,000	
-	-	62.8	24.6	76.2	11.7	34.5	21.7
+25	-	63.2	23.3	75	13.3	16.7	39.7
+50	-	62.9	22.8	77.7	12.6	7.4	41.8
+100	-	62.6	22.1	74.9	8.5	53.4	31.4
+25	-15	63.3	22.6	76.1	9.5	12.8	34.3
+50	-15	63.1	21.6	75.7	14.5	6.2	29.0
+100	-15	62.5	20.9	76.1	9.0	61.5	30.2
+25	-30	63.4	21.0	73.6	8.9	17.1	19.6
+50	-30	63.2	20.1	73.7	14.5	5.2	32.2
+100	-30	62.4	20.0	72.2	9.3	37.6	20.9
t distribution's degrees of freedom		9	6	6	9	5	9

The error distributions reported in Table 8-12 are consistent with the analysis done throughout this Chapter. Type I errors are high for the very weakly identifiable classes. Yet, surprisingly, the Type II errors (probability of choosing RRM-SS under RUM-SS) is non-dismissible even for these weakly identifiable cases.

The error distribution for the identifiable cases shows that even when identifiability is strong, smaller standard deviations (in the case of the RUM-SS) may generate smaller bias when forecasting. This is similar to the Bias-Variance trade-off phenomenon discussed in Section 2.3.2 (Hastie et al., 2001; McElreath, 2012). Even though the RUM-SS model might be biased, since it does not consider the real underlying heuristic, its smaller variance in the estimates generates more accurate forecasts. Our results suggest that this effect is enhanced in forecasting when the context changes.

8.3. Conclusions on Identification versus Forecasting

We analysed the Bias-Variance trade-off in the case where two models entailing different heuristics might compete in goodness of fit. In particular, we analysed the case where RUM-SS and RRM-SS are plausible candidates. To analyse the performance of the models three in-sample metrics (likelihood, DIC and BIC) and two out of sample techniques (cross validation and response analysis) were used.

In relation to the Bias-Variance trade-off, the likelihood only assesses the bias component; whereas DIC penalises the variance but weaker than BIC. Indeed, the former two exhibit similar behaviour given that the RRM-SS model has only one additional parameter; whereas

BIC penalised heavier the additional parameters given the sample sizes considered. The likelihood and DIC tended to identify the real underlying heuristic even when the model exhibited weak identifiability. On the other hand, BIC only identified the correct heuristic for the strong identifiability cases.

The out of sample techniques evaluate the Bias-Variance trade-off in different contexts. First, we performed a cross validation to evaluate the performance of the models in the same context where the model was estimated. In this case, the weakly identified models (RRM-SS) performed worse, even when they modelled the real underlying heuristic, when compared to its strongly identified model counterpart with the wrong heuristics (RUM-SS). When both models where strongly identified, each of them performed better when the underlying heuristic was coincident with the evaluated model. Therefore, the conclusions of the cross validation are consistent with those of the BIC analysis.

In the case of response analysis, we tested several different contexts varying positively and negatively different attributes. We conclude that as the context changes, the variance element in the Bias-Variance trade-off increases in importance. Indeed, the more the context changed, worse was the performance of the model with higher variance in its estimates (in our case the RRM-SS). Therefore, if the context changes importantly, a model with small deviations in its estimates might be preferable than a strongly identifiable model with the correct heuristics (identified due to higher likelihood) but with larger variance.

Our main conclusion relates the purpose of the model to the type of model selected and, therefore, the metric to be tested. If the objective is understanding the underlying process, then, less robust but more insightful models could be chosen. The associated metric to choose this kind of model is one that penalises less the extra parameters, like DIC. However, even if the underlying behaviour is correctly captured but the estimates pose big uncertainty, the forecasting performance might not be adequate. Therefore, if the objective is forecasting, our results suggest that more robust models should be chosen, even if it means not capturing the underlying mechanisms. To identify these models, ideally out of sample analysis should be performed; if not available due to data constraints, then, our analysis indicates that BIC could be a good candidate metric.

Even though our analysis gives interesting insight of the Bias-Variance trade-off phenomenon in multiple heuristic models, it should be taken cautiously given its limitations. We studied exclusively two combinations of choice heuristics. Indeed, given that the difference in the number of parameters was only one, few conclusions were obtained in the difference between likelihood and DIC. Further analysis must be done to completely understand the behaviour of multiple heuristics models in forecasting; this has only been the first step.

9. CONCLUSIONS

Discrete choice modelling is a powerful econometric tool designed for understanding and forecasting individual choices upon discrete choice sets. The kernel of a discrete choice model is its choice heuristic; understanding it is crucial to represent adequately decision makers' behaviour.

An extensive list of different choice heuristics has been proposed in the literature. To organise this list, we developed a framework that allows to organise the choice heuristics around three concepts: (i) how individuals process the numerous attributes of the alternatives, (ii) how they work with multiple alternatives, and -upon these two elements- (iii) how they build preferences. This framework enables us to understand the degree of similarity of various choice heuristics across different dimensions and to organise any choice heuristic around these three concepts, understanding how it relates to other choice heuristics. The heuristics analysed in this thesis were organised according to this framework.

Among the heuristics examined, we developed two contributions related with Elimination by Aspects (EBA) and Satisficing. EBA estimation is not straightforward; the best estimation procedure proposed in the literature considers the simulation of choice probabilities. In this thesis we developed an analytical approach that is exact - not as the current approach - and faster for low to mid complexity choice sets. Another important issue in the EBA model is the estimation of thresholds for continuous attributes. We presented a methodology that allows to estimate such thresholds. Although it does not guarantee finding the global optimum, we show that it gives good solutions which might perform even better than the underlying heuristic.

The Satisficing heuristic is a long standing one with clear principles stated in the work of Simon (1955). Surprisingly, before our work there was no model that wholly implemented the Satisficing principles exclusively with data about the alternative's profiles and the chosen alternative. We developed the Stochastic Satisficing (SS) model, which manages to use this normally available data and wholly implement Simon's principles. We analytically developed the SS model and explored its useful properties. This model allows to explain attribute saturation, attribute non-attendance, and estimates either constant or flexible marginal rates of substitution. Hence, this model is a now a new alternative in the practitioner's toolkit.

The core of this thesis was understanding the estimation of multiple heuristics discrete choice models. In practice, multiple heuristics models have shown several identifiability issues, which are even more extreme when the class membership function is more explicative than a simple constant population-wide (as in current practice).

To understand the identifiability phenomenon, we developed a theoretical framework which allows us to explain the empirical findings of our experiments. We start by analysing the first order condition of the optimization problem that produces the maximum likelihood estimates. Our analysis mathematically represents that to be able to identify several choice heuristics it is necessary that the increase in likelihood due to the incorporation of a new heuristic surpasses the loss of likelihood of decreasing the proportion of the previous heuristics. If this necessary condition is met, then the model might be estimated.

