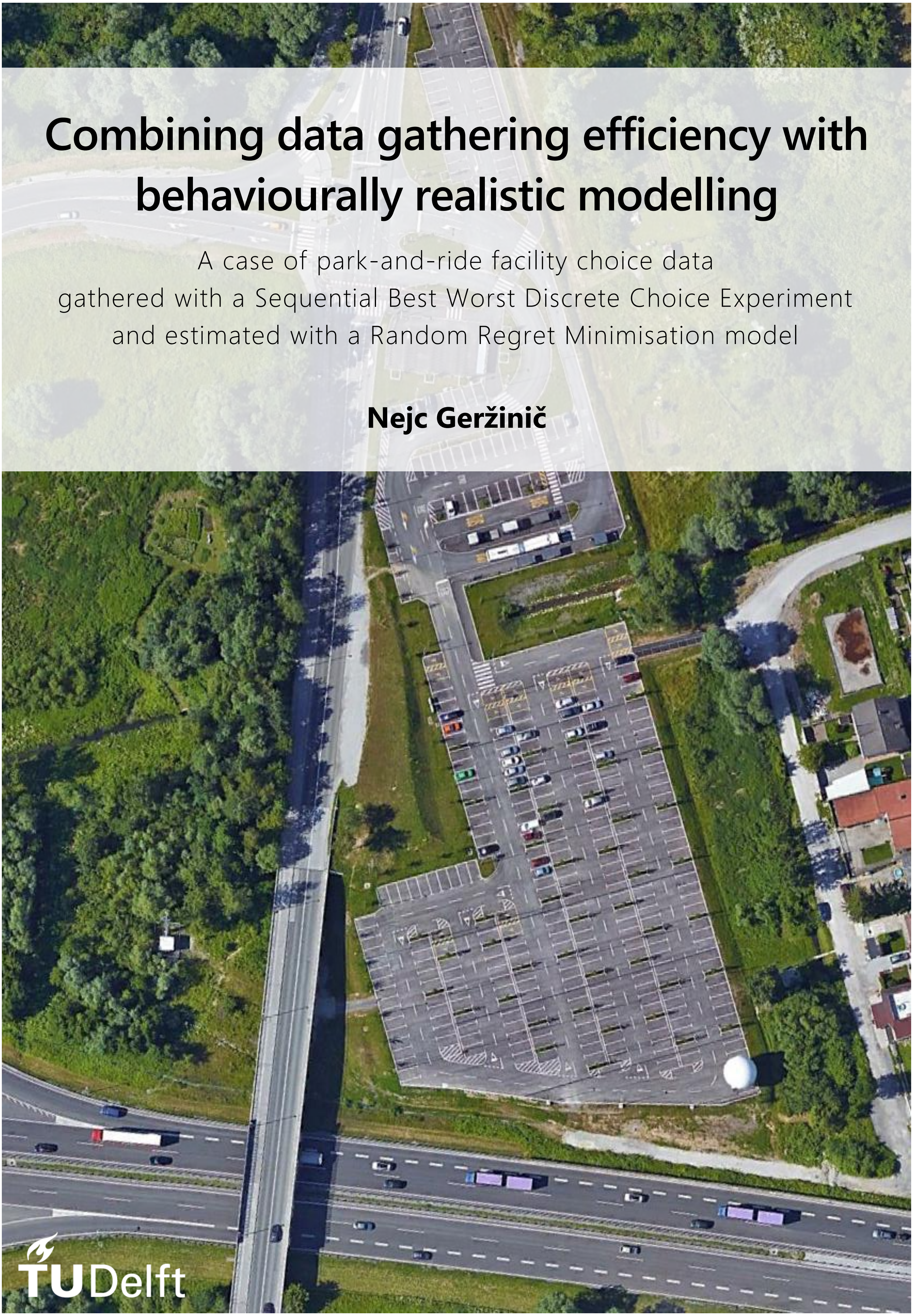


Combining data gathering efficiency with behaviourally realistic modelling

A case of park-and-ride facility choice data gathered with a Sequential Best Worst Discrete Choice Experiment and estimated with a Random Regret Minimisation model

Nejc Geržinič

TU Delft

Title image obtained from Google Earth (2016)

Combining data gathering efficiency with behaviourally realistic modelling

A case of park-and-ride facility choice data gathered with a Sequential Best Worst Discrete Choice Experiment and estimated with a Random Regret Minimisation model

Nejc Geržinič

December 2018

in partial fulfilment to obtain the degree of

Master of Science
in Transport, Infrastructure and Logistics

at the Delft University of Technology
Faculty of Civil Engineering & Geosciences (CEG)
Faculty of Technology, Policy and Management (TPM)

Evaluation Committee:	Prof.dr.ir. Caspar G. Chorus	TPM – TU Delft	Chairman
	Dr.ir. Sander van Cranenburgh	TPM – TU Delft	Supervisor
	Dr. Oded Cats	CEG – TU Delft	Supervisor

Preface

Dear reader,

For the past nine months I have been working on my thesis and most of the work I have carried out in this period is presented in this report. The topic of my thesis was gathering data using sequential best-worst discrete choice experiments and estimating that data with random regret minimisation modelling. I worked on the thesis at the Delft University of Technology, faculties of Civil Engineering and Technology, Policy and Management. For the purpose of the research, data gathering was carried out in my hometown of Ljubljana, Slovenia, on the topic of park-and-ride facility choice for daily commuters to the city.

First and foremost, I would like to express my gratitude to my thesis committee. Daily supervisors Sander van Cranenburgh and Oded Cats with whom I have met on a bi-weekly basis and who have provided guidance and detailed feedback at every meeting. And the chair of my committee, Caspar Chorus, I would like to thank for starting the idea of this thesis and for the motivation since before the official work even started until the very end. To my entire committee, thank you very much for your advice, support and enthusiasm from the start of the project over a year ago and throughout the ups and downs of my research.

I would like to thank Professor Emily Lancsar from the Australian National University for her time and her contribution to this research with her detailed commentary of the survey and report. Given her expertise in choice modelling and SBWDCE in particular, her input was invaluable.

I also want to thank all the respondents who took time out of their lives to provide invaluable observations for the stated choice survey and especially those who even went a step further and shared the survey with others as well, without which this research would not have been possible.

Many thanks also go to my family and friends who have stood by me throughout the entire graduation process, helping me with carrying out the research and by distracting me when I needed a break. I want to thank my fellow TIL students and students from the hok for keeping me motivated during my graduation. Special thanks go to Hanna and Ruben for making me laugh every single day of my graduation, Eva and Larissa for the brunches, dinners, games and conversations that have we have shared over the year, Alessio for always being there to talk about news in the public transport world and for our weekly Sunday lunches in Bagels & Beans, my cousins Nastja and Manja and my friends in my hometown, who kept me company and made me want to go back home as often as I did. Most of all, I would like to thank my mother, for always standing by me, inspiring me and giving me emotional support every step of the way.

Finally, I would like to dedicate this thesis to my father, who unfortunately could not see me finish my Master program and graduate but has always supported my interests and has motivated me to do my very best for as long as I can remember. I will be forever grateful for that.

Nejc Geržinič
28.11.2018

Executive summary

Transportation planners have an array of methods at their disposal for analysing trends, making projections on future demand for services and evaluating policy measures. In recent decades, discrete choice modelling (DCM) has become a key methodology used by transport planner around the world. Its development began around the mid-20th century with a simple logit model of binary mode choice, with alternatives described only with travel time and travel cost. Choice models have since been advanced and expanded with multiple alternatives, preference and taste heterogeneity, use of different decision rules, latent classes etc. Advancements of the first-choice random utility maximisation (RUM) multinomial logit (MNL) model have been developed as a response to the model lacking realistic behavioural representation for a certain aspect of user behaviour. Two such advancements, one in the realm of efficient data gathering and one in behaviourally realistic modelling, are used in this research.

Sequential best worst discrete choice experiments (SBWDCE) are a relatively new data gathering method for obtaining a larger number of observations from fewer respondents, saving time and money as well as allowing choice models to be estimated for smaller niche groups. The stated preference survey presents respondents with multiple alternatives and asks them to alternatingly select the best / worst alternative, until the choice set is exhausted (only a single alternative remains). The observations can be modelled in the same order they were stated by the respondents, with an SBWMNL model or using an implied rank ordered logit (ROL), which takes the implied order of the alternatives. In the SBWMNL, best choices are modelled in the normal MNL fashion, while worst choices are modelled using the opposite values of utility (disutility) and calculating the probability of the selected worst alternative.

Random regret minimisation (RRM) models are also a relatively new method in discrete choice modelling, challenging the notion of respondents assessing alternatives in a fully compensatory manner. Compensatory behaviour refers to how people trade-off the performance of alternatives.

- **Fully compensatory behaviour**: implies a bad performance on one attribute can be compensated with good performance of another attribute
- **Non-compensatory behaviour**: worse performance of an attribute reduces the attractiveness of an alternative, while a better performance on a different attribute adds no value
- **Semi-compensatory behaviour**: a middle ground, where better performance helps in an alternative's attractiveness, but to a lesser extent than worse performance of the same scale

The μRRM model is a generalised model with a parameter that varies on a scale from zero to infinity, translating to non-compensatory (P-RRM) and fully compensatory (RUM) behaviour respectively.

The SBWDCE data collection technique and μRRM modelling present the opportunity to combine them with the aim of improving model fit through higher behavioural realism and more significant parameters. According to image theory, the scale parameters of this combined SBWMNL μRRM model should result in fully compensatory behaviour for best choices and non-compensatory behaviour for worst choices. Image theory postulates that respondents first reduce choice sets by excluding the worst options in a non-compensatory fashion and then select the best in the reduced choice set by trading-off in a compensatory manner. This leads to the formulation of a research goal and hypothesis of this research.

Research goal: *Combining the SBWDCE method for gathering stated choice data and analysing that data using a μRRM model to estimate taste parameters and scale parameters.*

Hypothesis: *In the context of best worst choice experiments, people choose in line with RUM when selecting the best alternative and in line with (P-)RRM when choosing the worst alternative.*

Methodology

The first step in the research was to carry out a detailed analysis on both the SBWDCE data gathering technique and the μRRM model. As they are both specific and complex methods, understanding the theories, specifications, application examples, experimental design setups etc. was crucial for the experiment to be constructed so that the model could be estimated.

The second step was constructing a stated choice survey. From the literature overview, a detailed list of requirements for the experimental design was put together. The SBWDCE requires using a larger number of alternatives so that multiple best and worst choices can be made. Using the (μ)RRM model requires at least three alternatives, described by generic attributes with three or more attribute levels. Not meeting these constraints, the RRM model reduces to a RUM model. Given the requirements, the topic of park-and-ride facility choice in the Central Slovenia Statistical Region (CSSR) was selected for the survey and model application. The Bayesian efficient experimental design contained twelve choice sets, with five unlabelled alternatives, described by five attributes, four with three levels and a fifth dummy variable. To make the model robust towards any decision rule, a (new) method for generating experimental designs that calculates design efficiency irrespective of the underlying decision rule was utilised. The survey was implemented in the Survey Gizmo online environment and spread through social media, municipalities, interest groups and employers in Ljubljana. 108 responses were obtained.

The third step was determining the models to be estimated and compared. The models were first tested on a synthetic dataset to determine if taste and scale parameters, used to generate the dataset, can be obtained from the stated choice data. Four models (I.1-I.4) with implied decision rule combinations (RUM and P-RRM) for best and worst choices were specified (Figure 0.1). Scale parameters were also estimated in five models developed in this research (differing in the number of scale parameters and the number choice set size constants). They were compared to three reputable models: first-choice RUM, first-choice μRRM and SBWMNL RUM (Figure 0.1). To test the sample for latent groups, seven latent class models were also estimated, both with implied and estimated scale parameters. The models were first estimated without employing socio-demographic data in the class allocation function and later with it. LC models were compared amongst themselves, as well as to the previously estimated models.

The fourth step was to apply a few of the models (E.1, E.3 and E.7) on a case study of park-and-ride facility choice. For this purpose, a corridor to the southwest of Ljubljana was selected, where three P+R facilities are in operation and more planned. The models were used to evaluate how the evolution of the current plan would impact attractiveness of individual parking lots. A more sustainable development plan was proposed, utilising the link-and-ride concept of P+R facilities, and tested in the same way as the current plan, comparing how the attractiveness of individual facilities changes with different policy measures, such as improved public transport services at either satellite or remote P+R, the addition of new P+R facilities, changing prices and a new railway service.

Model estimation results

Models with an implied scale parameter showed that models assuming a RUM decision rule for best choices (I.1 and I.3) achieved a better model fit. Regarding the worst choice decision rule, the difference in model fit was minimal, with a slight advantage for the RUM decision rule (I.1) over the P-RRM (I.3), although the Ben-Akiva & Swait test proved that model I.3 has a 34% probability of being the true model in the population, while having a slightly worse fit (0,08 log-likelihood points less).

The results of models where the scale parameter was estimated are shown in Figure 0.1. The first striking difference is a much higher rho-squared of the first-choice only models E.1 and E.2, possibly indicating more reliable first choices. Comparing the SBW models, all the models accounting for decision rule variation performed better than model E.3. The best performing model is uncoincidentally the model with the most estimated parameters. Even when accounting for the extra parameters in the LRT (comparing E.5 and E.6), E.6 has only a 5% probability of having achieved a better fit due to coincidence. Models E.7 and E.8 however are very similar, with the LRT stating that E.8 achieving a better model fit is almost entirely due to coincidence (p-val = 0,96). The models with multiple choice set size constants (E.5 and E.6) performed better than those with a single constant adjusted for the size (E.7 and E.8).

Models with implied / fixed scale parameters

Model	Best choices	Worst choices	Rho-squared
I.1	RUM	RUM	0,1384
I.2	P-RRM	P-RRM	0,1310
I.3	RUM	P-RRM	0,1384
I.4	P-RRM	RUM	0,1303

Models with estimated scale parameters

Model	μ(Best (1))	μWorst (1)	μBest 2	n° Λ	Rho-squared
E.1	--- *First-choice RUM model* ---				0,2573
E.2	53,50**	--- *First-choice μRRM model* ---			0,2575
E.3	--- *SBWMNL RUM model* ---				0,1301
E.4	10,00′			3	0,1384
E.5	10,00′	0,43**		3	0,1393
E.6	10,00′	0,43**	0,50*	3	0,1397
E.7	10,00′	0,36**		1	0,1340
E.8	10,00′	0,36*	12,31	1	0,1340

* significant at a 95% level ** significant at a 99% level ′ fixed value

Latent class models

Model	Class 1 μBest	Class 1 μWorst	Class 2 μBest	Class 2 μWorst	Rho-squared
LC.0	61,45**	0,72	2,03**	2,04*	0,2090
LC.1	RUM	RUM	P-RRM	P-RRM	0,1973
LC.2	RUM	RUM	RUM	P-RRM	0,2077
LC.3	RUM	RUM	P-RRM	RUM	0,1983
LC.4	P-RRM	P-RRM	RUM	P-RRM	0,1969
LC.5	P-RRM	P-RRM	P-RRM	RUM	0,1955
LC.6	P-RRM	RUM	RUM	P-RRM	0,1980

* significant at a 95% level ** significant at a 99% level

Figure 0.1. Summary of all estimated models with model outcomes

Investigating the scale parameter values, the first-best choices are always made utilising a RUM decision rule. The first-worst choice is also largely consistent across models with 0,43 (E.5 and E.6) or 0,36 (E.7 and E.8). The biggest difference between the models utilising a single choice set size constant or multiple is the decision rule of the second-best choice (μBest 2), taking the value of 0,50 in E.6 and 12,31 in E.8. The first implies non- to semi-compensatory behaviour while the second represents fully compensatory behaviour. The difference was determined to be a consequence of the way choice set size was accounted for. It and the scale parameter are rooted in one another and thus changing one will result in the other changing as well. It was observed however, that differences in model fit were smaller when switching between decision rules for each additional choice: RUM or P-RRM made a big difference for the first choice (over 50 LL points) and less for the second (5-10 LL points) and third choice (2-10 LL points). That is also why in model E.4, where a single scale parameter value was used, it converged to a RUM decision rule across all choices made.

The LC models gave no significant information regarding different decision rules within the sample, with insignificant parameters and similar model fits across all decision rule combinations, with models implying a RUM decision rule for best choices doing slightly better. All LC models still achieved a significantly better model fit, but that was most likely due to the different values of attribute parameters, as in all cases one class had significantly higher parameter ratios for all trip segments. The two classes were also very similar in size, regardless of the decision rule combination. Adding socio-demographic data provided no added value, possibly because there is no correlation with the class membership or because the sample was too small to draw meaningful conclusions.

Model application results

The results of the case study show that satellite P+R facilities are much more attractive to consumers, having a large majority of the market share. Among all the tested measures, increasing the cost of satellite P+R and reducing the cost of remote parking proved the most successful to sway commuters to use remote P+R. A good railway service also provided benefits to remote facilities. A similar benefit can be obtained through more frequent and / or faster bus services. This, combined with the pricing policy shows that a lot can be achieved with relatively low investments, compared to a new railway.

With respect to model performance, the fact that the developed model (E.7) had a fully compensatory decision rule meant that the major difference between models would be based on their respective taste parameters. While all three tested models (E.1, E.3 and E.7) had roughly similar parameter ratios, model E.7 was most sensitive to change, predicting the largest differences in market share. E.3 on the other hand was the least sensitive. Model E.7 predicted the highest stake for satellite P+R facilities, with model E.3 being the opposite: putting more emphasis on remote P+R. Model E.1 (first-choice RUM) was a middle ground both with respect to sensitivity and emphasizing either remote or satellite P+R.

Conclusion

This research has shown that different decision rules are indeed applied in best-worst choice tasks, with the decision rules for the most part being in line with image theory: best choices are made using fully compensatory behaviour and worst choices using non-compensatory behaviour. Because of this, models accounting for such variation performed better in terms of model fit and validation. Latent class models, despite having a better model fit, seem to be driven by different taste parameters and marginal rate of substitution, rather than differences in decision rules.

While image theory may explain best / worst choice decision rules, it does not explain why the second-best choices were once estimated as compensatory and once not. This is due to successive decisions being less reliable compared to the first choice and the fact that taste parameters are liable to be less intense and as these were the same across all choices, the scale parameter had to compensate for this.

The inaction effect may have also played a role in this, stating that when choosing not to act (which can be assumed when selecting worst or even second-best), respondents are more likely to experience regret and thus decide with anticipated regret in mind.

The use of best-worst style surveys is contentious, as although it offers far more observations from a lower number of respondents, it is criticised for being an unnatural way of decision-making. Even some respondents of the survey commented on this issue. Each individual field of application may also be more or less suitable for best-worst tasks, with transport possibly falling under the less suitable. Best-worst tasks are widely used in healthcare and SBWDCE prove very useful in cases of very small populations to which certain surveys are relevant (a small group of patients suffering from a rare medical condition). An interesting potential for SBWDCEs is obtaining prior parameter values from a pilot survey. For efficient experimental designs, researchers need prior parameter values, which often cannot be obtained from literature. Surveys for obtaining priors are often much more limited compared to the main stated choice surveys. Using SBWDCE would therefore allow researchers to obtain significant prior parameter estimates from a small number of respondents.