If the necessary condition of the framework is met, then the optimal point contains more than one heuristic; however, this does not guarantee that the optimal point is unique. To obtain unique and, therefore, identifiable estimates, the hessian matrix of the model must be non-singular and, ideally, with a large determinant. Our analysis indicates that the higher is the behavioural difference of the choice heuristics in the data, higher is identifiability of the model. Moreover, the higher the proportion of one choice heuristic and the larger is the sample size, higher is the plausibility of identifying it in the model. Finally, we were able to relate these three elements and quantify them into an expression that is readily interpretable.

The connection between the theoretical analysis and identifiability in practice was a graphical analysis. We analysed the behavioural difference of the tested choice heuristics in our dataset assuming parameter values for them. Our analysis indicated that in our choice context, the popular random utility maximization (RUM) heuristic's behaviour does not differ importantly from random regret minimization (RRM). On the other hand, EBA and SS have significant behavioural differences. Indeed, our graphical analysis correctly identified the identifiability of the multiple heuristics model analysed later.

We tested identifiability empirically for models considering RUM and either RRM, EBA, or SS heuristics. We analysed these models across several dimensions considering different proportion of choice heuristics, correlation structures and sample sizes. The theoretical analysis indicated the importance of the behavioural difference and the proportion of choice heuristics, which we confirmed as key precursors of identifiability. The correlation structure tested did not have an important influence in identifiability. Across all experiments RRM could never be identified from RUM, SS was highly identifiable under favourable conditions, and EBA was identifiable from RUM across all contexts. We finally tested a context where EBA, SS, and RUM were present simultaneously. We showed that in our favourable context tested, the three heuristics model can be correctly identified even with explicative class membership functions.

The latent class approach is useful for modelling the multiple heuristics case because it allows to understand the factors affecting the probability of choosing each heuristic and its sensitivities. However, the latent class model requires that the heuristics are properly selected and that both the class membership function and the heuristic's formulation are properly modelled; failing in formulating one of these elements may affect the performance of the whole model. We proposed the Mixed Heuristics Model as a way of understanding the heuristics present in the dataset and finding the correct formulation for them. The Mixed Heuristics Model was capable to successfully identify the underlying heuristics if they were properly modelled without modelling the class membership functions. Indeed, its accuracy is significantly higher than the latent class model approach.

Once several successful models are available, a criterion must be used to choose one among them. We concluded that depending of the objective of the modeller a different criterion must be chosen. If the objective is understanding underlying behaviour, a criterion that penalises weakly the additional parameters, such as DIC, might be desired. Whereas if the objective is forecasting, ideally an out of sample validation should be performed. If it is not possible due to the amount of data or further complexities, then BIC might be used.

Through our work, we show that it is possible to estimate multiple heuristics models. Also, we provide tools that allow to find the most explicative models and, among them, choose the most useful one. Therefore, after this thesis, the complexity surrounding the use of multiple heuristics models should decrease.

However, even though the thesis provides useful insights into multiple heuristics model estimation, it does not handle every aspect entailing this type of models. We did not explore complicated class membership functions in depth nor test extreme and diverse degrees of correlation between class membership level and choice level. Furthermore, more research should be done to understand the degree of identifiability in different contexts, so that general conclusions regarding the selection of heuristics can be obtained.

Finally, we expect that this thesis contributes to the use of models better suited for the wide variety of individuals in society. This way, by better representing people with different behaviour, public policy will be more effective.

REFERENCES

Adamowicz, W.L., Swait, J.D., 2013. Are food choices really habitual? Integrating habits, variety-seeking, and compensatory choice in a utility-maximizing framework. American Journal of Agricultural Economics 95(1), 17–41.

Adler, T., Falzarano, C., Spitz, G., 2005. Modeling service trade-offs in air itinerary choices. Transportation Research Record: Journal of the Transportation Research Board (1915), 20–26.

Aguiar, V.H., Boccardi, M.J., Dean, M., 2016. Satisficing and stochastic choice. Journal of Economic Theory 166, 445–482.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723.

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle, in: Petrov B.N., and Csaki, F. Editors, Second International Symposium on Information Theory. pp. 267–281.

Anderson, D.R., Burnham, K.P., White, G.C., 1998. Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. Journal of Applied Statistics 25(2), 263–282.

Araña, J.E., León, C.J., Hanemann, M.W., 2008. Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly. Journal of Health Economics 27(3), 753–769.

Aribarg, A., Otter, T., Zantedeschi, D., Allenby, G.M., Bentley, T., Curry, D.J., Dotson, M., Henderson, T., Honka, E., Kohli, R., Jedidi, K., Seiler, S., Wang, X., 2017. Advancing Non-compensatory Choice Models in Marketing. Customer Needs and Solutions.

Balbontin, C., Hensher, D. a., Collins, A.T., 2017. Integrating attribute non-attendance and value learning with risk attitudes and perceptual conditioning. Transportation Research Part E: Logistics and Transportation Review 97, 172–191.

Beck, M.J., Chorus, C.G., R., J.M. Hensher, D.A., 2013. Vehicle purchasing behavior of individuals and groups: Regret or reward? Journal of Transport Economics and Policy 47(3), 475–492.

Bekhor, S., Freund-Feinstein, U., 2006. Modeling passengers' preferences on a short-haul domestic airline with rank-ordered data. Transportation Research Record: Journal of the Transportation Research Board (1951), 1–6.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short-term travel decisions, in: Hall, R.W. (Ed.), Handbook of Transportation Science, International Series in Operations Research & Management Science. Springer US, Boston, MA, pp. 470–471.

Ben-Akiva, M., de Palma, A., McFadden, D., Abou-Zeid, M., Chiappori, P.A., de Lapparent, M., Durlauf, S.N., Fosgerau, M., Fukuda, D., Hess, S., Manski, C., Pakes, A., Picard, N., Walker, J., 2012. Process and context in choice models. Marketing Letters 23(2), 439–456.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of Choice and Latent Variable Models, in: In Perpetual Motion. Elsevier, pp. 431–470.

Berndt, E.R., Hall, B.H., Hall, R.E., Hausman, J., 1974. Estimation and Inference in Nonlinear Structural Models. Ann. Econ. Social Measurement 3, 653–665.

Bettman, J.R., 1979. An information processing theory of consumer choice. Addison-Wesley Pub, California.

Bhat, C.R., 2008. The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. Transportation Research Part B: Methodological 42(3), 274–303.

Bhat, C.R., 2005. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. Transportation Research Part B: Methodological 39(8), 679–707.

Bhat, C.R., 1998. Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice. Transportation Research Part B: Methodological 32(7), 455–466.