Although providing more behaviourally realistic model outcomes, the developed models have a few downsides as well. By introducing additional parameters, models become more complex and the time needed to estimate results increases. Of the developed models, flexibility for application was also found to be an issue, as because the scale parameter and choice set size constant are linked, only one pair of the parameters can be used for each choice set size. This was slightly remedied by using a single choice set size constant, but this area of modelling requires further research. Another reoccurring issue is RRM models not being applicable for project appraisal in the same way as RUM models, due to their inability to determine the net welfare effect (the value of the whole choice set), although researchers have made progress in this field as well.

This research adds to the understanding of respondent behaviour in stated preference situations and in the field of SBWDCEs in particular, showing that respondents do indeed utilise different decision rules, as is postulated in image theory. It also opens up many new possibilities to be investigated. Given the specific nature of SBWDCEs, efficient experimental designs specifically for best-worst tasks should be analysed and a new / different D-error that incorporates all possible choices (not only the first-best) could be determined. Examining whether random errors are choice set specific, alternative specific or a combination of both would greatly improve the understanding of consumer behaviour and researches using synthetic data to evaluate models would benefit from it as well. The identified relationship between the scale parameter and choice set size constant also warrants further attention in the scientific community. As with many newly developed models, the SBWMNL μRRM model would benefit from further applications in a variety of fields. Making conclusions based on a single sample is not accurate and by applying it in many different fields could help in either proving or rejecting the findings of this research.

Table of content

Table of figures

Table of tables

Table of equations

Glossary

BIC	....	Bayesian information criterion
BWDCE	....	Best worst discrete choice experiment
BWS	....	Best worst scaling
CSSR	....	Central Slovenia Statistical Region
DCE	....	Discrete choice experiment
DCM	....	Discrete choice model
H&S	....	Hub-and-spoke
L+R	....	Link-and-ride
LC	....	Latent class
LL	....	Log-likelihood
LOS	....	Level-of-service
ML	....	Mixed Logit
MNL	....	Multinomial Logit
P+R	....	Park-and-ride
PT	....	Public transport
RDA LUR	....	Regional Development Agency of the Ljubljana Urban Region
ROL	....	Rank ordered logit
RP	....	Revealed preference
RRM	....	Random regret minimisation
P-RRM	....	Pure random regret minimisation
µRRM	....	Mu random regret minimisation
RUM	....	Random utility maximisation
SBWDCE	....	Sequential best worst discrete choice experiment
SBWMNL	....	Sequential best worst multinomial logit
SE	....	Standard error
SP	....	Stated preference
VMT	....	Vehicle miles travelled
VoT	....	Value of time
WtP	....	Willingness to pay


2 ZELENA JAMA

1 Introduction

Choice modelling has been a key method for researchers and developers in the field of transport planning for the past few decades and is an essential tool in every transport planner's toolbox (van Cranenburgh, 2017). Its development began in the first half of the 20th century, with first applications in the 1960's (Ben-Akiva & Lerman, 1994). Until then, aggregate demand forecasting models, such as the gravity and maximum entropy models dominated transport planning research. Despite both being powerful and commonly used methods, they cannot predict individual travel behaviour, especially with respect to new services and/or infrastructure (McFadden, 2000). A move towards disaggregate models (such as choice modelling) was also seen in the slogan of the Travel Demand Forecasting Project (1972):

"Zones don't travel; people travel!"

First applications of discrete choice models (DCMs) were in a simplistic binary form, asking respondents to choose among two alternatives, for which only the time and cost of travel were given. This allowed researchers to determine the value of time (VoT) for travellers and use it to assess the value of investments in transportation services/ infrastructure (Ben-Akiva & Lerman, 1994). In the 70's, DCMs became more complex: multiple alternatives were added, more attributes were used, socio-demographics were included, and the area of application spread from mode choice to include destination choice, housing choice, trip frequency etc. (Ben-Akiva & Lerman, 1994). That period also saw a major application for the Bay Area Rapid Transit (BART) project, the metro of San Francisco. A survey was carried out three years before operations began, and the forecast was compared to actual ridership in 1975, with high levels of accuracy. By proving its success, DCEs spread to other sectors: energy, telecommunications, healthcare, environment,… (McFadden, 2000).

Today, choice models are a widely used method for demand projections in a plethora of fields, from healthcare, politics, marketing, environment, tourism etc. (Chorus, 2016). In the sphere of transport, where the method originated, DCMs are a key method used in many transport models for project appraisal, evaluation of alternatives, planning of future services, necessary infrastructure investments etc. It has been applied to virtually any aspect of transport that involves consumer or even decision-maker choice making (Chorus, 2016).

The widespread applicability of DCMs pushed researchers to develop new, more accurate and more realistic ways to model people's behaviour in different circumstances. Although still most commonly used, the traditional Random Utility Maximisation Multinomial Logit (RUM MNL) model is not without its flaws. While the rather straightforward application led to its widespread success, more advanced models have focused on improving the model accuracy and reliability, at the cost of simplicity and tractability. Each improvement focused on one or more aspects of the RUM MNL model that was deemed overly simplified or not truly representative of people's behaviour.

In this research, we consider two relatively new additions to the choice modelling field of science. The first is a more efficient discrete choice data gathering technique, an improvement of traditional first-choice discrete choice experiments (DCEs): Sequential Best Worst Discrete Choice Experiments (SBWDCE) (Louviere et al., 2008) (Lancsar, Louviere, Donaldson, Currie, & Burgess, 2013). The second is a different approach of modelling the data by assuming a different underlying decision rule used by the respondents to make choices: instead of the traditional RUM model, the Random Regret Minimisation ((µ)RRM) model is used (Chorus, Arentze, & Timmermans, 2008) (Chorus, 2010) (van Cranenburgh, Guevara, & Chorus, 2015).

These advancements give the opportunity of carrying out a stated choice analysis that combines both the higher data gathering efficiency of SBWDCEs and higher behavioural realism of RRM models. There is reason to believe that combining the two can lead to more realistic modelling outcomes, as respondents may utilise different decision rules when selecting best and worst alternatives in an SBWDCE setting. If the hypothesis of this research is proven to be true, it could lead to better understanding of respondent's decision-making and behaviour in best-worst choice tasks. SBWDCEs are a powerful tool for choice modellers to obtain a larger number of observations from fewer respondents, but the modelled behaviour should be accurately accounted for.

This chapter starts by examining literature to pinpoint the problem to be tackled in this research. Following are the statement of the research objective and formulation of the research questions. Subchapter 1.3 justifies both the scientific and societal relevance of this research. The chapter is concluded with the research approach and the full outline of the report.

1.1 Problem statement

The most widely used approach in discrete choice modelling in practice is creating a first-choice DCE and estimating the results with a RUM MNL model. This is very tractable and straightforward, making it easy to understand and apply to a variety of research topics. Nevertheless, more advanced approaches are constantly being developed and applied in various fields of research as a response to limitations of the traditional line of research. Two such methods are SBWDCEs, which offer higher data gathering efficiency and RRM models, which model respondents' behaviour differently. We investigate here if there is a benefit in modelling realism and gaining knowledge on respondent's behaviour if they are combined. Choice modelling analyses can be broken down into a multi-step approach. Different researches (Hanley, Mourato, & Wright, 2001) (Reed Johnson et al., 2013) (Bennett, 2005) have proposed a variety of approaches with a different number of steps, but all include the same basic process, shown in Figure 1.1. The highlighted parts of the process form the focus area of this research.

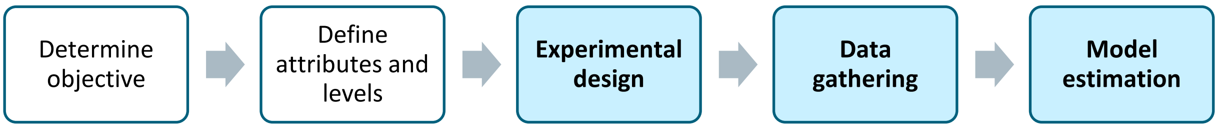


Figure 1.1. Choice modelling approach

SBWDCE rely on respondents alternatingly selecting best and worst alternatives within the same choice set, allowing researchers to obtain a larger number of observations from a small number of individuals. It is an approach that influences both the Experimental design stage and the data gathering stage in Figure 1.1. The (μ)RRM model is a way of modelling obtained data and was developed to better capture behavioural realism of the utilised decision rule of the respondents. It is a model estimation method that has been gaining attention in research since its development less than a decade ago.

There is reason to believe that there may be a difference in the way respondents make those decisions in a best-worst setting. Ben-Akiva, Morikawa, & Shiroishi (1992) investigated the reliability of stated preference ranking data in a set of four alternatives. By analysing the taste parameters from each level of the rank separately, they concluded that the intensity of parameters decreases with successive ranking levels. This means that the choice for best alternative yielded stronger parameters than second-best, third best etc. This can be accounted for with scale parameters, but the authors still warn when going further than two rank levels. Investigation of the intensity of taste parameters in a best-worst experiment was undertaken by Dyachenko, Reczek, & Allenby (2014) and the results were in line with the findings of Ben-Akiva et al. (1992), stating that first choices have stronger taste parameters than second choices, regardless of them being for the (second) best or worst alternative.

Differences between best and worst choices may also be a consequence of different decision-making strategies for best and worst choices. One theory that investigates consumer behaviour and may provide an explanation for these differences is Image theory (Beach & Mitchell, 1987). According to the theory, consumers undertake a two-step process in decision-making when faced with multiple alternatives. First, they evaluate the alternatives based on their compatibility and exclude those that do not meet a subjective minimal standard. Once the choice set has been reduced to only include acceptable alternatives, these are evaluated on their profitability. According to Beach & Mitchell (1987), compatibility is a non-compensatory decision-making process, excluding alternative that perform below a subjective threshold, while profitability is compensatory as consumers wish to maximise their 'profit' (utility) by trading-off the different characteristics (attributes) of alternatives. The definition of compatibility of Beach & Mitchell (1987) is similar to the decision-making process of Elimination by aspects (Tversky, 1972), where the decision-maker selects a certain attribute and removes all the alternatives that do not meet a certain aspect (a minimum performance on that attribute). This process continues until a single alternative remains. Compatibility and profitability also draw strong parallels with RUM and P-RRM decision rules with respect to the level of compensatory behaviour. Given image theory, it is reasonable to assume that when selecting the worst alternative, only the first stage of the two-step decision-making process of image theory is undertaken by consumers.

Image theory was reinforced and its understanding expanded by Ordóñez, Benson III, & Beach (1999) who went a step further in explaining the first step of the process by concluding that consumers have a certain subjective threshold and for each alternative, they add up its violations with respect to an expected performance. If the sum of the violations exceeds the threshold, the alternative is excluded from the second evaluation stage. As is shown in chapter 2, this is very reminiscent of the P-RRM model formulation, where only worse performance is considered and added up. Meloy & Russo (2004) further added to the understanding of image theory by concluding that people prefer to choose by selection rather than rejection, reinforcing the second stage of image theory. They also found that the first stage of the two-step approach does not seem to be the same in all cases. Exclusion of under-performing alternatives was found to be dominant in difficult situations, where a single correct answer is expected. Conversely, in a setting where more subjective judgement needs to be passed, inclusion rather than exclusion seems to be the dominant thought process of narrowing down a choice set.

These findings motivate the statement that respondents utilise different thought processes in selecting a best or a worst alternative. As exclusion of worst alternatives seems to be accomplished with non-compensatory decision rules, (P-)RRM can be expected to represent worst choices, while best choices can be assumed to be selected with the help of RUM, as the consumer wishes to maximise their utility. If this hypothesis is true, combining SBWDCEs and RRM modelling provides greater realism in modelling best-worst choice behaviour. How both techniques have developed is presented in the following subchapters, with SBWDCEs covered in subchapter 1.1.1 and the RRM model in subchapter 1.1.2.

1.1.1 Efficiency in obtaining observations for DCEs

In the stage of gathering discrete choice observations, there is constant strive to gather more observations and / or capture the trade-offs respondents make more efficiently. For the latter, efficient designs have made substantial advancements in recent years. For the former Lancsar et al. (2013) have pointed out two main ways of attaining additional observations:

1. Asking more people to take part in the survey
2. Asking the participants to answer more questions (more choice situations)

While the first approach is very straightforward, the main issue of it is that it will cost more money and take more time to obtain those additional responses. This may be problematic, as researches are often monetarily and time-wise restricted by a tight budget. The second option may at first seem rather easy to implement and without any negative impacts, yet several researches have suggested otherwise.

A popular belief is that respondents can handle up to seven attributes in a choice set with two alternatives (Molin, 2016a). This statement is debatable and has been investigated quite substantially, partially also because the number of choice sets / alternatives / attributes a respondent can handle without being overburdened depends on the context of the choice task and how familiar the respondent is with it. Amaya-Amaya, Kalleh, Ryan, & Odejar (2004) concluded that there seems to be a U-shaped relation between the survey complexity and the accuracy of the results. Initially, choice consistency of individuals is increasing with more attributes being added, but once it becomes overwhelming for the respondent the consistency decreases. A similar study was also performed by T. Arentze, Borgers, Timmermans, & DelMistro (2003), who varied the number of attributes (three or five) and alternatives (two or three) per choice set. What they found is that the change from two to three alternatives did not influence the results, while an increase in the attribute number did. An older study by Malhotra (1982) compared far larger numbers of attributes and determined the overload for respondents starts at 15 attributes, which seems very high compared to other findings. For alternatives, they summarised their own findings and those of other papers of the time to determine the maximum number per choice set to be between four and six. Interestingly, they found no significant interaction between the number of alternatives and the number of attributes presented, which seems counterintuitive.

Studies on respondent burden have also included the survey length (number of choice sets). Hess, Hensher, & Daly (2012) analysed the accuracy of respondents and how it changes over the length of the survey. Curiously, they found no evidence pointing towards participant fatigue and have criticised older studies for doing so. Their explanation was in a learning effect, that people get used to trade-offs from the first few choice tasks. This is supported by more cost evasive behaviour at first: the lowest willingness to pay (WtP) is present in the first choice task and then progressively increases over the following tasks. The perception of survey complexity was also found to be largely subjective and easily influenced by the researcher through guidance and information provision (Yu, Fricker, & Kopp, 2015).

A review of actual practice revealed that most researches solved this by limiting the survey in a way that it only focused on one major issue being tackled in the research, implying the necessity of making trade-offs (Timmermans & Molin, 2009). An optimal number of choice sets and alternatives was suggested by Chung, Boyer, & Han (2011) to be six and five respectively, although they acknowledge that this cannot be generalised to all SP surveys, as a single dataset from the food industry in South Korea was used.

The extensive literature on respondent burden also points to why efficient designs have been and still are a topic of experimental design research. As stated previously, researchers wish to gather sufficient observations to estimate models. However, these observations need to be of high quality as well, for the models to be estimated. The most important aspect of observation quality are the trade-offs that people are asked to make in presented choice sets. Trading-off the attributes of different alternatives, gives insight into respondents' preferences and can infer the value of estimated parameters, used to determine the value-of-time, value of transfers, waiting, modes etc. (Walker, Wang, Thorhauge, & Ben-Akiva, 2018).

Orthogonal designs have been the norm for stated choice tasks for a long time. Their main advantage is that all attribute levels appear an equal number of times in such a way that correlations between all attributes equal zero. A consequence of this is that some choice situations will be uninformative, as there will be no trade-off present, due to dominant alternatives, especially in unlabelled choice situations. A dominant alternative is one that is superior to another in at least one attribute and is not inferior in any attribute. For example, if route A takes 10 min and costs 2€, while route B takes 15 min and costs 3€, a rational decision maker should choose route A, which is both cheaper and faster. While being cognitively easier for the respondent, it provides no or only limited information to the researchers on the trade-offs the respondent had to make. In most choice tasks however, dominance is not as straightforward, for example with labelled alternatives, qualitative attributes, in cases where omitted attributes influence the decision process etc. Although consensus has not yet been reached for whether dominance is beneficial or not for stated choice surveys, a method that avoids dominant alternatives are efficient designs, which require prior parameter values to exclude choice sets with dominant alternatives (Molin, 2016b). In a way, efficient designs also reduce respondent burden, by making a smaller number of choices, yet providing researchers with almost the same information: by omitting choice sets that provide limited information and allowing respondents to concentrate their cognitive abilities on what is more important.

We can see that although researchers generally prefer not to overburden their respondents, no agreement has so far been reached on whether-or-not respondent fatigue is caused by complex surveys and to what extent, partly also because the context differs with each choice task. Knowing this, it seems logical to avoid the issue if possible. An interesting approach to avoid it has been proposed by Lancsar et al. (2013) in the field of healthcare. Healthcare in particular wishes to reduce the burden on respondents because (1) the questions posed to them can often be difficult as they concern personal health and life-or-death situations and (2) the number of respondents available to participate in such surveys can be quite limited, so researchers need to make the most of what they have. They propose using a larger number of alternatives and getting the respondents to keep making decisions, choosing the best or worst alternative (alternatingly) among the remaining alternatives, until the choice set is exhausted (only one alternative remains). The argument for this is that people are already familiar with the choice set and it is easier to make further choices while at the same time providing the researcher with more observations per choice set.

1.1.2 Respondent decision rule in a stated choice setting

In his lecture, Chorus (2016) summarised two major issues of the RUM MNL model and presented the advanced choice models that tackle these limitations. The first limitation is an unrealistic distribution of the error term in MNL models. Among other things, this causes nesting issues of different alternatives and does not account for the panel effect. The latter means that the MNL does not consider the fact that a single respondent makes choices in multiple choice sets, but rather sees each choice as a new respondent. A proposed solution is the Mixed Logit (ML) model.

Secondly, Chorus (2016) questions the RUM model for its linear-additive nature, as it does not account for non- or semi-compensatory behaviour nor does it consider the effect of the choice set (other present alternatives). A solution proposed by Chorus et al. (2008) and later refined by Chorus (2010) is a change in decision rule: the Random Regret Minimisation (RRM) model, based on regret theory. The idea behind it is that travellers, when faced with decisions, rather than choosing the alternative with the highest utility, select the one with the lowest regret. Although the intuition behind both models is similar, the difference lies in the mathematical formulation of the performance of each alternative.

Since the introduction of regret theory into transport decision making, many new researches have been published and the effectiveness of RRM compared to RUM is discussed by Chorus, van Cranenburgh, & Dekker (2014). Their paper analyses empirical findings of 43 journal articles, in which the RUM and RRM models were compared. They conclude that although differences in model fit between both models were found, they tended to be rather small. It also occurred that a better fitting model performed worse on the hold-out dataset. While the model fit differences were rather small, the predictions (i.e. of market share) made by these models can be significantly different. This leads to substantially different implications for the work of decision makers. Chorus et al. (2014) state that a single best model is very difficult to choose and that a good option for producing robust predictions is using both models, if the difference between model fits is not significantly different. This coincides nicely with the statement made by Chorus et al. (2008), saying that the proposed RRM model is not intended to be superior to RUM-based models in any way, but is rather proposed as a different way of looking at choice modelling and people's decision making process.