Bierlaire, M., Hurtubia, R., Flötteröd, G., 2010. Analysis of Implicit Choice Set Generation Using a Constrained Multinomial Logit Model. Transportation Research Record: Journal of the Transportation Research Board 2175, 92–97.

Boeri, M., Longo, A., Grisolia, J.M., Hutchinson, W.G., Kee, F., 2013. The role of regret minimisation in lifestyle choices affecting the risk of coronary heart disease. Journal of Health Economics 32(1), 253–260.

Breiman, L., Spector, P., 1992. Submodel Selection and Evaluation in Regression. The X-Random Case. International Statistical Review / Revue Internationale de Statistique 60(3), 291.

Calastri, C., Hess, S., Daly, A., Carrasco, J.A., 2017a. Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. Transportation Research Part A: Policy and Practice 104, 1–20.

Calastri, C., Hess, S., Daly, A., Maness, M., Kowald, M., Axhausen, K., 2017b. Modelling contact mode and frequency of interactions with social network members using the multiple discrete–continuous extreme value model. Transportation Research Part C: Emerging Technologies 76(December), 16–34.

Cantillo, V., 2004. Modelacion de demanda incorporando umbrales mínimos de percepción y valoración de atributos. Pontificia Universidad Católica de Chile.

Cantillo, V., Ortúzar, J. de D., 2005. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. Transportation Research Part B: Methodological 39(7), 641–657.

Caplin, A., Dean, M., 2011. Search, choice, and revealed preference. Theoretical Economics 6(1), 19–48.

Caplin, A., Dean, M., Martin, D., 2011. Search and Satisficing. American Economic Review 101(7), 2899–2922.

Cardell, N., Reddy, B., 1977. A multinomial logit model which permits variations in tastes across individuals. Charles Rivers Associates Inc.

Carrier, E., 2008. Modeling the choice of an airline itinerary and fare product using booking and seat availability data. Massachusetts Institute of Technology.

Cascetta, E., Papola, A., 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. Transportation Research Part C: Emerging Technologies 9(4), 249–263.

Castro, M., Martínez, F., Munizaga, M.A., 2013. Estimation of a constrained multinomial logit model. Transportation 40(3), 563–581.

Chassang, S., 2013. Calibrated incentive contracts. Econometrica 81(5), 1935–1971.

Chin, A.T., 2002. Impact of frequent flyer programs on the demand for air travel. Journal of Air Transportation 7(2), 53–86.

Chorus, C., Rose, J., 2012. Selecting a date: a matter of regret and compromises, in: Choice Modelling: The State of the Art and the State of Practice. Hess, S. y Daly, A. (Eds.), UK.

Chorus, C.G., 2014. A Generalized Random Regret Minimization model. Transportation Research Part B: Methodological 68, 224–238.

Chorus, C.G., 2010. A new model of random regret minimization. European Journal of Transport and Infrastructure Research 10(2), 181–196.

Chorus, C.G., Arentze, T.A., Timmermans, H.J.P., 2008. A Random Regret-Minimization model of travel choice. Transportation Research Part B: Methodological 42(1), 1–18.

Chorus, C.G., Bierlaire, M., 2013. An empirical comparison of travel choice models that capture preferences for compromise alternatives. Transportation 40(3), 549–562.

Conlisk, J., 2014. Why Bounded Rationality ? Journal of economic literature 34(2), 669–700.

Cox, D.R., 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society. Series B (… 24(2), 406–424.

Daganzo, C., 1979. Multinomial Probit: The Theory and Its Application to Demand Forecasting. New York: Academic Press.

Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39(1), 19–31.

Daly, A., Zachary, S., 1978. Improved multiple choice models, in: Determinants of Travel Choice. Saxon House, Westmead, UK.

Davidson, R., MacKinnon, J.G., 1981. Several Tests for Model Specification in the Presence of Alternative Hypotheses. Econometrica 49(3), 781.

Debreu, G., 1954. Representation of a preference ordering by a numerical function, in: Decision Processes. Thrall, Davis and Coombs, eds.

Denstadli, J.M., Lines, R., de Dios Ortúzar, J., 2012. Information processing in choice-based conjoint experiments. European Journal of Marketing 46(3/4), 422–446.

Donoso, P., Ortuzar, J. de D., 1982. Calibracion de modelos logit simple para el corredor Las Condes - Centro. Pontificia Universidad Católica de Chile.

Drabas, T., Wu, C.L., 2013. Modelling air carrier choices with a Segment Specific Cross Nested Logit model. Journal of Air Transport Management 32, 8–16.

Durbach, I., 2009. On the estimation of a satisficing model of choice using stochastic multicriteria acceptability analysis. Omega 37(3), 497–509.

Efron, B., Tibshirani, R., 1997. Improvements on cross-validation: The .632 plus bootstrap method. Journal of the American Statistical Association 92(438), 548.

Fader, P.S., McAlister, L., 1990. An Elimination by Aspects Model of Consumer Response to Promotion Calibrated on UPC Scanner Data. Journal of Marketing Research 27(3), 322–332.

Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2010. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Marketing Science 29(3), 393–421.

Foutz, R., Srivastava, R., 1977. The performance of the likelihood ratio test when the model is incorrect. The annals of Statistics 5(6), 1183–1194.

Fukushi, M., 2015. Detectando y Modelando El Efecto Decoy En Transporte. Universidad de los Andes.

Gabaix, X., Laibson, D., Moloche, G., Weinberg, S., 2006. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. American Economic Review 96(4), 1043–1068.

Gaudry, M.J.I., Jara-Diaz, S.R., Ortuzar, J. de D., 1989. Value of time sensitivity to model specification. Transportation Research Part B: Methodological 23(2), 151–158.

Gaundry, M.J.I., Dagenais, M.G., 1979. The dogit model. Transportation Research Part B: Methodological 13(2), 105–111.

Gelman, A., Carlin, J., Sern, H., Dunson, D., Vehtari, A., Rubin, D., 2013. Bayesian Data Analysis, 3rd editio. ed. Chapman & Hall/CRC.

Gensch, D.H., Ghose, S., 1992. Elimination by Dimensions. Journal of Marketing Research 29(4), 417–429.

Gilbride, T.J., Allenby, G.M., 2006. Estimating Heterogeneous EBA and Economic Screening Rule Choice Models. Marketing Science 25(5), 494–509.

Gilbride, T.J., Allenby, G.M., 2004. A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. Marketing Science.

Godoy, G., Ortuzar, J. de D., 2008. On the estimation of mixed logit models, in: Transportation Research Trends. pp. 289–310.

Gonzalez-Valdes, F., Heydecker, B.G., Ortuzar, J. de D., 2017. Identificabilidad de modelos de elección discreta con mecanismos de elección heterogéneos, in: 18º Congreso Chileno de Ingeniería de Transporte. La Serena-Coquimbo, Chile.