1.2 Research objective and research questions

SBWDCEs and RRM modelling have each made significant contributions to their respective areas of choice modelling and as stated in the previous subchapter, there may be benefit in combining them. The benefit may come from a hypothesis that people choose best and worst alternatives differently, possibly utilising different decision rules. That is why the main objective in this research is combining the two methods to evaluate the hypothesis. If the decision rule indeed differs, the model proposed in this study should achieve a better model fit compared to models not allowing for this variation (intra-respondent decision rule variation). To our knowledge, this has not yet been attempted, possibly due to SBWDCEs and RRM models both being a relatively new setting in the field of choice modelling.

Additionally, the proposed model(s) and already established models will be applied on a case study of park-and-ride (P+R) facility choice to see if they produce different demand projections and if so, to what extent do they differ. The case has been selected to allow for the collection of stated preference data on the subject, to showcase the functionality of the proposed model on a real-world case and to contribute to the wide array of choice modelling applications in the field of transportation that have already been carried out. The models already proven in research that will be used as a benchmark in this research are the first-choice RUM model, first-choice μRRM model (van Cranenburgh, Guevara, et al., 2015) and the SBWMNL model (Lancsar et al., 2013). These models are specified in detail in chapter 2.

From the problem statement and the presented model integration/combination opportunity the two methodologies provide, the goal of this research is formulated below. The SBWDCE and μRRM methods are described further in chapter 2.

Research goal: *Combining the SBWDCE method for gathering stated choice data and analysing that data using a μRRM model to estimate taste parameters and scale parameters.*

On the way to achieving the presented research goal, new insights into respondents' decision-making strategies can be obtained. To better structure these insights, a hypothesis (underpinned in subchapter 1.1) on the behaviour of respondents was established. The different decision rules, such as RUM and P-RRM that are used in the hypothesis are introduced in subchapter 2.3.1. From the formulated hypothesis, four research questions are proposed and summarised in the following paragraphs.

Hypothesis: *In the context of best worst choice experiments, people choose in line with RUM when selecting the best alternative and in line with (P-)RRM when choosing the worst alternative.*

The first question to be dealt with in this research is whether-or-not such a complex model can even be estimated. With a large number of parameters accounting for variations in taste, scale and choice set size, this potential issue should be explored.

RQ1: *Can a μRRM model extract accurate taste and scale parameters from data that was obtained in a sequential best-worst discrete choice experiment setup?*

The second research question (RQ2) is concerned with the extent of the difference between the proposed model and already developed models with respect to model fit and parameter estimates, and what these differences mean for the modelled behaviour of individuals.

RQ2: *To what extent does accounting for intra-person decision rule variability between making best and worst choices improve the model fit (if at all) when compared to already established discrete choice models?*

Similar to the possibility of the presence of intra-person decision rule variability (the same respondent using different decision rules in different choice settings), inter-person variability in decision rule utilisation may also be present among respondents (different respondents using different decision rules in the same choice situation). To investigate this issue, research question 3 (RQ3) is formulated.

RQ3: *How significant is the presence of inter-personal variability of decision rule utilisation in the setting of best-worst discrete choice experiments?*

Using the developed models on a case study of park-and-ride facility choice, research question 4 (RQ4) investigates the predictive abilities of the proposed model and compares the outcomes to other established models.

RQ4: *In what way does accounting for different decision rules in a best-worst experimental setting differ in predicting (future) market share when compared to already established discrete choice models?*

1.3 Scientific and societal relevance

Modelling and demand forecasting in the field of transportation are of major importance for transport planners, as they allow them to assess infrastructure investments, transportation services, innovative solutions etc. and choose one that will meet projected future demand. Choice modelling is often used in these studies and stated preference surveys in particular are a good tool for modelling new alternatives / solutions / modes that are not (yet) present. Their high behavioural realism and relative ease of use has also contributed to their widespread application in transportation. Nevertheless, this has not stopped researchers developing more realistic models, that capture people's behaviour even better. Several improvements of the traditional RUM MNL choice model have seen success and applications in transportation. Among them is also the (μ)RRM, while the SBWDCE has seen only a single application to our knowledge, in the field of flight choice (Louviere et al., 2008).

Scientifically, this thesis is therefore relevant as a push to even greater behavioural accuracy of choice models that focus on capturing stated choice behaviour more efficiently (SBWDCE) or for increasing efficiency of choice models that capture more realistic behaviour in terms of decision rule, as well as providing an application of a newly developed model to the field of transportation. As mentioned, there is constant stride in transportation research to develop ever better and more realistic behavioural models of individual choice behaviour and this thesis will do so by combining the RRM model with the SBWMNL model. SBWDCEs are a relatively new entrant in the field of DCEs and provide essential benefits for estimating choice models from smaller samples of respondents. By adding a (μ)RRM model to SBWMNL, we can observe if people utilise different decision rules when choosing best and worst alternatives. An application on a real-world case of park-and-ride facility choice will serve as a good proof of concept for the newly developed model. Although the SBWDCE and μRRM have both already been applied in transportation (the latter substantially more) it is always beneficial for a model to see further applications as well, as it proves its reliability and versatility. This research also adds to the scientific community by introducing the SBWMNL modelling technique of SBW data to the field of transport. Additionally, the developed SBWMNL μRRM model is also a contribution of this research and it can prove useful in other fields of choice modelling as well.

Since choice models are commonly used in practice as well as in research, their impact on society is clearly relevant. By showcasing the development and successful application of a newly developed or improved choice model, applications in practice are more likely. By both proposing an improved choice model and implementing it alongside other models, the relevance to society is clear and significant.

1.4 Research approach

Before the models could be estimated and the required data gathering could begin, a literature review needed to be carried out to gain more insight into the state-of-the-art and the state-of-practice in the field of choice experiments and choice modelling. An extensive literature overview looked at discrete choice analysis in general and focused on the two main areas of interest for this research: sequential best worst discrete choice experiments and random regret minimisation modelling.

To answer the presented research questions, a discrete choice experiment was carried out. Although revealed preference (RP) may yield more realistic results for the sake of the case study, a stated preference (SP) survey was carried out, as the research investigated people's choice behaviour in best-worst stated choice experiments. The attributes and alternatives were chosen based on literature study and a pilot survey, which determined the attributes and alternatives respondents are likely to consider. Given the specific nature of the estimated choice models, the survey had to be constructed to allow the models to be estimated. Once the attributes and alternatives were determined, choice sets were generated using Ngene software.

The survey took place in the Survey Gizmo online environment and was spread to residents of Ljubljana and the wider region through social media, by the help of the PR department of the Municipality of Ljubljana and other municipalities in the region, as well as employers located in the city.

The results of the survey were used to build a DCM in Matlab software. Various DCMs were constructed and evaluated on their model fit. The literature overview determined more specifically which DCMs were compared on their model fit and demand projections. Four models with different implied decision rule combinations for best and worst choices were estimated and compared with the help of statistical tests. Decision rules were also freely estimated using a variety of models with different numbers of scale parameters and choice set size constants, due to different modelling approaches, as explained further in chapter 3.1.2. The outcomes were used to accept or reject the hypothesis proposed in subchapter 1.2 and to answer the stated research questions in the same subchapter. Latent class (LC) choice models evaluated whether there are groups of respondents that utilise different decision rule combinations in the same situation and how many respondents can be classified into each group.

Finally, the developed models and the model outcomes were applied on a P+R case study in the Central Slovenia Statistical Region. The case study evaluated the effect of potential improvement of public transport services. This was used to make policy recommendations for the municipality, region and country regarding personal and public transportation and for sustainable mobility policies in the future.

1.5 Report outline

Based on the research approach described in subchapter 1.4, the proposed report outline is presented in Figure 1.2. This takes the approach into account and makes sure that the stated research questions will be answered clearly. Each small rectangular box represents a step taken in this research, with the corresponding chapter number. After the experimental design is generated, it will be tested in a controlled environment through Monte Carlo simulation, to make sure the model performs as planned. Should there be any difficulties, they will be accounted for by either adjusting the survey as necessary (the rectangle with the dashed border and dashed lines ('*Survey adjustment*')) or taking this information into consideration when applying the developed model on the collected data. The oval called '*Survey execution*' refers to the action of gathering data for the choice experiment. After obtaining the data, choice models will be estimated, applied on a case study and finally, the whole research is concluded and reflected upon in the final chapter.

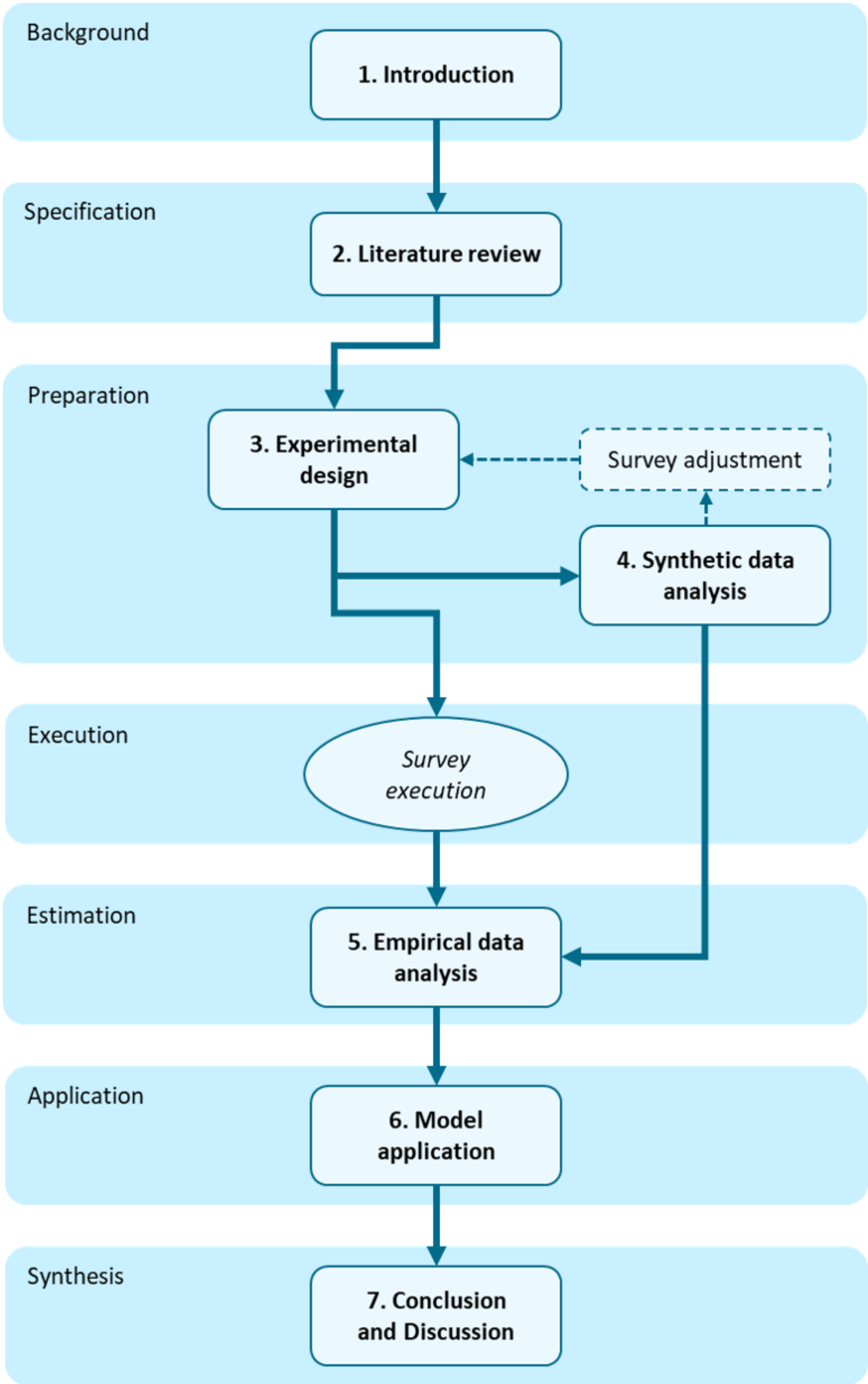


Figure 1.2. Outlined structure of the thesis report


13
SOSTRO
obvoz
MAN

2 Literature review

Since the adoption of discrete choice models for disaggregate travel demand prediction, countless new and improved versions have been proposed, with varying degrees of success. One common way of gathering data for discrete choice models is through stated preference surveys, where respondents are asked to select an alternative they prefer most, i.e. the alternative they perceive as the best (first-best data). The observations of all the respondents are then processed, and parameters are estimated by using a multinomial logit model with a random utility maximisation decision rule (Chorus, 2016).

In this chapter, an alternative way of conducting discrete choice experiments and an alternative way of analysing the data are presented: SBWDCEs and RRM modelling. SBWDCEs are used to gather a larger number of observations from a smaller pool of respondents and to answer the research questions, the obtained data will be estimated with the help of an RRM model. Because these two research methods will be combined within this research, good understanding of their properties is needed. The chapter will start with a general description of discrete choice analysis and the most commonly applied data gathering and model estimation techniques. The second subchapter will give an overview of SBWDCEs, with RRM models being covered in the third subchapter. Due to advancements made in the field of stated choice survey design and the fact that RRM modelling is used in this research, subchapter 2.4 will look into experimental designs that are robust towards the decision rule of the respondents.

2.1 Discrete choice analysis

Discrete choice analysis is a behaviour modelling method, where data on people's choices from corresponding choice sets are used to parameterise the utility functions of alternatives. This subchapter starts by giving an overview of the basic elements of discrete choice analysis, continues with the two main data sources used for estimating choice models and concludes with the theory how choice models are constructed and estimated.

2.1.1 Terminology

For better understanding of discrete choice experiments and models, five key elements of discrete choice analysis are presented and defined here: the decision maker, choice situation, alternative, attribute and decision rule (Ben-Akiva & Lerman, 1994).

Decision maker

Decision makers are usually people but can also be groups of people (households), or organisations (companies). As the name implies, their main task in the scope of discrete choice analysis is to make decisions (Ben-Akiva & Lerman, 1994). In the case of this research, decision makers are the individuals taking part in the stated preference survey on park-and-ride facility choice.

Choice situation / Choice set

A choice set constitutes a group of collectively exhaustive and mutually exclusive alternatives, among which the decision-maker must (usually) choose the best one. A choice set is a part of a larger universal set containing all existing alternatives. However, the decision maker never considers all of them, but only the alternatives that are feasible and known to the decision maker (Ben-Akiva & Lerman, 1994).

Alternative

Multiple alternatives make up a choice set, from which the decision maker chooses. They are described by attributes that are either observable or unobservable. P+R facilities are naturally discontinuous alternatives as one needs to be selected fully (discrete choice). In other words, the facilities cannot be only partially selected so they are not continuous (Ben-Akiva & Lerman, 1994).

Attribute

Attributes represent the attractiveness of an alternative, which allows the decision maker to evaluate different alternatives and select the most attractive one. The best alternative is one with the greatest utility or lowest regret (Ben-Akiva & Lerman, 1994).

Decision rule

The decision rule is a process by which the decision maker evaluates alternatives (based on their attributes) and selects one that (s)he deems best. Several decision rules have been proposed in literature, which were organised by Ben-Akiva & Lerman (1994) into four groups:

The first group of decision rules is dominance. An alternative is considered dominant if it performs better than all other alternatives in at least one attribute and does not perform worse in any of the others. However, this has three main limitations. Firstly, (1) it assumes that the presented attributes are the only ones the decision maker considers. That is a gross simplification of decision making. Secondly, (2) there almost never exists an alternative that is, for example the cheapest, fastest, most comfortable etc. The decision maker will almost always have to make trade-offs. Thirdly, (3) it is very likely that decision makers have some thresholds of differences in attribute levels: a difference in cost of i.e. 0,50€ might not be large enough to consider the alternatives to be sufficiently different (Ben-Akiva & Lerman, 1994).

Satisfaction decision rules postulate that decision makers have a certain upper (lower) limit for attributes and they will only consider the alternatives that meet these criteria. *Satisfaction* decision rules are often combined with other decision rules to come to a final choice (Ben-Akiva & Lerman, 1994).

A lexicographic decision rule is based on the theory that individuals rank attributes based on their importance. An alternative is then chosen based on its performance on the most important attribute. If two or more alternatives perform equally well, the second most important attribute is considered and so on (Ben-Akiva & Lerman, 1994).

Possibly the most extensively used decision rule in human behaviour models is utility. Unlike the other decision rules, utility considers compensatory behaviour, meaning that i.e. a more expensive alternative can still be chosen, if the travel time is sufficiently lower in comparison to a second alternative. Utility formulation means that all alternatives are reduced to a single value, based on their performance on all attributes and the decision maker chooses the one with the highest utility: (s)he wishes to maximise her/his utility (Ben-Akiva & Lerman, 1994).

In this research, a newer addition to the decision rule group is also considered: regret. When considering the four groups of decision rules by (Ben-Akiva & Lerman, 1994), the characteristics of the regret decision rule seem to place it into the *utility group* (which could and should then be renamed). The reasoning behind this is that they both reduce alternatives to a single number, corresponding to either the utility or regret of an alternative. In case of regret, the decision maker wishes to choose an alternative with the lowest regret. The main difference between utility and regret is how the value (utility, regret) characterising an alternative is computed. Because of the way regret is computed, it is semi-compensatory, or non-compensatory in the case of pure-RRM (P-RRM). With respect to that, the regret decision rule is more similar to the first three groups of decision rules, as they are all non-compensatory (Ben-Akiva & Lerman, 1994).

2.1.2 Data collection paradigms

Discrete choice models, as the name suggests, utilise data in the form of discrete choices of individuals for certain products or services. There are two main types of choice data used for estimating discrete choice models: (1) revealed preference (RP) and (2) stated preference (SP), both having their fair share of successful applications. The vast majority of discrete choice models utilise first-best data, meaning researchers only have the information on which is the best alternative in a specific choice set. Both RP and SP data are well suited for this, each with their own benefits and drawbacks (Molin, 2016a).

In this research however, the focus is on more than just first-best choice data. As explained in more detail further in this chapter, SBWDCEs use an implied order of all alternatives within a choice set. This order is obtained by asking respondents to make alternatingly best and worst choices within the same choice set until all alternatives have been ranked. As the study focuses on SP data, the choice of using RP or SP is irrelevant in this case, regardless of the benefits and drawbacks of either paradigm.

2.1.3 Discrete choice modelling

Once the choice data has been obtained, the model can be formulated and estimated. To do so, the first step is specifying the parameterised utility functions of alternatives, also called the systematic utility. This consists of the "*observable independent variables and unknown parameters*" (Ben-Akiva & Lerman, 1994, p. 2), the values of which are estimated in said model. The calculation of the systematic utility is presented in Equation 2.1. Because specifying a model that will result in correct predictions in every case is impossible, an error component is added to the parameterised utility function. The total utility of an alternative is obtained by adding up the systematic utility and the error term (shown in Equation 2.2).

Equation 2.1. Systematic utility formulation (Ben-Akiva & Lerman, 1994)

$$V_i = \sum_{k \in K} (\beta_k \cdot x_{ik})$$

where: V_i *systematic utility of alternative i*
β_k *taste parameter of attribute k*
x_{ik} *attribute level of attribute k in alternative i*

Equation 2.2. Total utility formulation (Ben-Akiva & Lerman, 1994)

$$U_i = V_i + \varepsilon_i$$

where: U_i *total utility of alternative i*
ε_i *error term of alternative i*

Following the utility formulation, the next step is determining the choice probabilities of individual alternatives. If the error term in Equation 2.2 is (1) independently & (2) identically distributed (IID property) and (3) Gumbel-distributed with mode η and scale μ, then the choice probability can be formulated as shown in Equation 2.3 (Ben-Akiva & Lerman, 1994).