Gonzalez-Valdes, F., Heydecker, B.G., Ortúzar, J. de D., 2018. On the identifiability of multiple heuristic discrete choice models. In revision in Transportation Research Part B: Methodological.

Gonzalez-Valdes, F., Ortúzar, J. de D., 2017. The Stochastic Satisficing Model. International Choice Modelling Conference 2017, Cape Town, South Africa.

González-Valdés, F., Ortúzar, J. de D., 2018. The Stochastic Satisficing model: A bounded rationality discrete choice model. Journal of Choice Modelling 27, 74–87.

Gonzalez-Valdes, F., Raveau, S., 2018. Identifying he presence of heterogeneous discrete choice heuristics at an individual level. Journal of Choice Modelling (in press).

Gonzalez-Valdes, F., Raveau, S., 2017a. Identifying the presence of heterogeneous discrete choice mechanisms at an individual level, in: Symposium of the European Association for European Research in Transportation. Haifa, Israel.

Gonzalez-Valdes, F., Raveau, S., 2017b. Identifying indiviual choice mechanism profiles for air travel behaviour, in: 5th International Choice Modelling Conference. Cape Town, South Africa.

Gonzalez-Valdes, F., Raveau, S., 2017c. Modelling air travel behaviour with heterogeneous choice mechanisms at an individual level, in: Euro Working Group on Transportation. Budapest, Hungary.

Greene, M., Mora, R.I., Figueroa, C., Waintrub, N., de, J., 2017. Towards a sustainable city: Applying urban renewal incentives according to the social and urban characteristics of the area. Habitat International 68, 15–23.

Greene, W.H., 2003. Econometric Analysis, Fifth edit. ed. Prentice Hall, Upper Saddle River, New Jersey.

Grünwald, P., 2007. The Minimum Description Length Principle. MIT Press.

Guevara, C.A., 2016. Mode-valued differences of in-vehicle travel time savings. Transportation 44(5), 977–997.

Guevara, C.A., Chorus, C.G., Ben-akiva, M.E., 2016. Sampling of alternatives in random regret minimization models. Transportation Science 50(1), 306–321.

Guevara, C.A., Fukushi, M., 2016. Modeling the decoy effect with context-RUM Models: Diagrammatic analysis and empirical evidence from route choice SP and mode choice RP case studies. Transportation Research Part B: Methodological 93, 318–337.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY.

Henningsen, A., Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. Computational Statistics 26(3), 443–458.

Hensher, D. a., Collins, A.T., 2011. Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? Journal of Choice Modelling 4(1), 62–89.

Hensher, D.A., Greene, W.H., Chorus, C.G., 2013. Random regret minimization or random utility maximization: an exploratory analysis in the context of automobile fuel choice. Journal of Advanced Transportation 47(7), 667–678.

Hess, S., Stathopoulos, A., 2013. A mixed random utility - Random regret model linking the choice of decision rule to latent character traits. Journal of Choice Modelling 9(1), 27–38.

Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: An approach and four case studies. Transportation 39(3), 565–591.

Horowitz, J., 1982. Specification tests for probabilistic choice models. Transportation Research Part A: General 16(5–6), 383–394.

Hsiao, C., 1983. Chapter 4 Identification, in: Griliches, Z., Intrilligator, M. (Eds.), Handbook of Econometrics. North Holland, Amsterdam, pp. 223–283.

Hurvich, C.M., Tsai, C., 1989. Regression and time series model selection in small samples. Biometrika 76(2), 297–307.

Jang, S., Rasouli, S., Timmermans, H., 2017. Bias in random regret models due to measurement error: formal and empirical comparison with random utility model. Transportmetrica A: Transport Science 0(0), 1–30.

Jara-Díaz, S.R., Ortúzar, J. de D., 1989. Introducing the expenditure rate in the estimation of mode choice models. Journal of Transport Economics and Policy 23(3), 293–308.

Jedidi, K., Kohli, R., 2008. Inferring latent class lexicographic rules from choice data. Journal of Mathematical Psychology 52(4), 241–249.

Jedidi, K., Kohli, R., 2005. Probabilistic Subset-Conjunctive Models for Heterogeneous Consumers. Journal of Marketing Research 42(4), 483–494.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47(2), 263–292.

Kaplan, S., Prato, C.G., 2012. The application of the random regret minimization model to drivers' choice of crash avoidance maneuvers. Transportation Research Part F: Traffic Psychology and Behaviour 15(6), 699–709.

Kass, R.E., Carlin, B.P., Gelman, A., Neal, R.M., 1998. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. The American Statistician 52(2), 93.

Keane, M.P., Wolpin, K.I., 2007. Exploring the Usefulness of a Nonrandom Holdout Sample for Model Validation: Welfare Effects on Female Behavior*. International Economic Review 48(4), 1351–1378.

Kent, J., 1982. Robust properties of likelihood ratio tests. Biometrika 69(1), 19–27.

Kivetz, R., Netzer, O., Srinivasan, V., 2004. Alternative Models for Capturing the Compromise Effect. Journal of Marketing Research 41(3), 237–257.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference on Artificial Intelligence 14(12), 1137–1143.

Kohli, R., Jedidi, K., 2017. Relation Between EBA and Nested Logit Models. Operations Research 65(3), 621–634.

Kohli, R., Jedidi, K., 2015. Error Theory for Elimination by Aspects. Operations Research 63(3), 512–526.

Kullback, S., Leibler, R., 1951. On information and sufficiency. The annals of mathematical statistics.

Lancaster, K.J., 1966. A New Approach to Consumer Theory. Journal of Political Economy 74(2), 132.

Leong, W., 2014. Thesis: Embedding decision heuristics in discrete choice models: Assessing the merits of majority of confirming dimensions, extremeness aversion, and reference revision. The University of Sidney.

Leong, W., Hensher, D., 2012a. Is Route Choice a Matter of Regret Minimisation or Relative Advantage Maximisation? International Conference on Travel Behaviour Research (July), 15–20.

Leong, W., Hensher, D.A., 2012b. Embedding multiple heuristics into choice models: An exploratory analysis. Journal of Choice Modelling 5(3), 131–144.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. The BUGS Book. CRC Press, New York.

Manrai, A.K., Sinha, P., 1989. Elimination-By-Cutoffs. Marketing Science 8(2), 133–152.

Manski, C.F., 2017. Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment. Theory and Decision 83(2), 155–173.

Manski, C.F., 1977. The structure of random utility models. Theory and Decision 8(3), 229–254.

Manzini, P., Mariotti, M., 2014. Stochastic Choice and Consideration Sets. Econometrica 82(3), 1153–1176.

Martínez, F., Aguila, F., Hurtubia, R., 2009. The constrained multinomial logit: A semi-compensatory choice model. Transportation Research Part B: Methodological 43(3), 365–377.

Matzkin, R.L., 2007. Chapter 73: Nonparametric identification, in: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. North Holland, Amsterdam, pp. 5307–5368.

McCain, K.W., 2015. Mining full-text journal articles to assess obliteration by incorporation: Herbert A. Simon's concepts of bounded rationality and satisficing in economics, management, and psychology. Journal of the Association for Information Science and Technology 66(11), 2187–2201.

McElreath, R., 2012. Statistical Rethinking. UC Davis, California.

McFadden, D., 1981. Econometric Models of Probabilistic Choice, in: Structural Analysis of Discrete Daa with Econometric Applications. pp. 198–269.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior, in: Zarembka, P. (Ed.), Frontiers of Econometrics. New York: Academic Press, New York.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15(5), 447–470.

McFadden, D., Train, K., Tye, W.B., 1977. An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model. Transportation Research Record 637, 39–46.

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. Environmental and Resource Economics 51(4), 599–616.

Munizaga, M., 1997. Implicancias de la naturaleza de los datos en modelos de elección discreta. Pontificia Unviersidad Católica de Chile.

Newell, A., Simon, H.A., 1972. Human problem solving. NJ: Prentice-Hall, Englewood Cliffs.

Ortuzar, J. de D., Donoso, P., Hutt, G.A., 1983. Codificación, validación y evaluación de información para la estimación de modelos desagregados de elección discreta, in: IV Congreso Latinoamericano Sobre Métodos Computacionales En Ingeniería. Santiago.

Ortúzar, J. de D., Espinosa, A., 1986. Influencia del ingreso y la tasa de motorización en la partición modal para el viaje al trabajo, in: Tercer Congreso Latino-Iberoamericano de Investigación Operativa e Ingeniería de Sistemas. Santiago.

Ortúzar, J. de D., Fernández, J.E., 1985. On the stability of discrete choice models in different environments. Transportation Planning and Technology 10(3), 209–218.

Ortuzar, J. de D., Ivelic, M.A., 1987. Effect of Using More Accurately Measured Level-of-Service Variables on the Specification and Stability of Mode Choice Models, in: Proceedings of the 15th PTRC Summer Annual Meeting. PTRC Education and Research Services Ltd., London, pp. 117–130.

Ortuzar, J. de D., Willumsen, L., 2011. Modelling Transport. John Wiley & Sons.

Paleti, R., 2015. Implicit choice set generation in discrete choice models: Application to household auto ownership decisions. Transportation Research Part B: Methodological 80(683), 132–149.

Palma, D., Ortúzar, J. de D., Rizzi, L.I., Guevara, C.A., Casaubon, G., Ma, H., 2016. Modelling choice when price is a cue for quality a case study with Chinese wine consumers. Journal of Choice Modelling 19, 24–39.

Papi, M., 2012. Satisficing choice procedures. Journal of Economic Behavior & Organization 84(1), 451–462.

Peterson, R.A., Kerin, R., Ross, I., 1979. An information processing theory of consumer choice. Journal of Marketing 43(3), 124–126.

Pinjari, A.R., Bhat, C., 2010. A multiple discrete–continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. Transportation Research Part B: Methodological 44(4), 562–583.

Plummer, M., 2016. RJags: Bayesian Graphical Models using MCMC. R package version 4-6.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd Internaitional Workshop on Disbtributed Statistical Computing.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Radner, R., 1975. Satisficing, in: Optimization Techniques IFIP Technical Conference. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 252–263.

Raveau, S., Alvarez-Daziano, R., Francisca Yáñez, M., Bolduc, D., Ortúzar, J. de D., 2010. Sequential and Simultaneous Estimation of Hybrid Discrete Choice Models Some New Findings. Transportation Research Record (2156), 131–139.

Recker, W.W., Golob, T.F., 1979. A non-compensatory model of transportation behavior based on sequential consideration of attributes. Transportation Research Part B: Methodological 13(4), 269–280.

Rezaei, A., Puckett, S., Nassiri, H., 2011. Heterogeneity in Preferences of Air Travel Itinerary in a Low-Frequency Market. Transportation Research Record: Journal of the Transportation Research Board (2214), 10–19.

Richardson, A., 1982. Search models and choice set generation. Transportation Research Part A: General 16(5–6), 403–419.

Rothenberg, T., 1971. Identification in parametric models. Econometrica 39(3), 577–591.

Russo, J.E., Dosher, B.A., 1983. Strategies for multiattribute binary choice. Journal of Experimental Psychology: Learning, Memory, and Cognition 9(4), 676–696.

Saelensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological Economics 59(3), 331–340.

Savage, L.J., 1951. The Theory of Statistical Decision. Journal of the American Statistical Association 46(253), 55–67.

Schwarz, G.E., 1978. Estimating the dimension of a model. Annal of Statistics 6, 461–464.

Simon, H.A., 1956. Rational choice and the structure of the environment. Psychological Review 63(2), 129–138.

Simon, H.A., 1955. A behavioral model of rational choice. The Quarterly Journal of Economics 69(1), 99–118.

Simonson, I., 1992. The Influence of Anticipatin Regret and Responsibility on Purchase Decisions. Journal of Consumer Research 19(1), 105–118.

Stan Development Team, 2015. Stan: A C++ Library for Probability and Sampling. Version 2.8.0, .

Stüttgen, P., Boatwright, P., Monroe, R.T., 2012. A satisficing choice model. Marketing Science 31(6), 878–899.

Theis, G., Adler, T., Clarke, J.P., Ben-Akiva, M., 2006. Risk aversion to short connections in airline itinerary choice. Transportation Research Record: Journal of the Transportation Research Board (1951), 28–36.

Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Train, K.E., 1998. Recreation Demand Models with Taste Differences over People. Land Economics 74(2), 230–239.

Tversky, A., 1972a. Choice by elimination. Journal of Mathematical Psychology 9(4), 341–367.

Tversky, A., 1972b. Elimination by aspects: a theory of choice. Psychological Review 79(4), 281–299.

Tversky, A., 1969. Intransitivity of preferences. Psychological Review 76(1), 31–48.

Tversky, A., Kahneman, D., 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. The Quarterly Journal of Economics 106(4), 1039–1061.

Tversky, A., Sattath, S., 1979. Preference trees. Psychological Review 86(6), 542–573.

Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgment and choice. Psychological Review 95(3), 371–384.

Tversky, A., Simonson, I., 1993. Context-Dependent Preferences. Management Science 39(10), 1179–1189.

Tyson, C.J., 2008. Cognitive constraints, contraction consistency, and the satisficing criterion. Journal of Economic Theory 138(1), 51–70.

van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transportation Research Part A: Policy and Practice 74, 91–109.

van Eggermond, M.A., 2007. Consumer choice behaviour and strategies of air transportation service providers. Delft University of Technology.