Equation 2.3. Initial multinomial choice probability (Ben-Akiva & Lerman, 1994)

$$P(i) = \frac{e^{\mu \cdot V_i}}{\sum_{j \in J} e^{\mu \cdot V_j}}$$

where: *P(i)* *choice probability of alternative i*
μ *scale parameter of the Gumbel-distributed error term*

The total utility (*U*) of an alternative is based on both the systematic utility (*V*) and the error term (ε), as shown in Equation 2.2. The difficulty with this is that both the systematic utility and the error term have parameters that can be estimated to obtain the final choice probabilities: β in the systematic utility and μ in the error term. The latter, as stated previously, is the scale of the Gumbel distribution, which has a variance of $\pi^2/6\mu^2$. This implies that having a small taste parameter yields the same results as having a large variance and vice-versa. Consequently, both parameters cannot be estimated simultaneously, so the μ is normalised to 1, meaning that the variance of the error term is fixed to $\pi^2/6$ (Chorus, 2016). The choice probability formulation from Equation 2.3 is therefore reduced to Equation 2.4.

Equation 2.4. Final multinomial choice probability (Train, 2009)

$$P(i) = \frac{e^{V_i}}{\sum_{j \in J} e^{V_j}}$$

Having formulated the alternatives' choice probabilities, taste parameters can now be estimated. By knowing the choice sets in which respondents were asked to make choices and knowing those choices, we can determine the likelihood of an alternative being chosen. For a single choice set, the likelihood is equal to the choice probability of the chosen alternative (as observed in a survey). By knowing the choices people have made in a survey, taste parameters should be determined so that they represent the observed choices as closely as possible, meaning the likelihood should be as high as possible. Since surveys are usually carried out with more than one choice set and multiple respondents, the total likelihood of the dataset is calculated as shown in Equation 2.5.

Equation 2.5. Calculation of the likelihood of observed choices (Train, 2009)

$$L(\beta) = \prod_{n \in N} \prod_{i \in J} P_n(i|\beta)^{y_n(i)}$$

where:

$L(\beta)$	*Likelihood of vector β*
β	*vector of taste parameters*
$P_n(i\|\beta)$	*choice probability of alternative i in observation n*

$$y_n(i) = \begin{cases} 1 & if\ alterative\ i\ was\ chosen\ in\ observation\ n \\ 0 & otherwise \end{cases}$$

For many observations, the likelihood gets very small, so for numerical reasons, the logarithm of the likelihood is considered: the log-likelihood (LL), as shown in Equation 2.6. Through an iterative process, the taste parameters that make the data most likely are estimated by maximising the log-likelihood.

Equation 2.6. Calculation of the log-likelihood of observed choices (Train, 2009)

$$LL(\beta) = ln\left(\prod_{n \in N} \prod_{i \in J} P_n(i|\beta)^{y_n(i)}\right) = \sum_{n \in N} \sum_{i \in J} \left(y_n(i) \cdot ln\left(P_n(i|\beta)\right)\right)$$

2.2 Sequential Best Worst Discrete Choice Experiments (SBWDCE)

The sequential best worst discrete choice experiment is the third and newest addition to the Best Worst Scaling (BWS) family (Lancsar et al., 2013). The first two are (1) best worst object scaling and (2) best worst attribute scaling. In best worst object scaling, respondents are asked to choose a best / worst object (alternative) among a selection, but the objects are not described by any characteristics, such as attributes. This case looks mostly into people's beliefs, principles and opinions. Best worst attribute scaling on the other hand provides individuals with a single alternative and a set of attributes with corresponding attribute levels describing it. Respondents are asked to choose the best / worst attribute.

BWDCEs were introduced by Louviere et al. (2008) and are a combination of both object and attribute scaling BWSs: they have a choice set of multiple alternatives, among which respondents pick the best / worst (similar to object scaling), but the alternatives are described by attributes (as in attribute scaling), so respondents do not have to go only by their beliefs. As defined by Lancsar et al. (2013): "*BWDCE is the type of BWS closest to traditional DCEs. Like standard DCEs, BWDCEs involve respondents making repeated choices between alternatives offered in choice sets, each described by a number of attributes*." Unlike traditional DCEs however, they give more information to researchers, as respondents make additional alternating best and worst decisions within the same choice set until it is exhausted, meaning that a single alternative is left.

This is a very valuable property of BWDCEs, as obtaining a sufficient number of responses / observations is of crucial importance for estimating choice models that are used for demand projections of future products and services. Other ways of obtaining a larger number of observations from first-choice experiments (traditional DCEs) are either getting more individuals to take part in the survey, which can be costly, or providing the respondents that do take part in the survey with a larger number of choice sets, which can be overburden the respondents mentally (Lancsar et al., 2013).

2.2.1 Modelling SBWDCE with SBWMNL

From a modelling perspective, the data output from SBWDCEs is an implied ranking of alternatives within each choice set. An implicit order of alternatives can be modelled by a rank ordered logit (ROL) or exploded logit, as each choice set is "exploded" into *J-1* choice sets, where *J* is the number of alternatives in the original choice set. The implied order of alternatives can thus be modelled as separate multinomial logit models, where in each round the best performing alternative from the previous choice set is removed until only one is left. For example, in an initial set of five alternatives (A, B, C, D, E), with the implied ranking of $V_A > V_B > V_C > V_D > V_E$, where V_i represents the structural utility of alternative i, the probability of observing that ranking can be computed as show in Equation 2.7.

Equation 2.7. Model formulation of a ranked order logit model (Lancsar et al., 2013)

$$P\ (V_A > V_B > V_C > V_D > V_E) = \frac{e^{V_A}}{\sum_{j=A,B,C,D,E} e^{V_j}} \cdot \frac{e^{V_B}}{\sum_{j=B,C,D,E} e^{V_j}} \cdot \frac{e^{V_C}}{\sum_{j=C,D,E} e^{V_j}} \cdot \frac{e^{V_D}}{\sum_{j=D,E} e^{V_j}}$$

However, as Lancsar et al. (2013) pointed out, this does not accurately represent the way the data was gathered. Respondents in the BWDCE were not asked to rank the alternatives from best to worst, but rather to alternatingly choose the best and worst option from the choice set. They therefore present a modified formulation of the choice probabilities: The Sequential Best Worst Multinomial logit (SBWMNL). By modelling the choice set in the same way it was presented to the respondents, with the best and worst choices being considered, a slightly different formulation, for the same implied ranking of alternatives ($V_A > V_B > V_C > V_D > V_E$) was constructed and is presented in Equation 2.8.

Equation 2.8. Model formulation of the SBWMNL model (Lancsar et al., 2013)

$$P\ (V_A > V_B > V_C > V_D > V_E) = \frac{e^{V_A}}{\sum_{j=A,B,C,D,E} e^{V_j}} \cdot \frac{e^{-V_E}}{\sum_{j=B,C,D,E} e^{-V_j}} \cdot \frac{e^{V_B}}{\sum_{j=B,C,D} e^{V_j}} \cdot \frac{e^{-V_D}}{\sum_{j=C,D} e^{-V_j}}$$

As can be seen, the choice sets reflect the same order as the survey. What is important to note is that the worst alternatives are indeed modelled as worst. This is modelled in a relatively straightforward way, by simply taking the negative of all the utilities. Since the worst alternative in a given choice set has the lowest utility for the respondent, that means it has the highest disutility (disutility being the negative of utility), meaning it will have the highest probability of being observed as the worst alternative. This is the key difference between modelling an ROL and an SBWMNL.

2.2.2 Applications of BWDCE

Being a relatively new entrant in research, best worst discrete choice experiments have seen fairly limited application and empirical examples. In addition to providing this new type of preference elicitation method, Louviere et al. (2008) presented two empirical examples of their newly developed model. One example was for the choice of two types of food products: delivered pizzas and packaged juices. The second empirical example, more relevant in the field of transport, is the choice among different flight options. Although done on a very small sample of only 12 student volunteers, it is a good proof of concept. Each respondent was presented with 16 choice sets, containing four generic alternatives each. Flight options were described by three four-level attributes (the return cost, number of stops and travel time) and three two-level attributes (aircraft type, airline and whether food was provided as well or only drinks). Respondents then had to make three decisions within each choice set: first selecting the best alternative, then the worst and finally the best of the two remaining.

A more recent application was carried out by Lancsar et al. (2013) on the topic of cardiac arrest treatment in public places. The experimental design consisted of 16 choice sets, of five alternatives, described by three attributes. One of the alternatives was the status quo (meaning the attribute levels were constant throughout the survey), while the other four were varied. Two of the attributes had four levels and one had two levels. Because the same attributes were used in the status quo and other alternatives, all attributes were generic. The data was analysed using first-best data with the traditional MNL approach and using the full dataset with the ROL and SBWMNL model. All three models were expanded to mixed logit, allowing for preference heterogeneity and for both preference and scale heterogeneity. For all three types of data (first-best, implied ROL, SBW) the preference heterogeneity ML model performed best, having the lowest AIC and BIC values, while the other two models performed equally well.

2.3 Random Regret Minimisation

In the field of choice modelling, by far the most recognisable is the random utility model, known and used for its ease of application and understanding. However, as stated in subchapter 0, its fully compensatory behaviour has often been put into question. One solution to provide semi-compensatory behaviour was provided in the form of the random regret minimisation model by Chorus (2010) .

RRM models are also a relatively new choice modelling approach like the SBWMNL. Unlike the SBWMNL however, they have seen substantial attention in the scientific community, with various researches comparing it to the RUM model in different contexts. Several extensions and generalisations of the RRM have also been proposed. The first random regret minimisation model was proposed by Chorus et al. (2008), but was later approximated to an improved model, derived by Chorus (2010). The initial model used two sets of max operators and while it was correct in terms of behavioural representation, it presented computational difficulties (function discontinuity at 0) and possibly most importantly, could not be used to determine the willingness to pay (WtP) or elasticities (Chorus et al., 2014).

The improved approximated RRM model still compares the attributes between each two alternatives and performing better on one attribute still does not cancel out performing worse on another attribute to the same degree, but in the approximated model, the regret function is continuous. The model formulation is presented in Equation 2.9.

Equation 2.9. Random regret minimisation model (Chorus, 2010)

$$R_i = \sum_{j \neq i} \sum_{k=1 \ldots K} ln\left(1 + exp\left(\beta_k \cdot (x_{jk} - x_{ik})\right)\right)$$

where:	R_i	*systematic regret of alternative i*
	β_k	*taste parameter of attribute m*
	x_{ik}, x_{jk}	*attribute levels of alternatives i and j*

Like with utilities, there are two types of influences on the behaviour of people in RRM theory: observed and unobserved. Combining the observed and unobserved regret of people, we get total random regret, as presented in Equation 2.10, where i represents an alternative within a choice set.

Equation 2.10. Obtaining random regret from systematic regret and unobserved random regret (Chorus, 2010)

$$RR_i = R_i + \upsilon_i$$

where:	RR_i	*total random regret of alternative i*
	υ_i	*unobserved random regret of alternative i*

To then obtain choice probabilities, in which the random aspect of RRM is included (similarly as it is with structural utility), the negative regret values are put into the logit function in the same way as for a RUM model. Using negative values of the systematic regret is the only difference between the computation of choice probabilities from utilities and from regret. Unlike utility, larger regret makes alternatives less appealing, so to estimate the best alternative, negative regret is taken. The mathematical formulation is presented in Equation 2.11 (with the same use of notation as in Equation 2.9).

Equation 2.11. Logit model for RRM (Chorus, 2010)

$$P_i = \frac{e^{-R_i}}{\sum_{j=1 \ldots J} e^{-R_j}}$$

2.3.1 RRM generalisations

Following the development of RRM models, researchers have already put forward improvements of the model. Here, we focus on two generalisations made to RRM models, as they gave the model a far greater applicability and allow the quantification of compensatory behaviour.

The first is the Generalised RRM (G-RRM) model (shown in Equation 2.12), which has an additional variable: the regret weight γ_m in place of the number 1 in the regular RRM model. This allows the compromise effect for each attribute to vary independently between a RUM and an RRM model, meaning that the researcher does not need to infer a priori the decision rule, but can allow the data to speak for itself (Chorus, 2014a). A γ value of 0 is analogous to fully compensatory behaviour (like in a RUM model), while a value of 1 represents semi-compensatory behaviour (RRM model). Depending on the value between 0 and 1 that the γ takes, the decision behaviour can be evaluated.

Equation 2.12. The G-RRM model formulation (Chorus, 2014a)

$$R_i = \sum_{j \neq i} \sum_{k=1 \ldots K} ln\left(\gamma_k + exp\left(\beta_k \cdot (x_{jk} - x_{ik})\right)\right)$$

where:	γ_k	*regret weight*

An even more generalised version of the RRM model was proposed by van Cranenburgh, Guevara, et al. (2015) with the μRRM model. In addition to allowing the attribute decision rule to vary between RUM and RRM (as in the case of G-RRM), the μRRM also includes Pure RRM (P-RRM). The latter is a version of RRM, where performing better on one attribute yields no rejoice for the alternative overall. The formulation of μRRM is shown in Equation 2.13. The P-RRM, regular RRM and RUM model are all defined by the value of the scale parameter μ, which is estimated in the model along with the taste parameters. Examples of all three situations are presented in Table 2.1. By allowing such flexibility of the data, the researcher does not need to make any assumptions on how compensatory the behaviour of respondents is, the model can handle anything from non-compensatory to fully compensatory behaviour.

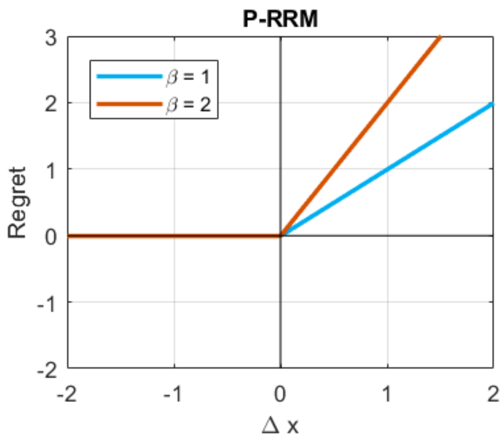
At this point, a distinction needs to be made between the scale parameter used traditionally in RUM models and the scale parameter of the μRRM model. The scale parameter in the μRRM model is interpreted as the level of compensatory behaviour utilised by the respondents for a certain attribute or across all choices. This scale parameter is one that is referred to throughout the report. In RUM models however, the scale parameter is related to the variance of the error term, which is defined as var(ε) = $(\pi^2/6)$ / μ^2. As this cannot be jointly estimated with taste parameters (β), the variance is fixed by fixing the μ scale parameter to one (Train, 2009). When var(ε) is large (μ is small), choices become more or less random, whereas a small var(ε) (large μ) means very deterministic choices (Chorus, 2016). The scale parameter is therefore related to choice consistency.

Equation 2.13. The μ-RRM model formulation, which extends the regular RRM model with the variable μ (van Cranenburgh, Guevara, et al., 2015)

$$R_i = \sum_{j \neq i} \sum_{k=1 \dots K} \mu_k \cdot ln\left(1 + exp\left(\frac{\beta_k}{\mu_k} \cdot (x_{jk} - x_{ik})\right)\right)$$

where: μ_k *scale parameter*

Table 2.1. Comparison of P-RRM, RRM and RUM decision rules (for a single attribute) (Chorus, 2016)

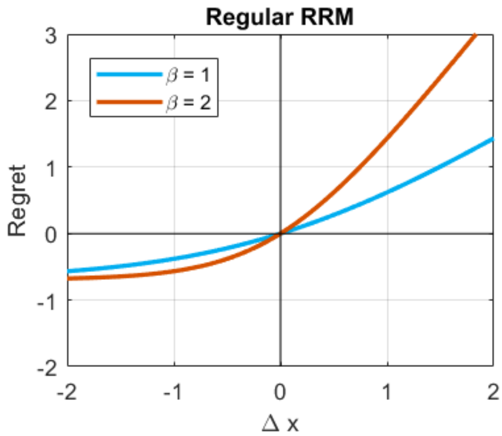


Pure Random Regret Minimisation model

Non-compensatory behaviour

μ → 0

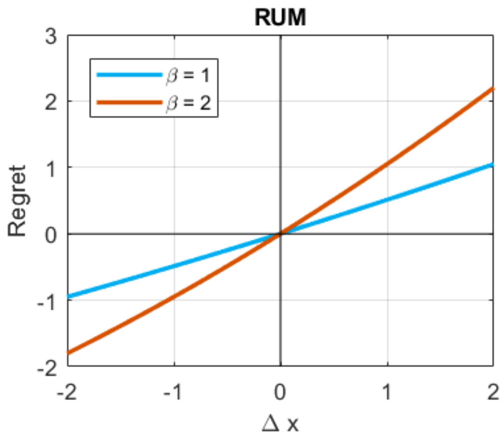
(a value of 0,01 is used for μ in the graph, as μ never reaches 0)



Random Regret Minimisation model

Semi-compensatory behaviour

μ = 1



Random Utility Maximisation model

Fully compensatory behaviour

μ → ∞

(a value of 100 is used for the μ in this graph)

2.3.2 Applications of RRM

Since their development, RRM models have been used, compared and analysed in countless researches. Chorus, Rose, & Hensher (2012) investigated how hybrid decision rule models compare to single decision rule models in terms of model fit. RUM and RRM models were compared on consumer preference for alternative fuel vehicles, with the results highlighting that although the model fits are similar, the policy implications are quite different, especially due to the compromise effect, captured by the RRM model (Chorus, Koetse, & Hoen, 2013). The different decision rules were also compared by Hess & Stathopoulos (2013), who found that different people utilise different decision rules in the same choice situations (subconsciously). Chorus (2014b) analysed the benefits of adding alternatives to the choice set of both RUM and RRM models, the effect of having opt-out alternatives on both decision rules was investigated by Hess, Beck, & Chorus (2014) and the robustness of RRM models to missing relevant attributes was assessed by van Cranenburgh & Prato (2016). Additional insight into the value of time derivation for RRM models, a crucial element for designing policies, was provided by Dekker (2014). Mai, Frejinger, & Bastin (2015) compared the performance of RUM and RRM for the specific case of route choice. This was also done on a larger transport model scale by van Cranenburgh & Chorus (2017), who found considerable differences in the projections of both models.

Clearly, much research has been done on the topic of random regret minimisation modelling. Chorus et al. (2014) created an overview, to which the interested reader is referred for a more complete summary of applications. The paper assesses 43 applications of both RUM and RRM models and compares the empirical evidence.