Vovsha, P., 1997. Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. Transportation Research Record: Journal of the Transportation Research Board 1607, 6–15.

Vuong, Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica 57(2), 307.

Wallace, C.S., 2005. Statistical and Inductive Inference by Minimum Message Length, Information Science and Statistics. Springer-Verlag, New York.

Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Journal of Machine Learning Research 11(Dec), 3571–3594.

Wen, C.H., Lai, S.C., 2010. Latent class models of international air carrier choice. Transportation Research Part E: Logistics and Transportation Review 46(2), 211–221.

White, H., 1982. Maximum Likelihood Estimation of Misspecified Models. Econometrica 50(1), 1–25.

Williams, H.C.W.L., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. Environment and Planning A 9(3), 285–344.

Williams, H.C.W.L., Ortuzar, J. de D., 1982a. Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research Part B: Methodological 16(3), 167–219.

Williams, H.C.W.L., Ortuzar, J. de D., 1982b. Travel demand and response analysis—some integrating themes. Transportation Research Part A: General.

Young, W., Richardson, A.J., Ogden, K.W., Rattray, A.L., 1983. An inter-urban freight mode choice model. Transportation Planning and Technology 8(1), 61–80.

Zeelenberg, M., 1999. The use of crying over spilled milk: A note on the rationality and functionality of regret. Philosophical Psychology 12(3), 325–340.

Zeelenberg, M., Pieters, R., 2007. A Theory of Regret Regulation 1.1. Journal of Consumer Psychology 17(1), 29–35.

Zhao, C.-L., Huang, H.-J., 2016. Experiment of boundedly rational route choice behavior and the model under satisficing rule. Transportation Research Part C: Emerging Technologies 68, 22–37.

APPENDICES

Appendix A Glossary of terms

In this thesis the nomenclature used is presented in Table A-1.

Table A-1 Nomenclature

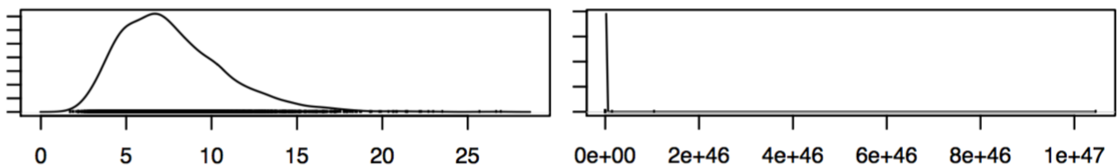
Term	Meaning
q, Q	DM or set of DMs
i, I	Alternative or set of alternatives
j, J	Secondary alternative or set of secondary alternatives
k, K	Alternative’s attribute or set of attributes
s, S	Simulation or set of simulations
h, H	Choice heuristic or set of heuristics
β	Choice heuristic’s parameters
θ	Class membership function parameters
c, C	Choice set or sets of choice sets.

Appendix B Identification of the EBA model

In this thesis we analyse the identification of multiple choice heuristic models. To study it, the identification of a single choice heuristic should not be a problem. Unfortunately, the identifiability of some choice heuristics is not straightforward which is the case of the EBA. Different phenomena can cause non- identifiability of a model such as correlated variables or an inefficient design. For the most common models, such as linear regression models, the relationship between correlation and identification has already been studied (Greene, chap. 4, 2003); however, for the EBA model the phenomenon under the identifiability is blur. In the case of the EBA model we do not expect to develop a methodology, rather show different EBA structure tested showing that the one used is an identifiable one.

For this analysis, we estimate the EBA model using Bayesian estimation. In this framework the non-identifiability is identified as obtaining posterior distributions extremely biased and with high coefficient of variation. The diagnosis is evident when plotting the density function of the parameters. For example, Figure B-1 shows the estimated posterior density function of a parameter. The left panel shows an identified parameter, whereas the right panel shows a non-identified parameter which range tends to infinity. When the parameter is identifiable, the density function is described in a reasonable range. Conversely, when the parameter is not identifiable the posterior samples through a big scope of unfeasible values.

Figure B-1 Posterior density function of the estimated parameters



We run several EBA models finding no identifiability. Correlation between the aspects was discarded since no correlation out bounded the 40%. Moreover, higher correlation among parameters had no relationship with the identified parameters. Several EBA experiments where done to determine the reason of certain non-identifiability as stated in Table B-1. The last model was simulated starting only from alternative specific constants and then gradually incorporating additional parameters. Results suggests that the nature of the identifiability is that certain weights were oversized.

Table B-1 EBA model parameters and mean identifiability of parameters

Variable		Model A	Model B	Model C	Model D
Thresholds	In-vehicle time	15 min	15 min	15 min	15 min
	Wait time	5 min	3 min	2 min	3 min
	Walk time	5 min	5 min	5 min	5 min
	Cost	60 and 100 CLP	50 and 90 CLP	60 and 100 CLP	40 and 60 CLP
Weights	In-vehicle time	8	8	8	4
	Wait time	16	16	16	8
	Walk time	28	20	20	10
	Cost	8 and 20	4 and 20	4 and 20	4 and 4
	ASC 1 to 9	1.0 to 1.8	1.0 to 1.8	1.0 to 1.8	1.0 to 1.8
Mean identifiability of parameters		6.8/13	6.5/13	6.8/13	13/13

Appendix C Detailed results on Singapore travel models

In this appendix the detail of the results of Singapore air travel model are presented. The model that considers EBA and RUM heuristics are presented in Table C-1. The results of the second and third model, considering EBA-RUM stops and EBA-RUM fare as alternative mechanisms to RUM are presented in Tables C-2 and C-3 respectively. The fourth model that contains RRM and RUM is presented in Table C-4. The results of the fifth model containing SS and RUM are presented in Table C-5. Finally, the sixth and last model is the base RUM model, which results are presented in Table C-6.

Table C-1 Model parameters for the latent class and mixed heuristic models. Heuristics RUM and EBA being represented.

	LC		MH	
Parameter	Value	SD	Value	SD
RUM – Cost	-7.63	0.62	-7.61	0.64
RUM – Time	-0.35	0.11	-0.36	0.11
RUM – Stop time	-0.71	0.08	-0.70	0.08
EBA – No stops	4.13	0.46	4.17	0.43
EBA – Cost 1	2.13	1.54	1.87	1.56
EBA – Cost 2	3.37	1.00	3.32	1.01
EBA – Regular carrier (fixed)	0	-	0	-
RUM - μ base	2.88	0.61	3.09	0.58
RUM - σ standard deviation	-	-	0.83	0.53

Table C-2 Model parameters for the latent class and mixed heuristic models. Heuristics RUM and EBA-RUM stops being represented.