2.4 Constructing experimental designs

Selecting a survey design can have big implications on the outcome of the model estimation (van Cranenburgh, Rose, & Chorus, 2018), so the design needs to be chosen carefully, with the model estimation step in mind. In the past, orthogonal designs were the most widely used for stated choice surveys. Orthogonal designs preserve attribute level balance and minimise the correlations among attributes. This reduces standard errors of parameter estimates which are reliable and unbiased (Molin, 2016a). Recently, efficient designs have become the norm in choice modelling studies. By using prior values for parameters, they avoid presenting the respondents with dominant alternatives and they help minimise the standard errors of estimates. However, as pointed out by Walker et al. (2018), efficient designs assume that the priors used for constructing the survey are perfect (or close to perfect) and only perform best if that is true. As the difference between the prior parameter values and the actual values increases, efficient designs become more inefficient. For larger differences between the prior and true parameter value, Bayesian designs have been found to be more robust, while for even larger differences, orthogonal designs have proven to be the most robust. In their study, Walker et al. (2018) used a prior VoT of 20$/h and found that D-efficient designs (the most widely used of the efficient designs) are at their best only when the true population VoT is between 10$/h and 30$/h. Bayesian designs performed better but still degraded above the values of 50$/h for smaller variances and above 80$/h for larger variances. Above those values, orthogonal designs without dominant alternatives are the best choice.

Using a Bayesian design therefore seems preferable when some uncertainty is assumed. However, all efficient designs, including Bayesian designs, are biased towards RUM, because the designs are evaluated based on their D-errors, which are calculated on the basis of a RUM decision rule. Van Cranenburgh et al. (2018) determined that (RUM based) efficient designs can be inefficient if (P-)RRM is the true population decision rule and vice-versa. They developed a methodology for generating efficient designs that do not make assumptions towards the decision rule of the respondents (RUM or (P-)RRM). This is done by determining both the RUM and P-RRM D-errors. In their paper, they provide the method to construct the Fisher information matrix for a P-RRM decision rule, from which the respective efficiency of a design can be determined (through the AVC matrix). The formulation for deriving the Fisher information matrix is presented in Equation 2.14. The difference in doing it for a P-RRM-efficient design is that the choice probabilities need to be calculated with P-RRM (Chorus et al., 2008), as opposed to using RUM and the attribute levels (x) change to the summation of all the differences in attribute levels that contribute to experiencing regret. Different formulations are used for two generic attributes, two alternative specific attributes or a generic and an alternative specific attribute. Because the experimental design in this research will only contain generic attributes, only the relevant equation is given. The interested reader is referred to Rose & Bliemer (2005) and Choice Metrics (2018) for deriving the AVC matrix for RUM-efficient designs and to van Cranenburgh et al. (2018) for deriving RRM-efficient designs.

Equation 2.14. Derivation of the Fisher Information matrix for two generic attributes (van Cranenburgh et al., 2018)

$$\frac{\partial^2 L(\beta^2,\beta)}{\partial\beta *_{k1}\,\partial\beta *_{k2}} = -\sum_{s=1}^{S}\sum_{j=1}^{J}\left(x *_{jk1s}\cdot P_{js}\cdot\left(x_{jk2s}-\sum_{j=1}^{J}(P_{is}\cdot x *_{ik2s})\right)\right) \quad \forall k_1,k_2\in K *$$

$$x *_{iks}=\begin{cases}\sum_{j\neq i} max(0,x_{jks}-x_{iks})\ if\ \beta_k>0\\ \sum_{j\neq i} min(0,x_{jks}-x_{iks})\ if\ \beta_k<0\end{cases}$$

where:	β^*_{k1}	*parameter of generic attribute k_1*
	x^*_{ik1s}	*sum of attribute levels that contribute to experiencing regret for attribute k_1 in alternative i and choice set s*
	P_{is}	*choice probability of alternative i in choice set s*
	K^*	*set of generic attributes*
	β_k	*parameter value of attribute k*

Having outlined the method to determine the D-error based on the P-RRM decision rule, van Cranenburgh et al. (2018) present a way to evaluate experimental designs based on both their RUM and P-RRM D-error, by introducing the composite D-error, calculated as shown in Equation 2.15. The most efficient design without making any assumptions on the decision rule is thus one with the lowest composite D-error.

Equation 2.15. Composite D-error of an experimental design (van Cranenburgh et al., 2018)

$$D_{composite}=\frac{1}{2}\cdot D_{RUM}+\frac{1}{2}\cdot D_{P-RRM}$$

where:	$D_{composite}$	*composite D-error of both RUM and P-RRM*
	D_{RUM}	*RUM-based D-error*
	D_{P-RRM}	*P-RRM based D-error*

2.5 Conclusion

This chapter has given an overview of discrete choice analysis and several of its components that contribute to the ever-expanding pool of knowledge in this field of research. The relatively new fields of choice modelling: SBWDCEs and RRM models were investigated in detail. SBWDCEs are interesting as they provide more information to the researcher on the trade-offs a person has made, yet at the same time trying not to overburden the respondents. However, when considering different types of decisions being made (best and worst choices), the question arises of whether respondents may utilise different decision rules for accomplishing those decisions. That can be evaluated using a combined RRM and SBWMNL model to determine if it improves the model's ability to capture human behaviour. The µRRM model is particularly well suited for this as it allows its scale parameter to vary between RUM, RRM and P-RRM (from fully compensatory to non-compensatory behaviour respectively).

The chapter concluded by analysing the advancements in experimental design construction, in particular when wishing to remain unbiased towards the decision rule. A method of constructing and evaluating experimental designs on their RUM and P-RRM robustness was described and is used in this research for the generation of the choice situations.


715 - 121

3 Methodology

This chapter provides the framework for how the research questions will be answered, and the hypothesis tested. The first subchapter looks at the model estimation and testing technique(s) to be utilised for answering the presented research questions, as well as how the model(s) will be validated. The descriptions of several different models are presented and a list of models to be estimated is given. Given the model characteristics, an opportunity to evaluate the model behaviour will be analysed in the estimation process.

Subchapter 3.2 then gives an overview of the case study selected for conducting the survey and the estimation of parameters. The selected topic for the survey is park-and-ride facility selection and the area will be focused on the Central Slovenia Statistical Region. After the model estimation, the obtained results will be used in the model application for forecasting park-and-ride facility choice.

Subchapter 3.3 gives a detailed description of the experimental design construction, starting with the required characteristics the survey needs to meet, due to the requirements of the utilised estimation methods. The stated choice survey is constructed by selecting the relevant attributes and corresponding attribute levels. Additionally, a regret scale that measures how much regret respondents experience is added in the interest of observing if respondents are maximisers or satisficers. This information can also prove valuable in carrying out latent class analyses. Socio-demographic and travel related questions are added to get insight into the background information of the respondents and again to be used in latent class analyses for the class allocation function. The subchapter concludes by reviewing the results of the pilot survey and choosing an experimental design for the main stated choice survey.

The data gathering technique, when and how the data was collected is described in subchapter 3.4.

The chapter concludes by summarising the socio-demographic characteristics of the respondents. The obtained socio-demographic data is analysed and compared to the typical population characteristics of the Central Slovenia Statistical Region. Although this is not the true population that is being investigated, it was the best population for which the data was available. The true population is made up of commuters from outside the city of Ljubljana that drive into the city on a daily basis.

3.1 Model estimation

For answering the research questions, the model estimation that will be performed on the obtained data is key, as this will provide the answer and allow us to accept or reject our hypothesis. Given the peculiarities of the gathered data, the variation of choice set size has to be accounted for, which is explained in subchapter 3.1.1. The main description of how the data will be modelled and which models will be estimated and compared is given in 3.1.2 with subchapter 3.1.3 investigating model behaviour, the predictive ability of different models and evaluating prior parameter values.

3.1.1 Choice set size variation in RRM modelling

As RRM models compute the regret of each alternative based on the performance of all alternatives in the choice set, the number of alternatives influences the level of regret. This can cause problems when using choice sets with a varying number of alternatives, between two and five in the case of this research. Van Cranenburgh, Prato, & Chorus (2015) investigated the possibility of using a correction factor that also incorporated the choice set size. In simpler examples, where all possible choice set sizes are know in advance, they propose using constants for all except one choice set size. The correction factor is simply added to the regret of each alternative. In the SBWMNL μRRM, this will be done by taking the choice set with two alternatives as the base and then choice sets with three, four or five alternatives will each be given a correction factor: Λ_3, Λ_4 and Λ_5 respectively. For more flexibility, a single constant that is adjusted according to the choice set size can also be modelled. The formulation for the single choice set size constant, proposed by van Cranenburgh, Prato, et al. (2015) is presented in Equation 3.1.

Equation 3.1. Accounting for the choice set size variation with a single estimate (van Cranenburgh, Prato, et al., 2015)

$$\frac{J_n}{\Lambda}$$

where:	J_n	*choice set size in observation n*
	Λ	*estimated constant for choice set size variation*

3.1.2 Model estimation and hypothesis testing

The primary model estimation technique to be employed in this research will be the SBWMNL μRRM model with two scale parameters (μ) and three choice set size constants (Λ). The model is a combination of the SBWMNL model (Lancsar et al., 2013) described in chapter 2.2.1 and the μRRM model (van Cranenburgh, Guevara, et al., 2015) explained in chapter 2.3.1. The formulations of both model estimation techniques do not change from Equation 2.8 and Equation 2.13 respectively. The two values for the scale parameter (μ) are:

- *μ(Best)* used for modelling the best choices in the choice set
- *μ(Worst)* used for modelling the worst choices in the choice set

By allowing the scale parameter to vary in such a way, its value will provide the resulting decision rule utilised by the respondents. When the value of the scale parameter approaches zero, it indicates a P-RRM decision rule (non-compensatory behaviour), a value towards infinity indicates a RUM (fully compensatory behaviour) and a value close to one means a regular RRM decision rule (semi-compensatory behaviour) (van Cranenburgh, Guevara, et al., 2015).

3.1.2.1 Implied decision rule combinations

Intra-person variability of decision rules in best-worst choice experiments will be investigated with two scale parameters (one for best and one for worst choices) that will be imposed rather than estimated in the model. That way, all four combinations of decision rules can be tested for their model fit. The four combinations are presented in Table 3.1.

Table 3.1. Four models with different combinations of implied decision rules

		I.1	**I.2**	**I.3**	**I.4**
Decision	**Best choices**	RUM	P-RRM	RUM	P-RRM
rule	**Worst choices**	RUM	P-RRM	P-RRM	RUM

These four models will be compared by means of different statistical tests. The first of these is the McFadden's' rho-squared, which is an indication of how much the given model has reduced uncertainty of the data (Hauser, 1978). This is done by comparing at the null and final log-likelihoods, as shown in Equation 3.2. The null log-likelihood is the likelihood of a model where all parameters are zero.

Equation 3.2. McFadden's' rho-squared formula (Hauser, 1978)

$$\rho^2 = 1 - \frac{LL_\beta}{LL_0}$$

where:	ρ^2	*rho-squared*
	LL_β	*final log-likelihood*
	LL_0	*null log-likelihood*

Secondly, comparing non-nested models can be done with the help of the Ben-Akiva and Swait test. The test gives the probability of a worse fitting model (model B) being the true model of the population, compared to a better fitting model A (Chorus, 2016). The formulation of the Ben-Akiva and Swait test is presented in Equation 3.3.

Equation 3.3. Ben-Akiva and Swait test for comparing non-nested choice models (Ben-Akiva & Swait, 1986)

$$Pr\ (\rho_B^2 - \rho_A^2 \geq z) \leq \Phi\left(-\sqrt{2 \cdot N \cdot z \cdot ln(J) + (K_B - K_A)}\right)$$

$$\rho^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(0)}$$

where:		
	Pr	*probability of model B being the better fitting model*
	ρ^2	*rho-squared model fit*
	Φ	*standard normal cumulative distribution function*
	N	*number of observations*
	z	*difference between model fits*
	J	*number of alternatives in the choice set*
	K	*number of parameters*
	$LL(\beta)$	*log-likelihood of the associated model*
	$LL(0)$	*null log-likelihood*

3.1.2.2 Estimation of decision rule

In the developed SBWMNL μRRM model, there are different options of accounting for choice set size variation and a different number of scale parameters can be used for certain choices. Choice set size can be accounted for either by utilising a single constant, that is adjusted according to the choice set size. The alternative is to assign each choice set size except one its own choice set size constant, that is estimated independently. This means that in this research, we can either use one or three choice set size constants. Regarding scale parameters, a single μ value can be imposed across all four choices within a choice set, two separate values can be used for best (first and third) and worst (second and fourth) choices, or each of the four choices can have its own scale parameter. For investigating the main research hypothesis, four models with either two or four scale parameters and one or three choice set size constants are developed and shown in Table 3.2. In addition, a model with a single scale parameter value is also estimated, with the goal of comparing it to the SBWMNL RUM model (Lancsar et al., 2013) to test if imposing a RUM decision rule across all choices is most valid if a single decision rule across all choices is assumed. The developed models are compared to three existing and prominent models in choice modelling: the first-choice RUM model (E.1), first-choice μRRM model (E.2) and the SBWMNL RUM model (E.3). All the models to be estimated and compared are summarised in Table 3.2.

Table 3.2. Models to be estimated and compared in this research

	E.1	**E.2**	**E.3**	**E.4**	**E.5**	**E.6**	**E.7**	**E.8**
	First-choice RUM model	First-choice μRRM model	SBWMNL RUM model	SBWMNL μRRM models				
# of μ				*1*	*2*	*4*	*2*	*4*
# of Λ				*3*	*3*	*3*	*1*	*1*

Model E.1 is estimated because it is the most commonly applied in choice modelling research. E.2 is estimated to see if model E.1 is correct to assume a RUM decision rule in first-best choices made by respondents. The benchmark model for SBWMNL modelling is model E.3, to which the remaining five SBWMNL models will be compared to. E.4 is estimated to evaluate if a single decision rule is inferred, is it correct to assume a RUM decision rule, as is done in E.3. Models E.5 to E.8 will evaluate the differences in compensatory behaviour between best and worst choices and will also investigate if the specification of choice set size variation has an impact on the model outcome.

The models will be compared using the rho-squared and the Ben-Akiva and Swait test, as mentioned in subchapter 3.1.2.1, with the formulation presented in Equation 3.2 and Equation 3.3 respectively.

Comparing the model outcomes will also be done with the help of the Likelihood Ratio Test, which compares two nested models. The test relies on comparing models with different numbers of attributes. In Equation 3.4, model A is the better fitting model and the LRT checks if this is due to sample peculiarities. The obtained likelihood ratio statistic (LRS) is used to find the significance level in the χ^2 table, where the difference in the number of estimated parameters represents the degrees of freedom. The obtained critical values from the χ^2 table give the threshold of significance (Chorus, 2016).

Equation 3.4. Likelihood ratio test formulation (Train, 2009)

$$\mathrm{LRS} = -2 \cdot (LL_B - LL_A)$$

where:	*LRS*	*likelihood ratio statistic*
	LL	*log-likelihood of the associated model*

Additionally, the models can also be compared with the Bayesian information criterion (BIC) (Stone, 1979). Unlike the other tests, the BIC is not a statistical measure but an arbitrary number. The BIC values of different models can be compared, and the lower BIC value indicates a better fitting and more parsimonious (with respect to the number of parameters) model. Including a larger number of parameters means models are more likely to have a better fit to the data, so including this variable in the test adds value. The formulation of the BIC is presented in Equation 3.5.

Equation 3.5. The formulation of the Bayesian information criterion (Stone, 1979)

$$BIC = ln(n) \cdot k - 2 \cdot LL$$

where:	*BIC*	*Bayesian information criterion*
	k	*the number of estimated parameters*
	n	*the number of observations (data points)*
	LL	*log-likelihood of the associated model*

Evaluating the rate of compensatory behaviour (fully to non), profundity of regret for attributes in models E.4 to E.8 will be computed and analysed. Profundity of regret (α_m) examines the extent of regret minimisation behaviour imposed by an RRM model on different attributes. When the value of α_m approaches the value of zero, the behaviour approaches fully compensatory, whereas a value close to one indicates non-compensatory behaviour (van Cranenburgh, Guevara, et al., 2015). The mathematical formulation is presented in Equation 3.6.

Equation 3.6. Formulation of attribute profundity of regret (van Cranenburgh, Guevara, et al., 2015)

$$\alpha_m = \frac{1}{|A_m|} \cdot \sum_{A_m} \left(\left| \frac{e^{\frac{\beta_m}{\mu} \cdot (x_{jmn} - x_{imn})} - 1}{e^{\frac{\beta_m}{\mu} \cdot (x_{jmn} - x_{imn})} + 1} \right| \right); \quad A_m = \{x_{jmn} - x_{imn} | x_{jmn} - x_{imn} \neq 0\}$$

where:	α_m	*Profundity of regret of attribute m*
	A_m	*set of non-zero differences of attribute m levels*
	β_m	*taste parameter of attribute m*
	μ	*scale parameter*
	x_{jmn}	*attribute level of choice set n, alternative j, attribute m*

3.1.2.3 Interpersonal variability

In addition to investigating decision rule variability on an individual level, differences between individuals and their use of decision rules can also be tested. Hess, Stathopoulos, & Daly (2012) investigated four separate datasets for decision rule variability between respondents and found significant improvements in model fit when accommodating for this variability, which was done with the help of LC models. One of the datasets was tested for respondents being either RUM or RRM decision makers and found that roughly 40% were regret minimisers, with the rest being utility maximisers. This finding encourages the investigation of interpersonal variability in this research as well. Although there exists several different models that capture such variability (most notably the mixed logit model), latent class modelling is used for its rather simple application (no need to make assumption on the distribution of heterogeneity), the possibility of using socio-demographic variables for class membership and ability to easily compare different classes (Greene & Hensher, 2003).

In latent class modelling, estimated parameters are latent-class specific, allowing them to vary for different groups of individuals. These groups each have a class membership probability, which is the probability of being part of a certain class (van Cranenburgh, 2017). The extension to the single-class choice model is presented in Equation 3.7.

Equation 3.7. Latent class model for cross-sectional data (Hess, Ben-Akiva, Gopinath, & Walker, 2008)

$$P_n(i|\beta) = \sum_{s=1}^{S} \left(\pi_{ns} \cdot P_n(i|\beta_s)\right)$$

where:	$P_n(i\|\beta)$	*Probability of decision-maker n, choosing alternative i, conditional to parameters b*
	β_s	*set of parameters in class s*
	π_{ns}	*probability of decision-maker n belonging to class s*

The class membership probability can be static (using only the constant δ_s) but can also be extended to allow socio-demographic variables to help predict class membership of individuals. Determining this is done by a logit model, which determines the class probabilities (van Cranenburgh, 2017). The formulation is presented in Equation 3.8. One of the class's 'utility' is fixed to zero as a reference point (usually class s=1), so δ_1 and γ_{1d} equal zero.

Equation 3.8. Class allocation function in a latent class choice model (Hess et al., 2008)

$$C_{ns} = \delta_s + \gamma_{sd} \cdot D_n$$

$$\pi_{ns} = \frac{e^{C_{ns}}}{\sum_{j=1...S} e^{C_{nj}}}$$

where:	C_{ns}	*utility of class s for decision-maker n*
	δ_s	*class specific constant*
	γ_{sd}	*parameter for socio-demographic D for class s*
	D_n	*socio-demographic value of decision-maker n*

As with the single class models, the LC models will be estimated with implied decision rule combinations (from LC.1 to LC.6) and with the scale parameter being estimated as well (LC.0 model). The overview of all the latent class models to be tested is shown in Table 3.3. With respect to including socio-demographic characteristics, only the LC.0 model will be extended to include the parameter for socio-demographic characteristics in the class allocation function.

The LC.0 model will be compared to single class models (E.3 and E.5) with the help of the rho-squared and by applying the LRT (Equation 3.4). With respect to comparing the six LC models with implied decision rules, they will be evaluated based on their rho-squared and with the Ben-Akiva and Swait test.

Table 3.3. List of all latent class models to be estimated in this research

		LC.0	LC.1	LC.2	LC.3	LC.4	LC.5	LC.6
Class1	Best choice	*estimated*	RUM	RUM	RUM	P-RRM	P-RRM	P-RRM
	Worst choice	*estimated*	RUM	RUM	RUM	P-RRM	P-RRM	RUM
Class 2	Best choice	*estimated*	P-RRM	RUM	P-RRM	RUM	P-RRM	RUM
	Worst choice	*estimated*	P-RRM	P-RRM	RUM	P-RRM	RUM	P-RRM

3.1.3 Model performance analysis

Along with estimating the models to analyse choice behaviour and answer the stated research questions, it is interesting to analyse the performance of the model, particularly due to its complex nature, large number of parameters and new entry onto the choice modelling field. The aspects to be analysed are the μ–Λ relationship, the model performance through validation and the evaluation of parameter values used as priors in the experimental design step.

3.1.3.1 μ-Λ relationship

This research offers a unique opportunity to study the behaviour of the relationship between the scale parameter (μ) and the choice set size constant (Λ). As van Cranenburgh, Prato, et al. (2015) have established, the values are rooted in one another and given that the models in this research will look at different ways of modelling choice set size variation and vary the number of scale parameters, interesting findings may emerge from investigating this relationship.

3.1.3.2 Model validation

Validation is an important step for testing the predictive capabilities of the developed choice model. For this research, hold-out validation will be performed. As the sample size is expected to be small and to avoid not using observations for estimation purposes, a k-fold cross-validation technique will be applied; in particular the 3-fold cross-validation (Refaeilzadeh, Tang, & Liu, 2009). Using three-fold cross-validations means that the data will be split up into three (roughly) equally sized datasets. The models will then be estimated (trained) three times, each time with a different part left out for validation (testing), as shown in Figure 3.1. This way, all data points can equally contribute to the model estimation process and the models' predictive abilities are still tested (Refaeilzadeh et al., 2009). The models will be evaluated on their hit-rate in the hold-out dataset. In addition, the log-likelihood of each of the hold-out datasets will reported and compared. The models to be validated are the E.3, E.5, E.6, E.7 and E.8. Additional statistical tests (as presented in 3.1.2.1) will not be used, as they utilise the log-likelihood value of the model and values such as the number of observations, attributes etc. which will be the same in all three instances of the three-fold cross-validation process.

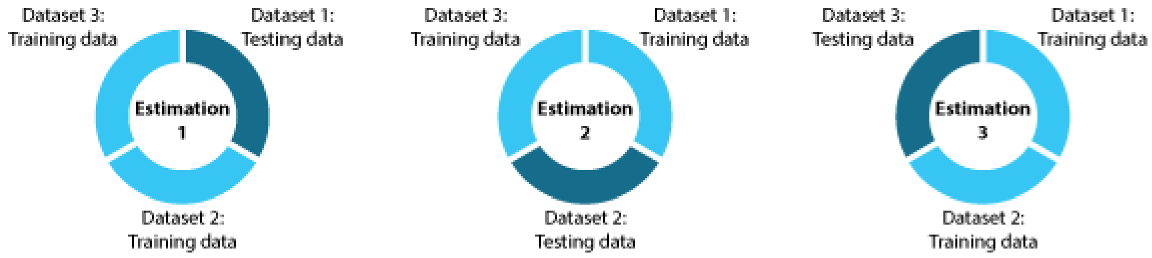


Figure 3.1. Representation of the three-fold cross-validation process

3.1.3.3 Evaluation of priors

The prior parameter values used in the experimental design generation step will be evaluated and compared to the parameters obtained from the model estimation. Comparing them with a statistical test is not reasonable, as there is no indication that the prior values are true for the population. It is more interesting to compare the parameter ratios and explore if they remained similar or if they were completely unsuitable for this survey. Based on this, the decision on the type of experimental design used for the survey can be evaluated.

3.2 Introduction to case study

The second subchapter of methodology describes the example case study that is used for the purposes of this research. The chosen topic is park-and-ride facility selection (dealt with in subchapter 3.2.1) and the chosen location is the Central Slovenia Statistical Region (CSSR), within which lies the capital and largest city Ljubljana, described in subchapter 3.2.2.

This case study is used for constructing the experimental design, by providing the attributes and attribute levels that can be associated with a P+R scheme in the region. Once the parameters are estimated, the models are applied on a case study on (a part of) the P+R scheme in the CSSR.

3.2.1 Park-and-Ride

It is interesting to research the choice of P+R facility in the light of moving towards sustainable mobility, especially since despite the widespread application of such parking lots, many studies have pointed out that the contribution of P+R is not always as positive as first thought. Many report of worsening road congestion and air quality (Meek, Ison, & Enoch, 2011) (Mingardo, 2013) (RPS, 2009), with city centres being the only to benefit from the measure (conditional to the application of stricter parking policies).

Spillar (1997, p. 9) defines park-and-ride lots as "*intermodal transfer facilities*". The basic format of P+R facilitates transfers between automobiles and public transport, by providing parking space and a good public transport connection. In a broader sense, P+R are also used as carpool lots, where multiple single-occupant vehicles are parked, and the passengers continue their journey with other drivers, improving the car load factor. The phenomenon of car passengers being dropped off at a transit station to board transit is called "*Kiss-and-ride (K+R)*" and is often present at P+R facilities as well. P+R can also be access by other modes, like feeder buses, bicycles or walking (Spillar, 1997), but these are not related to P+R as much as simply an access mode for public transport (in the case of walking or cycling) or a transfer in the case of a (feeder) bus service. P+R are attractive for such access modes and for K+R because they often offer public transportation connections with a higher level-of-service.

Their attractiveness for car users stems from abundant, low-cost parking, high accessibility (often located near highways) and solid public transportation links, which result in increasing levels of road congestion and air pollution. While their goal was redirecting traffic from city centres, they have in fact attracted new users into cars: (1) travellers who previously used public transportation for their whole journey and (2) new travellers who would not have made the trip otherwise. This means that although the traffic in the city stays the same or is reduced, the overall number of car journeys and the total vehicle distance travelled by everyone (by car) increases. The level of reducing congestion and air pollution in city centres seems more related to parking policies in city centres, rather than outside the centre (Spillar, 1997).

Despite the capability of P+R schemes to offer sustainable mobility, many current applications do not do so, but rather lead to an increase in the total car distance travelled. To better understand how the scheme needs to be developed, a classification of P+R lots in necessary. A detailed classification is provided by (Spillar, 1997). In this research, P+R facilities are divided into two broad categories, based on the much broader classifications of Spillar (1997):

- **Satellite P+R lots**
  Located at the edge of the city (built up area), often near a highway off-ramp, their main goal is to offer a cheaper alternative to parking in the city centre.
- **Local / Remote P+R lots**
  Positioned farther away from the city, usually located along a (higher quality) public transport corridor, they try to intercept commuters as close to their homes as possible.

Local and remote P+R lots could be separated, where local P+R lots would refer to smaller facilities found along a public transport corridor, while remote P+R lots are larger and located close to larger dormitory towns, from which many people commute to the respective regional centre.

City and regional P+R schemes are then made up of multiple parking lots and the sustainability of a scheme depends on the types of P+R lots and where they are located. In Europe, the traditional P+R concept (for the most part) means building large satellite P+R lots at the edge of cities, near highway off-ramps. According to Meek et al. (2011) such schemes are the least sustainable and contribute most to additional VMT (vehicle miles travelled). Meek et al. (2011) have evaluated five alternative concepts based on how much they reduce VMT and concluded that the Hub-and-Spoke (H&S) and Link-and-Ride (L+R) concepts resulted in the highest reduction of VMT from the current state. Three other concepts (Demand-led, Integrated and Remote site) did not result in the same level of VMT reduction. All five alternative concepts, along with the current concept can be seen in Figure 3.2.

The H&S concept proposes interchange facilities (P+R) farther from the city and with improved local public transport connections from rural areas to the interchange, where a high level-of-service (LOS) public transport is provided.

The link-and-ride concept, although the most different from the current practice (Meek et al., 2011), is more geared towards car users and thus may be more applicable in lower density rural areas. It envisages a high-quality public transport corridor with small interchanges (including P+R) located at regular intervals. This brings P+R closer to people's homes and allows for the largest reduction in VMT of all analysed concepts, despite not providing as much local public transportation as in the H&S concept for example.

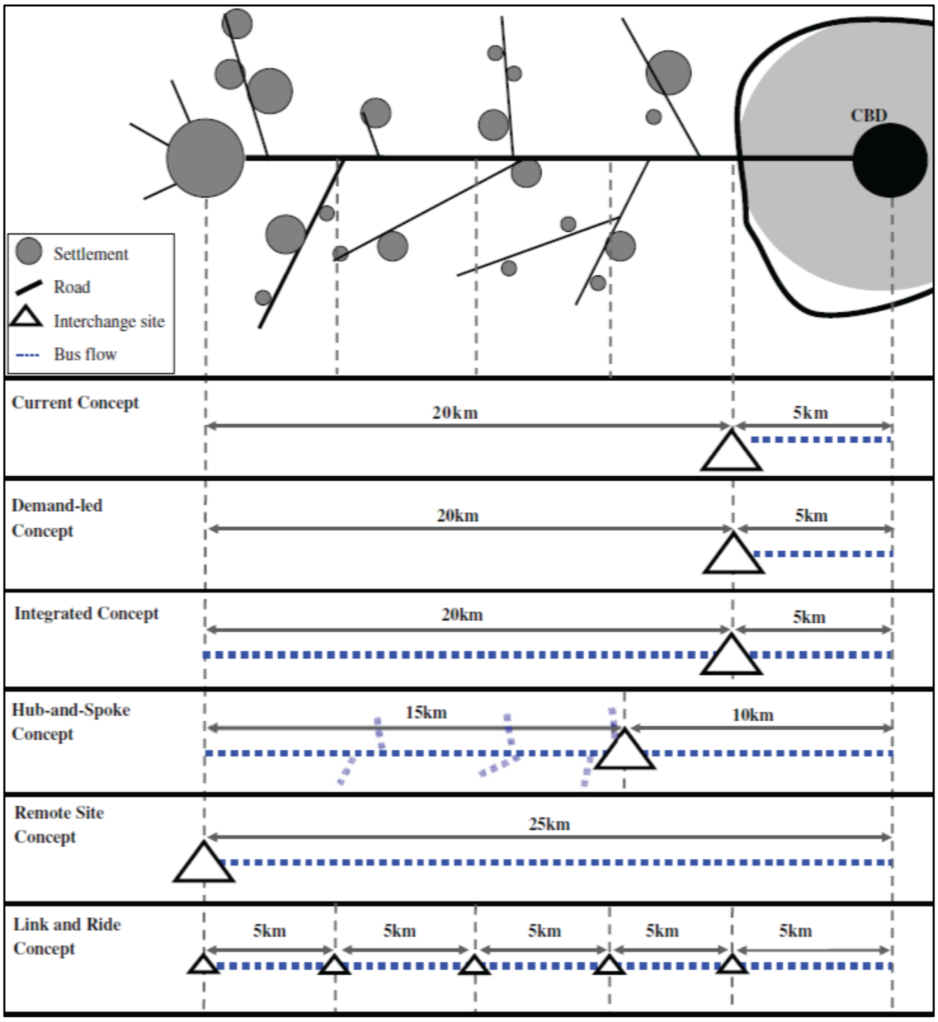


Figure 3.2. Alternative P+R concepts (Meek et al., 2011)

3.2.2 Central Slovenia Statistical Region

The Central Slovenia Statistical Region (CSSR) is located, as the name suggests, in the centre of Slovenia and containing within its borders the largest and capital city Ljubljana. A geographic representation, along with the population, size and population densities are given in Figure 3.3.

The central location at the intersection of major transport corridors makes Ljubljana the economic, educational, cultural and transportation centre of Slovenia, with 221.837 employees (Statistični Urad Republike Slovenije, n.d.-a) and over 40.000 students (Milanovič et al., 2017). Of those working in the capital, less than half also reside within the city municipality, while the rest drive from the CSSR, as well as from other regions, according to the work migration index (Statistični Urad Republike Slovenije, n.d.-a). The index considers the municipality where residents are registered and the municipality where they work. It does not include students, who also make up a significant share of daily commuters to the city. The work commute index for the relative share of municipal residents commuting to Ljubljana for work purposes is shown in Figure 3.4, with the regional work commute index shown in Figure 3.5.

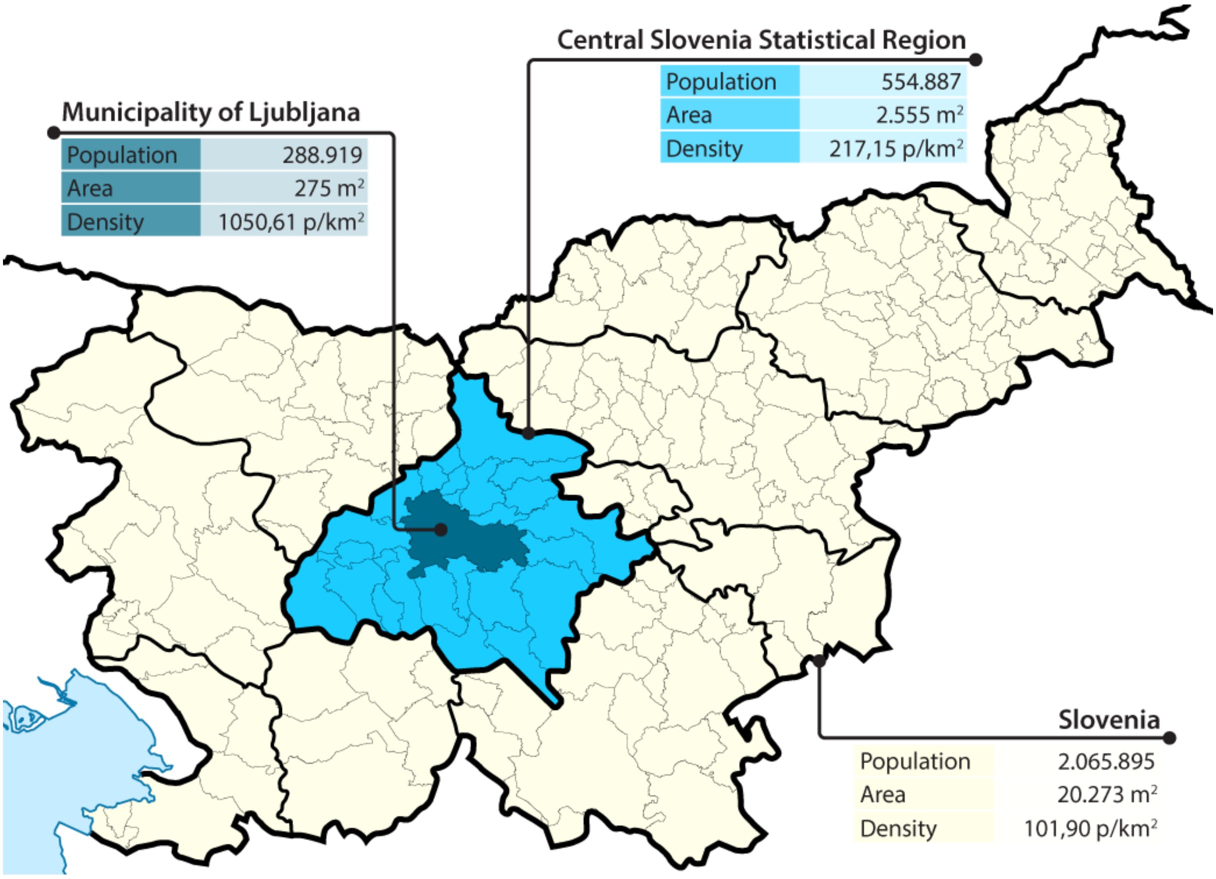


Figure 3.3. Location and statistics of the study area (STAGE, n.d.)

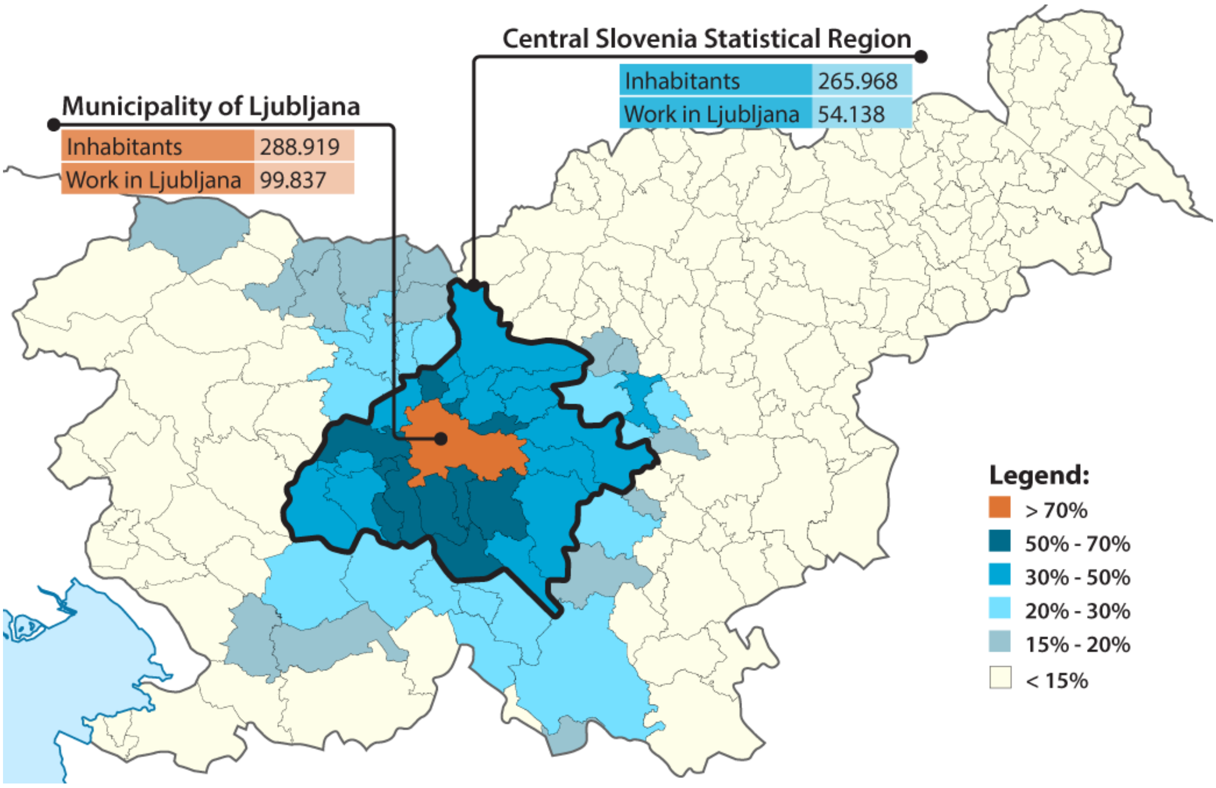


Figure 3.4. Municipal work commute index for the relative share of residents commuting to Ljubljana (STAGE, n.d.)