	LC		MH	
Parameter	Value	SD	Value	SD
RUM – Cost	-8.31	0.75	-7.74	1.04
RUM – Time	-3.26	1.19	-0.33	0.12
RUM – Stop time	-7.10	0.87	-0.65	0.10
EBA – RUM Cost	-5.35	7.18	-12.45	8.70
EBA – RUM Time	0.38	9.07	0.01	0.73
RUM - μ base	2.64	0.09	2.54	0.92
RUM - σ standard deviation	-	-	1.87	0.80

Table C-3 Model parameters for the latent class and mixed heuristic models. Heuristics RUM and EBA-RUM fare being represented.

	LC		MH	
Parameter	Value	SD	Value	SD
RUM – Cost	-8.20	0.63	-7.71	1.02
RUM – Time	-3.54	1.11	-0.36	0.11
RUM – Stop time	-7.70	0.91	-0.70	0.08
EBA – RUM Cost	-1.13	2.90	4.17	0.43
EBA – RUM Time	-0.86	2.95	1.87	1.56
EBA – RUM Stop time	-1.52	2.65	1.87	1.56
RUM - μ base	3.28	0.59	3.09	0.58
RUM - σ standard deviation	-	-	0.83	0.53

Table C-4 Model parameters for the latent class and mixed heuristic models. Heuristics RUM and RRM fare being represented.

	LC		MH	
Parameter	Value	SD	Value	SD
RUM – Cost	-8.70	1.16	-8.71	0.64
RUM – Time	-0.35	0.12	-0.35	0.12
RUM – Stop time	-0.79	0.08	-0.81	0.08
RRM – Cost	-0.92	1.58	0.81	1.60
RRM – Time	36.3	32.6	41.6	24.4
RRM – Stop time	-16.1	13.4	16.5	0.08
RUM - μ base	3.71	0.24	3.79	0.20
RUM - σ standard deviation	-	-	0.50	0.31

Table C-5 Model parameters for the latent class and mixed heuristic models. Heuristics RUM and SS being represented.

	LC		MH	
Parameter	Value	SD	Value	SD
RUM – Cost	-9.51	0.99	-6.13	1.31
RUM – Time	-0.40	0.13	-0.29	0.12
RUM – Stop time	-0.88	0.11	-0.70	0.10
SS – Cost sensitivity	-1.93	2.02	-0.38	0.76
SS – Cost threshold	0.11	3.26	0.17	3.20
SS – Time sensitivity	3.83	2.05	-1.08	1.01
SS – MRS Time & Stop time	2.35	0.81	2.45	1.15
SS – Time threshold	1.39	2.37	-0.83	2.77
RUM - μ base	2.34	0.44	2.63	3.60
RUM - σ standard deviation	-	-	2.37	0.47

Table C-6 Model parameters for the RUM model

Parameter	Value	SD
RUM – Cost	-8.11	0.57
RUM – Time	-3.27	1.04
RUM – Stop time	-7.34	0.73

Appendix D Constants of the experiment of identifiability versus forecasting

In this Appendix, we detail the alternative specific constants of the models detailed in Chapter 8.

Table D-1 Model alternative specific constants and degree of identifiability for the 10,000 sample size experiment with RRM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters	Very weak identifiability	Strong identifiability
	SS class	
ASC1	-0.23 (0.75)	-0.33 (0.66)
ASC2	-1 (fixed)	-1 (fixed)
ASC3	-0.49 (0.7)	-0.64 (0.67)
ASC4	0.35 (1.1)	0.63 (1.22)
ASC5	-0.5 (0.37)	-0.59 (0.36)
ASC6	0.16 (1)	-0.13 (0.86)
ASC7	-0.66 (0.35)	-0.72 (0.4)
ASC8	0.11 (1.08)	-0.21 (0.86)
ASC9	0.01 (0.84)	-0.16 (0.69)
	Non-SS class	
ASC1	0.03 (0.02)	0.6 (0.24)
ASC2	0 (0)	0 (0)
ASC3	0 (0.01)	0.18 (0.2)

ASC4	0.05 (0.02)	0.57 (0.25)
ASC5	0.05 (0.01)	0.77 (0.19)
ASC6	0.04 (0.02)	0.69 (0.23)
ASC7	0.01 (0.02)	0.22 (0.33)
ASC8	0.01 (0.02)	0.35 (0.23)
ASC9	0.01 (0.02)	0.47 (0.23)

Table D-2 Model alternative specific constants and degree of identifiability for the 10,000 sample size experiment with RUM-SS underlying heuristic

	RRM-SS		RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability
	SS class		
ASC1	-0.22 (0.42)	-0.4 (0.76)	-0.33 (0.66)
ASC2	-1 (0)	-1 (0)	-1 (0)
ASC3	-0.8 (0.33)	-0.68 (0.6)	-0.64 (0.67)
ASC4	-0.52 (0.37)	-0.7 (0.47)	0.63 (1.22)
ASC5	-0.58 (0.26)	-0.66 (0.35)	-0.59 (0.36)
ASC6	-0.6 (0.35)	-0.56 (0.67)	-0.13 (0.86)
ASC7	-0.69 (0.28)	-0.81 (0.4)	-0.72 (0.4)
ASC8	-0.81 (0.33)	-0.44 (0.88)	-0.21 (0.86)
ASC9	-0.52 (0.34)	-0.36 (0.74)	-0.16 (0.69)
	Non-SS class		

ASC1	0.02 (0.04)	0.2 (0.13)	0.6 (0.24)
ASC2	0 (0)	0 (0)	0 (0)
ASC3	-0.06 (0.04)	-0.02 (0.11)	0.18 (0.2)
ASC4	0.15 (0.05)	0.47 (0.15)	0.57 (0.25)
ASC5	0.06 (0.04)	0.34 (0.1)	0.77 (0.19)
ASC6	0 (0.05)	0.27 (0.13)	0.69 (0.23)
ASC7	-0.09 (0.06)	0.02 (0.18)	0.22 (0.33)
ASC8	-0.06 (0.05)	0.05 (0.14)	0.35 (0.23)
ASC9	-0.06 (0.05)	0.1 (0.14)	0.47 (0.23)

Table D-3 Model alternative specific constants and degree of identifiability for the 20,000 sample size experiment with RRM-SS underlying heuristic

	RRM-SS			RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability	Strong identifiability
		SS class		
ASC1	-0.61 (0.26)	-0.75 (0.24)	0.05 (0.03)	-0.81 (0.25)
ASC2	-1 (fixed)	-1 (fixed)	0 (fixed)	-1 (fixed)
ASC3	-0.72 (0.25)	-0.96 (0.23)	-0.02 (0.03)	-0.94 (0.24)
ASC4	-0.43 (0.26)	-0.57 (0.25)	0.11 (0.04)	-0.38 (0.29)
ASC5	-0.58 (0.18)	-0.62 (0.18)	0.1 (0.03)	-0.67 (0.19)
ASC6	-0.38 (0.28)	-0.76 (0.24)	0.08 (0.03)	-0.68 (0.26)
ASC7	-0.86 (0.2)	-0.72 (0.21)	-0.04 (0.04)	-0.84 (0.21)