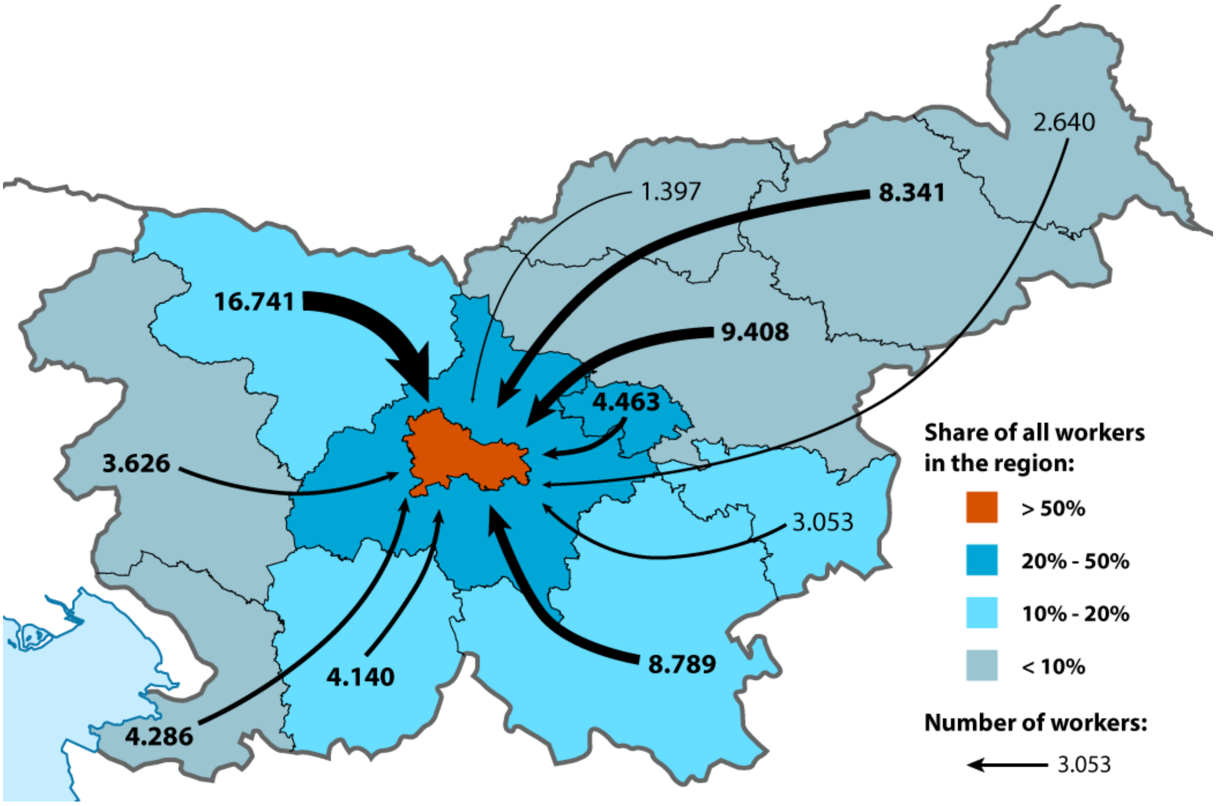


Figure 3.5. Regional work commute index for relative share and absolute number of residents commuting to Ljubljana (STAGE, n.d.)

Looking at the modal split, the vast majority of commuters from outside the Municipality of Ljubljana come to the city by car (over 80%), while only around 15% use public transport (Klemenčič, Lep, Mesarec, & Žnuderl, 2014). Coupled with the location at the intersection of major international transport corridors, this results in large traffic flows to and around the capital. The highway ring road is the most congested road in the country, with the highest annual average daily traffic (AADT) of 74.438 vehicles being measured on the western section of the ring road (Direkcija RS za Infrastrukturo, 2017). A map of AADT load on roads is presented in Figure 3.6, where the highways and the ring road can be clearly noticed.

By most estimates, over 100.000 cars enter the capital every workday, with some measurements reaching as high as 130.000 (Strniša, 2009), resulting in congestion on city streets as well, especially during morning and afternoon peaks.

The reasons for the current travel situation can largely be attributed to: low urbanisation level, high car ownership and the (lack of quality) spatial and transportation policies (Uršič, 2012). The urbanisation level is the lowest among EU countries at 49,6%, while the EU average stands at 75% (The World Bank, n.d.). Car ownership on the other hand scores above average at 523 per 1000 inhabitants, compared to the 497 EU average (Eurostat, 2017). All three factors influence each other, creating a vicious cycle where the absence of quality public transport means more people are dependent on cars for their mobility needs. With a car, they are more flexible and can move out of the city, reducing the urbanisation level and once living in the countryside, where public transport quality is even lower, they rely on cars even more as virtually the only means of mobility. This cycle was further reinforced through the construction of a highway network, while railways and public transportation did not see much government support.

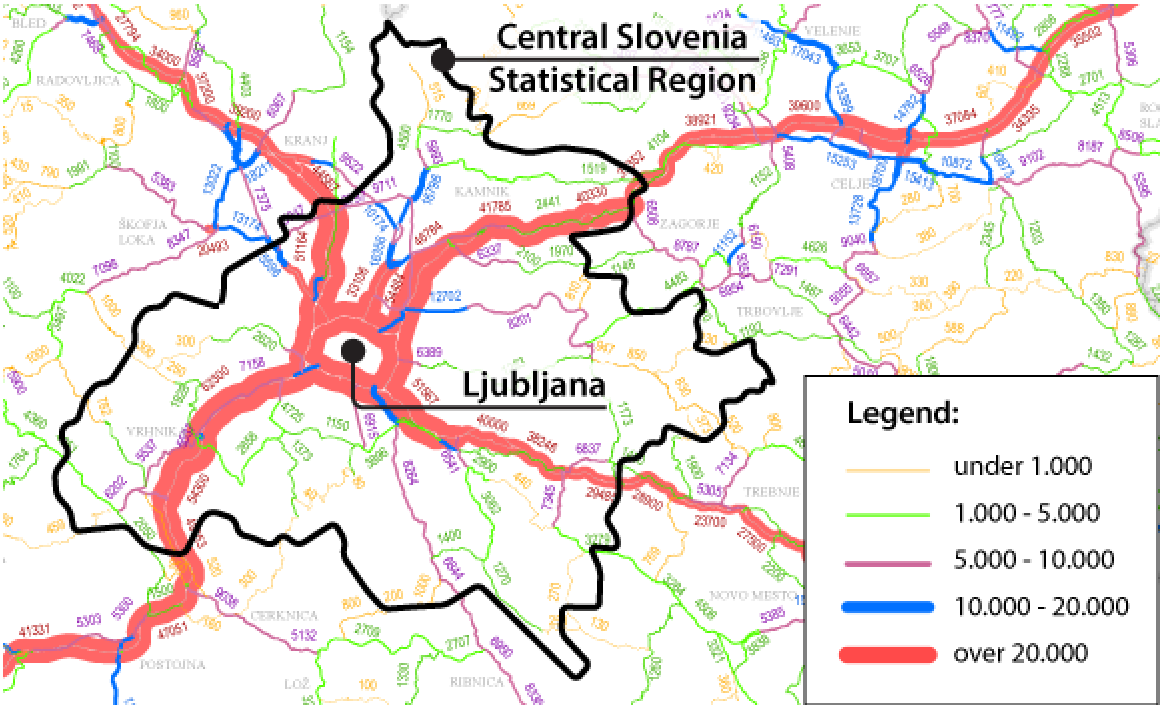


Figure 3.6. Map of Average Annual Daily Traffic (AADT) in and around the CSSR (Direkcija RS za Infrastrukturo, 2017)

In recent years, progress has been made. Long overdue railway upgrade works went underway, a national smart-card ticketing system (similar to the OV-chipkaart service in the Netherlands) is currently being piloted by students (Ministrstvo za Infrastrukturo, 2016) and the urban bus network has seen several changes and expansions since 2007, to better suit the changed activity pattern of residents.

One project is also the development of a park-and-ride scheme (RDA LUR, 2014). The project calls for 22 new parking lots, along with the redevelopment of an already existing P+R facility. Of these, a few would see reconstruction from a current unofficial P+R facility, but the majority would be new parking lots, some located at the edge of the city, others farther away. P+R in the CSSR are shown in Figure 3.7. For unofficial P+R lots, only those located at train stations are shown. A full list of official P+R facilities is given in Appendix A. The list does not include unofficial P+R, since although they are quite commonly found, they are hard to determine, as basically any parking lot can function as a P+R and determining a finite list of parking lots would require detailed research far beyond the scope of this research.

The development plan of the Regional Development Agency (RDA) can be seen as a hybrid of the traditional P+R concept, with large satellite P+R facilities surrounding the city and the link-and-ride concept with smaller parking lots dotted across the region along transit corridors. This plan provides a good opportunity to carry out an analysis of attracting users to L+R facilities as opposed to satellite P+R lots by offering very high levels of public transportation service.

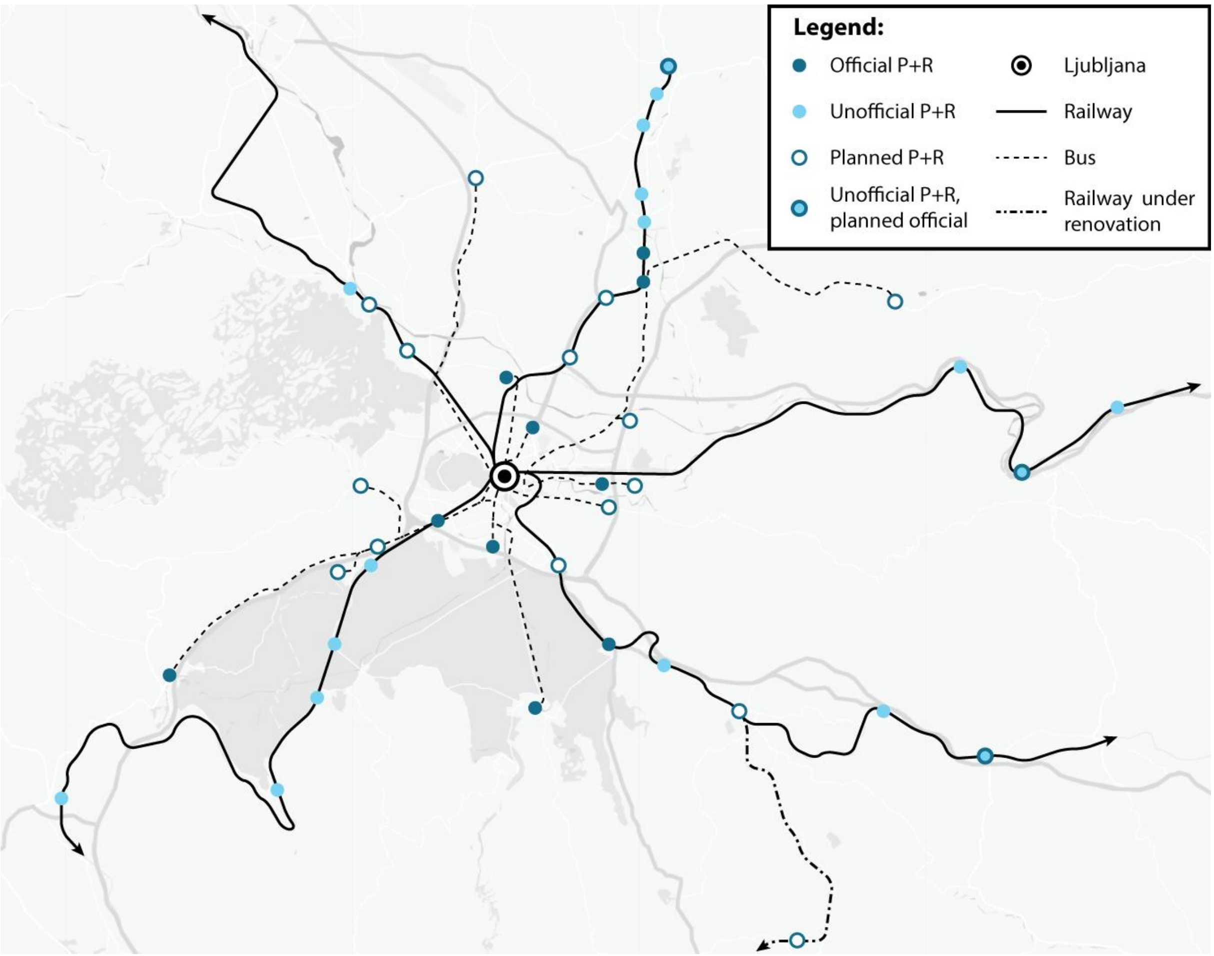


Figure 3.7. Official, unofficial and planned P+R facilities in the CSSR region

3.3 Design of experiment

An essential part of any choice modelling research is obtaining data to analyse. Some use data already collected in prior researches, saving both time and money, but can provide issues, if the way the data was collected in way that is not suitable for the method being investigated. Therefore, data for this research had to be collected. SBWDCEs are a very recent addition to the choice modelling family and the presence of this type of data is not abundant. The wish to apply the method in the field of transport further narrows that down. Choice data therefore needed to be collected for this research specially.

A common dilemma when collecting choice data is also whether to use RP or SP data. Because this research will investigate the sequential best and worst choices within the context of SBWDCE, the choice is irrelevant as this research investigates the behaviour in SP surveys.

Additional requirements for the experiment, based on the estimation technique are presented in subchapter 3.3.1. Subchapter 3.3.2 looks at the individual elements of the survey and how each of them have come together. The final subchapter deals with data collection, first with the pilot survey, followed by the final survey that was undertaken for this research.

3.3.1 Experimental design dimensions

The characteristics a survey needs to meet are usually influenced by the case study being investigated. This research however, uses a specific analysis method, so the survey needs to be tailored to the model: SBWMNL μRRM. For that, the requirements of both μRRM and SBWMNL models need to be outlined.

Attributes should be generic

An RRM model determines regret by comparing the same attributes across different alternatives (Chorus, 2012). Alternative specific attributes have no added value when using regret modelling. The alternatives should therefore include as many generic attributes as possible. The easiest way this can be accommodated is by having unlabelled alternatives (meaning no alternative specific constants) which are different from each other only based on their attribute levels.

Number of alternatives

The choice sets should have at least three alternatives for the RRM model to produce a result different to a RUM model (Chorus, 2016). Because the model will also look at additional choices within the same choice set, the initial number of alternatives should be higher than three. From the perspective of the SBWDCE and looking at the difference between best and worst choices, having the same number of observations of both best and worst decisions is preferred. This leads to the initial choice sets needing to have an odd number of alternatives. For example, a three-alternative choice set would produce one best and one worst choice, a five-alternative choice set would give us two best and two worst choices, etc. For simplicity sake and so as not to overwhelm the respondents, the number of alternatives should also not be too high. Considering all this, the number of alternatives chosen for the survey is five.

Number of attributes

Similarly, to the number of alternatives, the number of attributes should not be too high, so the respondents are not overwhelmed by the amount of data they need to process (chapter 0). At the same time, the number of attributes should also not be too low, so the respondents have enough information to make meaningful comparisons among alternatives. Taking the number of alternatives also into consideration, the number of attributes should be around five. As summarised in chapter 0, different studies have come to different conclusions regarding this issue and five alternatives with five attributes each is a number most of the papers found acceptable.

Number of attribute levels

Like the number of alternatives, the number of attribute levels also needs to be three or higher for RRM to yield a different result from a RUM model (Chorus, 2012). For this reason, dummy variables should be avoided if possible since they only have two levels (0 and 1).

Opt-out / status quo alternatives

Opt-out alternatives are useful because they add to the realism of a stated choice survey, as they do not force respondents to make a choice, if none is attractive enough (Ryan & Skåtun, 2004). This is analogous to choosing not to travel in the transportation sector. However, using opt-outs with RRM models can prove difficult, as these alternatives generally have no attributes, meaning there is no way to compare them to other alternatives with RRM modelling. Opt-out alternatives can also result in difficulties for either the RUM or RRM model, depending on how they are phrased (Hess et al., 2014) and respondents may choose opt-outs just to avoid mental effort (Ryan & Skåtun, 2004), although this is less of an issue in the SBWDCE where respondents still have to implicitly rank all alternatives. The performance of models not containing opt-outs was found to be similar to one containing it (Hess et al., 2014). In the case of a status quo alternative, this was used by Lancsar et al. (2013) and it did contain all the same attributes as other alternatives. Status quo alternatives can be problematic because they are labelled, a characteristic that we wish to avoid, as it requires alternative specific constants, which are not desired.

3.3.2 Survey construction

This subchapter presents the arrangement of the stated preference survey. The first two subchapters focus on selecting the attributes and their corresponding attribute levels for the twelve choice tasks present in the first part of the survey. This is followed by the regret scale in 3.3.2.3. Finally, the socio-demographic data that is asked of the respondents is presented in 3.3.2.4.

3.3.2.1 Attribute selection

Selecting the appropriate attributes in a stated choice survey is of crucial importance. Attributes are what characterises alternatives and the levels of these attributes are what respondents use to make their decisions. A common issue is that the number of attributes describing alternatives is greater than what can be captured in a stated choice survey, without overwhelming the respondents (T. A. Arentze & Molin, 2013). Different people may also consider different attributes when making decisions. Since the number of attributes in this survey was limited to around five (in subchapter 3.3.1) to keep the survey from being too complex, these five attributes need to be chosen with great care to make sure the most common characteristics are chosen.

In the case of park-and-ride parking lots, many attributes describing vastly different characteristics have been found in a literature overview. To get a better understanding and simplify the overview, grouping these attributes is logical. Bos, Heijden, Molin, & Timmermans (2004) proposed a division of attributes into five groups (Table 3.4). Although they also looked at car only trips (driving to the destination), these groups can still be used when only looking at P+R attributes.

Table 3.4. Grouping of park-and-ride attributes (Bos et al., 2004)

Parking	**P+R facilities**	**Connecting PT**	**Time**	**Cost**
General information on the P+R, such as capacity, walking distances, possibility of finding a space	Safety (lighting, supervision) and extra amenities like kiosks, supermarkets and cafes	Characteristics of the public transport serving the P+R	All aspects of time on the journey: in-car, in-PT, walking, waiting, …	Costs associated with using a P+R, including fuel, parking, toll, PT ticket

Park-and-ride facilities are also very often associated with a level of uncertainty and variability. Several studies have looked into traveller behaviour with respect to this. Bos et al. (2004) included the chance of finding a parking space and the time for finding one. The time to find a parking lot was also included by Zhao, Li, & Xia (2017), Chen et al. (2015) and Bekhor & Shiftan (2010), while the probability of finding a free space was investigated by Chen et al. (2015). A very commonly present attribute in literature was also the capacity of a parking lot, which can be seen as a proxy for the possibility of finding a free space (Pang & Khani, 2018) (Bergman, Gliebe, & Strathman, 2011) (Mahmoud, Habib, & Shalaby, 2014), especially when coupled with the time of arrival at the P+R or time of departure from home (Olaru, Smith, Xia, & Lin, 2014) (Chen et al., 2015).