ASC8	-0.57 (0.25)	-0.8 (0.24)	0.03 (0.03)	-0.82 (0.24)
ASC9	-0.57 (0.24)	-0.81 (0.22)	0.03 (0.03)	-0.78 (0.23)
		Non-SS class		
ASC1	0.02 (0.01)	0.07 (0.02)	-0.67 (0.25)	0.5 (0.16)
ASC2	0 (fixed)	0 (0)	-1 (0)	0 (fixed)
ASC3	0 (0.01)	0.01 (0.02)	-0.71 (0.24)	0.14 (0.13)
ASC4	0.03 (0.01)	0.08 (0.02)	-0.29 (0.29)	0.49 (0.16)
ASC5	0.03 (0.01)	0.08 (0.02)	-0.64 (0.18)	0.67 (0.12)
ASC6	0.02 (0.01)	0.07 (0.02)	-0.74 (0.24)	0.6 (0.15)
ASC7	0.01 (0.01)	0 (0.03)	-0.4 (0.24)	0.21 (0.16)
ASC8	0 (0.01)	0.01 (0.02)	-0.74 (0.23)	0.31 (0.15)
ASC9	0.01 (0.01)	0.04 (0.02)	-0.51 (0.23)	0.45 (0.15)

Table D-4 Model alternative specific constants and degree of identifiability for the 20,000 sample size experiment with RUM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters	Weak identifiability	Strong identifiability
	SS class	
ASC1	-0.75 (0.24)	-0.72 (0.24)
ASC2	-1 (fixed)	-1 (fixed)
ASC3	-0.96 (0.23)	-0.82 (0.24)
ASC4	-0.57 (0.25)	-0.49 (0.26)
ASC5	-0.62 (0.18)	-0.69 (0.18)

ASC6	-0.76 (0.24)	-0.65 (0.24)
ASC7	-0.72 (0.21)	-0.81 (0.21)
ASC8	-0.8 (0.24)	-0.71 (0.24)
ASC9	-0.81 (0.22)	-0.64 (0.23)
	Non-SS class	
ASC1	0.07 (0.02)	0.42 (0.15)
ASC2	0 (fixed)	0 (fixed)
ASC3	0.01 (0.02)	0.01 (0.12)
ASC4	0.08 (0.02)	0.62 (0.16)
ASC5	0.08 (0.02)	0.6 (0.12)
ASC6	0.07 (0.02)	0.5 (0.15)
ASC7	0 (0.03)	0.06 (0.18)
ASC8	0.01 (0.02)	0.17 (0.16)
ASC9	0.04 (0.02)	0.28 (0.16)

Table D-5 Model alternative specific constants and degree of identifiability for the 40,000 sample size experiment with RRM-SS underlying heuristic

	RRM-SS			RUM-SS
Parameters	Very weak identifiability	Weak identifiability	Strong identifiability	Strong identifiability
		SS class		
ASC1	-0.89 (0.15)	-0.87 (0.16)	-0.82 (0.17)	-0.97 (0.16)
ASC2	-1 (fixed)	-1 (fixed)	-1 (fixed)	-1 (fixed)
ASC3	-0.87 (0.15)	-0.91 (0.16)	-0.89 (0.16)	-1.03 (0.16)
ASC4	-0.74 (0.16)	-0.68 (0.16)	-0.71 (0.16)	-0.64 (0.17)

ASC5	-0.78 (0.11)	-0.8 (0.12)	-0.74 (0.12)	-0.81 (0.12)
ASC6	-0.66 (0.16)	-0.76 (0.17)	-0.81 (0.16)	-0.92 (0.16)
ASC7	-0.9 (0.14)	-0.86 (0.14)	-0.92 (0.14)	-0.94 (0.14)
ASC8	-0.77 (0.15)	-0.8 (0.16)	-0.79 (0.16)	-0.93 (0.15)
ASC9	-0.85 (0.14)	-0.71 (0.15)	-0.78 (0.15)	-0.89 (0.15)
		Non-SS class		
ASC1	1.03e-3 (0.22 e-3)	5.22e-2 (10.93 e-2)	0.11 (0.02)	0.64 (0.11)
ASC2	0 (fixed)	0 (fixed)	0 (fixed)	0 (fixed)
ASC3	0.26e-3 (0.18 e-3)	0.5e-2 (9.15 e-2)	0.01 (0.02)	0.25 (0.09)
ASC4	1.66e-3 (0.23 e-3)	6.98e-2 (11.37 e-2)	0.16 (0.02)	0.64 (0.11)
ASC5	1.36e-3 (0.16 e-3)	6.71e-2 (8.31 e-2)	0.15 (0.02)	0.79 (0.08)
ASC6	1.1e-3 (0.2 e-3)	5.55e-2 (10.92 e-2)	0.13 (0.02)	0.8 (0.1)
ASC7	0.49e-3 (0.22 e-3)	1.46e-2 (11.63 e-2)	0.06 (0.02)	0.33 (0.11)
ASC8	0.46e-3 (0.21 e-3)	2.24e-2 (10.95 e-2)	0.05 (0.02)	0.43 (0.11)
ASC9	0.67e-3 (0.2 e-3)	2.99e-2 (10.75 e-2)	0.07 (0.02)	0.55 (0.1)

Table D-6 Model alternative specific constants and degree of identifiability for the 40,000 sample size experiment with RUM-SS underlying heuristic

	RRM-SS	RUM-SS
Parameters	Strong identifiability	Strong identifiability
	SS class	
ASC1	-0.77 (0.16)	-0.82 (0.16)
ASC2	-1 (fixed)	-1 (fixed)
ASC3	-0.89 (0.15)	-0.94 (0.15)
ASC4	-0.74 (0.16)	-0.76 (0.16)
ASC5	-0.73 (0.11)	-0.77 (0.12)
ASC6	-0.75 (0.16)	-0.81 (0.16)
ASC7	-0.88 (0.13)	-0.91 (0.13)
ASC8	-0.77 (0.15)	-0.84 (0.15)
ASC9	-0.77 (0.14)	-0.84 (0.14)
	Non-SS class	
ASC1	0.4 (0.09)	0.5 (0.11)
ASC2	0 (fixed)	0 (fixed)
ASC3	0.03 (0.08)	0.1 (0.09)
ASC4	0.7 (0.1)	0.8 (0.12)
ASC5	0.58 (0.07)	0.71 (0.08)
ASC6	0.47 (0.09)	0.62 (0.11)
ASC7	0.14 (0.11)	0.24 (0.12)
ASC8	0.14 (0.1)	0.26 (0.11)
ASC9	0.27 (0.1)	0.41 (0.11)