As P+R can be described by so many attributes, a way to investigate all of them would be to carry out multiple experiments, as done by Bos et al. (2004) and T. A. Arentze & Molin (2013). However, since identifying and evaluating travel behaviour is not the main purpose of this research, but rather to investigate people's decision-making with respect to best and worst choices, a single experiment will be undertaken, but in a reduced format, similar to one of the sub-experiments of Bos et al. (2004) or T. A. Arentze & Molin (2013).

The focus is on the quality of public transportation at P+R and how this affects commuter behaviour, so the main three attribute groups are *Connecting PT*, *Time* and *Cost*. Uncertainty-related attributes are also not be used in the survey, further narrowing the scope of possible attributes. From a literature review of eleven papers on P+R research, nine were found that met the stated criteria (Table 3.5).

Table 3.5. Park-and-ride attributes in literature

	P+R access by car (time or distance)	**Costs (individual or grouped)**	**Travel time by public transport**	**Frequency of public transport (waiting time)**	**Mode of public transport**	**Number of transfers**	**Egress time from public transport**	**Quality / Comfort of public transport**	**Crowding on public transport (seat certainty)**
(Bos et al., 2004)	X	X	X	X	X	X		X	X
(Bekhor & Shiftan, 2010)	X	X	X		X		X		
(Bergman et al., 2011)	X		X				X		
(Qin, Guan, & Wu, 2013)	X	X	X	X		X	X	X	X
(T. A. Arentze & Molin, 2013)	X	X	X	X	X				
(Mahmoud et al., 2014)	X	X			X				
(Olaru et al., 2014)	X								
(Chen et al., 2015)		X							
(Zhao et al., 2017)	X	X	X	X		X	X		X
(Sharma, Hickman, & Nassir, 2017)	X	X	X	X	X				
(Pang & Khani, 2018)	X		X	X	X	X	X		
Σ	**10**	**8**	**8**	**6**	**6**	**4**	**5**	**2**	**3**

Looking purely at frequency of occurrences, the first five attributes should be chosen. Indeed, these five were selected, but not solely because of this. Table 3.5 is far from complete as there is a plethora of research papers on the topic of park-and-ride.

P+R access by car was chosen as the only indicator of accessing the parking lot. Similarly, cost is the only financial indicator, so for any valuation calculation a monetary attribute is required. It was decided to be combined for both parking cost and public transport ticket cost as that is how the system is organised in the CSSR, and the respondents are most likely to be familiar with such a system.

The remaining seven attributes all fall in the group of *Connecting PT*. Crowding was excluded as P+R are generally located outside of the city or even farther away, so getting a seat should be the norm in most situations. The quality / comfort of PT is seen as more important in a choice situation between car and PT, rather than two different PT options. Also, the attribute of mode already captures a level of the comfort aspect. Mode was also chosen because it added realism to the choice sets and allowed the respondents to make a clearer picture of different parking lots. Frequency was chosen as a typical PT service attribute, that can relatively easily be changed to attract more users. Transfers and egress time would both have been very valuable attributes to have, but the reasoning they were not chosen is that the model used for market share in the case study could not really account for it. The study area will be analysed by comparing the attractiveness of one P+R to that of others. The destination is assumed to be the city centre, where almost all public transport services go, and commuters can then reach most places within a short walking time. If a full-scale transport model had been used, the information of transfers and egress time would have been of much greater value for modelling commuter behaviour. To summarise, the five attributes chosen for the survey are:

- Car travel time (from home to the P+R)
- Cost (of parking and public transport)
- Public transport travel time
- Public transport frequency
- Public transport mode

3.3.2.2 Attribute level selection

Having determined the attributes in the previous chapter, the next step is determining the attribute levels for each of the attributes. This was done based on current travel times, costs and frequencies in the study area (Central Slovenia Statistical Region). A detailed overview of how the attribute levels were obtained can be seen in Appendix B.

Initially, five levels were planned for all attributes except for PT mode (binary), as Vanniyasingam et al. (2018) determined that high D-efficiency was found in experimental designs where the number of attribute levels was the same as the number of alternatives. This does seem logical and allows for attribute level balance as each level could appear once in the choice set.

However, for several reasons the number of attribute levels was subsequently reduced to three. The first problem was that having a larger number of attribute levels made it more difficult for respondents to evaluate the choice tasks. Despite the number of attributes and alternatives remaining unchanged, having fewer attribute levels means that attribute level overlap is more likely. Overlap is when two alternatives have the same attribute level on one (or more) of the attributes. This makes it slightly easier for the respondents, as they have one fewer attribute to trade-off. It also adds information on other, possibly otherwise less important attributes, which must now be traded-off by the decision-maker (Reed Johnson et al., 2013).

The second difficulty that arose in the experimental design were the rather long combined travel times (both car and public transport). Although realistic on their own in the setting of the CSSR, the combined travel time often resulted in exaggerated travel times. From the analyses of car and public transport travel times, the minimum and maximum travel times were obtained. In reality, a long car travel time is accompanied by a short public transport time (in the case of using a satellite P+R) or vice versa in the case of using a remote P+R facility. In the generated choice experiment however, the combined travel times regularly exceeded travel times found within the region, reaching as high as 90 minutes, if long car access times were often coupled with long public transport travel times. To avoid this issue, when reducing the levels from five to three, car and PT travel time levels were reduced by taking away the two highest levels, rather than the second and fourth as done for price and frequency. The final attribute levels, after being reduced from five to three, are shown in Table 3.6, with the removed levels indicated as red and crossed out

Table 3.6. Final attributes, attribute levels and prior parameter values

Variable	Attribute	Levels	Unit
Car	Car in-vehicle time	5, 15, 25, ~~35~~, ~~45~~	min
PT	Public transport in-vehicle time	10, 20, 30, ~~40~~, ~~50~~	min
Cost	Parking and public transport cost	1, ~~3~~, 5, ~~7~~, 9	€
Headway	Headway of public transport service	5, ~~10~~, 15, ~~20~~, 30	min
Mode	Mode of public transport	bus (0), train (1)	dummy

3.3.2.3 Regret scale

To evaluate how prone the respondents are to experiencing regret, a regret scale (Schwartz et al., 2002) was added after the choice sets were presented to the respondents. The scale is made up of five questions / statements, for which the respondents must decide, to what extent they agree with the statement or to what extent it applies to them. They do so on a 7-point scale from '*completely disagree*' to '*completely agree*'. The five questions of the regret scale are:

1. Once I make a decision, I don't look back.
2. Whenever I make a choice, I'm curious about what would have happened if I had chosen differently
3. Whenever I make a choice, I try to get information about how the other alternatives turned out.
4. If I make a choice and it turns out well, I still feel like something of a failure if I find out that another choice would have turned out better.
5. When I think about how I'm doing in life, I often assess opportunities I have passed up.

The first question is scored in reverse, as the '*Completely agree*' option is the least regretful, while in the other four, it is the other way around. This leads to a minimum score of five and a maximum score of 35, the former representing a respondent experiencing hardly any regret while the latter would mean that the respondent experiences a lot of regret when making decisions. The result is used as a socio-demographic variable to determine if the level of experiencing regret has an impact on the model. It may prove very useful in the class allocation function within the LC model.

3.3.2.4 Socio-demographics and travel behaviour

The personal characteristics of respondents were the final part of the survey. Although they cannot be varied in a DCE setting in the same way as attributes, they can be added in the modelling step, for example as interaction effects. Both socio-demographic data and current travel behaviour data of respondents was collected. As with the regret scale, this adds information on the individual's behaviour and can be particularly useful for the class allocation function in the LC models. The same socio-demographic data was collected for all respondents:

- Sex
- Age
- Education level
- Employment status
- Household income
- Household car ownership
- Household size
- Possession of driving licence
- Municipality of residence
- Region of employment / education

Based on the municipality of residence, travel behaviour data relevant for a specific group of respondents was collected as follows:

- *Residents in the city of Ljubljana were asked:*
    - Which mode do you use most when travelling in Ljubljana?
- *Residents residing outside the city of Ljubljana were asked:*
    - Have you ever used a P+R for your travel to Ljubljana?
    - How often do you travel to Ljubljana?
    - Which mode do you use most when travelling to Ljubljana? Data collection

Once the choice modelling survey elements have been selected, a design was constructed, tested and implemented. Ahead of making the final survey, a pilot survey was undertaken as explained in 3.3.2.5. Based on these responses, several designs were generated, using different methodologies: D-efficient designs and Bayesian designs (with small and large variances). They were then evaluated, and the best performing design was selected for the survey.

3.3.2.5 Pilot survey

The pilot survey was carried out ahead of the final survey with the goal of testing it and solving any potential issues. The survey was filled in by 13 respondents, eight females and five males. The average age of the respondents was 45 years old, with a standard deviation of 16 years. The minimal age was 16 and the maximum was 60 years old.

Respondents were given the full introduction text (Appendix C) with instructions and context description and asked if everything was understandable or if anything was left vague or unclear. They were also given the option to comment on the text. When presenting the choice tasks and attributes, they were also asked to share if they think of any additional characteristics that they feel should be included.

After the introductory text, the respondents were shown the same choice task in three different layouts (found in Appendix 0). After solving all three, they were asked which they found the best / most clear / least difficult to process. Layouts 1 and 2 both featured the choice set displayed in a table, with each column representing an alternative and each row an attribute. The difference between the layouts was that in Layout 1, all four questions (best, worst of the remaining four, best of the remaining three, worst of the remaining two) were asked at the same time. The advantage of this was that it made the survey feel shorter, as reported by the respondents. The main issue however, was that there was no possibility, within the software, to make sure that each alternative was only chosen once. A respondent could have, even by accident, have selected, for example Alternative 1 for all four questions.

Layout 2 sought to fix this problem, by asking only one question at a time. Once the question was answered, the next page contained only the answers not yet chosen. This way, it was made sure that a different alternative was picked every time.

Layout 3 had a slightly different presentation of the choice set, with each alternative having its own image. Like Layout 2, one question at a time was asked and the non-chosen alternatives were repeated in the follow-up question. The main difference from Layout 2 was the presentation, as the previously chosen alternative was no longer present. This was the preferred layout from the researcher's point of view, as it is most similar to how the choice sets are modelled: only considering the remaining alternatives in the choice probability calculation.

Layouts 1 and 3 scored relatively equally (39% and 46% respectively), while Layout 2 was lowest. In addition, it should be noted that all respondents who chose Layout 2 completed the survey on a smart phone, rather than a computer. The smart phone view distorted the layouts quite significantly, so a decision was taken to only make the survey available for a desktop layout. The respondents choosing Layout 2 also stated that their second choice would then be Layout 3.

Finally, at the end of the pilot survey, respondents were given an option to write any additional comments they may have regarding the survey in general or anything in particular. Some grammatical changes and different wordings were suggested. One added that they would only be interested in the combined travel time and not separated into car and public transport travel times. A few commented on the necessity of possessing a smart card for using the P+R. The most commonly brought up issue was the rather confusing nature of selecting best, then worst, then best and finally worst. In most cases, the respondents suggested just asking for the best alternative all the time, others suggested an explicit ranking solution. As the research is looking at the differences between best and worst choices, these suggestions could not be met.

3.3.2.6 Final survey

Generating experimental designs was done with the help of Ngene software (Choice Metrics, 2018). In line with recent trends, an efficient design was to be used for the survey. As Walker et al. (2018) pointed out, a D-efficient design is not necessarily the best way to go, so a Bayesian design was also considered. In the end three different designs were generated: D-efficient, Bayesian will small standard deviations and Bayesian with large standard deviations.

Constructing efficient designs requires prior parameter values to calculate choice probabilities and make more efficient choice tasks, where differences between structural utilities are small. The prior values were obtained from a detailed study on park-and-ride facility choice attributes done by Bos et al. (2004). As their survey took place in the Netherlands, there is less certainty in the accuracy of the prior values, since the survey for this research will be carried out in Slovenia. Both countries are in the European Union, so the differences are not expected to be drastic. Nevertheless, this reinforces the notion of using a Bayesian design rather than a D-efficient, as outlined by Walker et al. (2018). The exact procedure of how the priors were obtained is shown in Appendix E. The final values can be seen in Table 3.7.

Apart from the prior values of parameters, Bayesian designs also require standard deviations (SD) of each parameter value. For this research two different Bayesian designs are constructed: one with smaller and one with larger standard deviations. These standard deviations were calculated based on the prior value and its distance from zero, as passing that value would mean a change in sign (from – to + or vice versa). While the exact values of parameters are uncertain, their signs are typically expected. For example, the travel time and cost parameters are usually negative (longer travel times and higher travel costs would naturally seem to cause disutility). The difference between small and large deviations for two different Bayesian designs was distinguished by how many SDs away from zero the prior is. The smaller deviations were therefore determined as 1/3 of the distance between the prior and zero, with the larger deviations being 1/2 the distance. Considering a normal distribution, prior values with the smaller SDs are 99,9% certain to have the correct sign or while larger SD have are 97,7% certain. An overview of the prior values and standard deviations is given in Table 3.7.

Table 3.7. Prior values of parameters and standard errors for the experimental designs

	D-efficient	**Bayesian with small SD**		**Bayesian with large SD**	
	prior	**prior**	**SD**	**prior**	**SD**
Car	-0,0490	-0,0490	0,0163	-0,0490	0,0245
PT	-0,0535	-0,0535	0,0178	-0,0535	0,0268
Cost	-0,2575	-0,2575	0,0858	-0,2575	0,1288
Headway	-0,0990	-0,0990	0,0330	-0,0990	0,0495
Mode	0,1700	0,1700	0,0567	0,1700	0,0850

The three different designs were constructed in Ngene (syntaxes presented in Appendix F). The base utility function used in all three syntaxes is presented in Equation 3.9.

Equation 3.9. Utility function for constructing the choice experiment

$$U_i = \beta_{Car} \cdot Car_i + \beta_{PT} \cdot PT_i + \beta_{Cost} \cdot Cost_i + \beta_H \cdot H_i + \beta_{Mode} \cdot Mode_i$$

where:	i	*alternative*
	m	*attribute*
	x_{im}	*attribute levels of attribute m in alternative i*

Eight designs of each design type were generated. At the time, the software did not yet allow for generating designs that are both RUM and P-RRM efficient, as was implemented by van Cranenburgh & Collins (2019), so the designs were generated as RUM efficient (all three design types) and evaluated post-hoc on their D-error for P-RRM according to Equation 2.14 and the composite D-error which incorporates both decision rules. The results are presented in Table 3.8. In line with expectations, D-efficient designs perform best on the RUM D-error. Interestingly, they also performed best on the P-RRM D-error and hence on the composite D-error. However, as was stated previously in this chapter, D-efficient designs are not as reliable if the true parameter value differs from the prior parameter value. Comparing the Bayesian designs, those with small deviations perform better than ones with larger deviations for the same reason that D-efficient designs perform better than Bayesian designs. It was decided that the most efficient Bayesian design would be selected for the survey. In this case, this was design 9, a Bayesian design with small standard deviations. The full design can be seen in Appendix G.

Table 3.8. D-efficiency of generated experimental designs sorted by the composite D-error from most to least efficient

Design ID	Design type	RUM D-error	P-RRM D-error	Composite D-error
3	D-efficient	0,005621	0,001563	0,003592
1	D-efficient	0,005649	0,001596	0,003623
9	Small Bayesian	0,005735	0,001601	0,003668
2	D-efficient	0,005671	0,001688	0,003680
4	D-efficient	0,005718	0,001691	0,003705
10	Small Bayesian	0,005768	0,001744	0,003756
5	D-efficient	0,005761	0,001794	0,003778
22	Large Bayesian	0,005891	0,001816	0,003854
18	Large Bayesian	0,005802	0,002126	0,003964
8	D-efficient	0,006035	0,002055	0,004045
11	Small Bayesian	0,006214	0,001882	0,004048
14	Small Bayesian	0,006173	0,001949	0,004061
17	Large Bayesian	0,005869	0,002307	0,004088
7	D-efficient	0,006197	0,002183	0,004190
23	Large Bayesian	0,006554	0,00197	0,004262
15	Small Bayesian	0,006424	0,002169	0,004297
24	Large Bayesian	0,006728	0,002301	0,004515
12	Small Bayesian	0,007031	0,002161	0,004596
13	Small Bayesian	0,007342	0,00221	0,004776
20	Large Bayesian	0,007158	0,002506	0,004832
16	Small Bayesian	0,007948	0,002741	0,005345
19	Large Bayesian	0,008832	0,003502	0,006167
6	D-efficient	0,009369	0,00334	0,006355
21	Large Bayesian	0,00968	0,004065	0,006873

3.4 Data collection

Once the survey was designed, respondents were needed to fill in the survey and provide observations for the stated choice experiment. Obtaining data from a large number of people with different socio-demographic characteristics is key for robust model outcomes and significant results in traditional DCEs. As this research is applying the SBWDCE methodology, the limited time and budget of the research should not pose an issue. The survey was spread in the online environment through different channels. A snowballing method on Facebook was used throughout the 41 days of the survey being open, with reminders. The survey was spread among employees by several employers in the municipality of Ljubljana. Several municipalities in the Central Slovenia Statistical Region were also contacted if they could help spread the survey link through their PR channels.

The survey was opened on 3.6.2018 and stayed active until the 13.7.2018. During the 41-day period, 108 complete responses were recorded and a further 138 partial responses, resulting in a 43,9% completion rate. The survey was considered complete if all 48 choice tasks were accomplished. Some respondents answered the choice situations but then skipped some socio-demographic questions. Such replies are still considered complete, as choice situations were the focal point of the survey.

3.5 Socio-demographic characteristics

This subchapter investigates the socio-demographic data of the respondents and compares the sample characteristics with those of the population. The socio-demographic questions were asked in the second half of the survey, following the stated choice questions. As the questions were not compulsory, not all respondents answered them, although the vast majority did.

When comparing the sample with the population, the population needs to be defined to determine its characteristics. As the survey topic was daily commuters using P+R facilities in the CSSR, the population is made up of working adults and those in education who have a driver's licence and commute to Ljubljana on an almost daily basis. A different definition could also see the population as the working population and those in education in general. Both definitions exclude those without a driver's licence and those who do not commute daily to a bigger city (unemployed, pensioners …). Socio-demographics for either of the populations could not be determined and the closest population for which data was available is the regional population of the CSSR, regardless of their working status, age or possession of a driver's licence.

The socio-demographic questions started with five statements of the regret scale, completed by almost all respondents (106 out of 108). The distribution of the results can be seen in Figure 3.8 with the summary statistics shown in Table 3.9. The summary statistics in the table show that the data is left skewed. Both the median and mode are right at the centre point of the regret scale (at 20) that rangers from five to 35, indicating that on average the respondents were not very prone to experiencing either a lot of or minimal regret. This also indicates that the sample as a whole was distributed somewhat normally across the scale. The largest bin in Figure 3.8 is 20-25, meaning the respondents experience slightly above average regret levels. The skewness indicates that although the largest group experience slightly above average regret, very few experience a lot of regret, while groups with lower levels of regret experience, the difference in size is smaller. The mean value also shows this, as it is slightly below the middle point of 20, at 18,86. The sample is therefore slightly less prone to experiencing regret but does resemble a normal distribution across the scale.

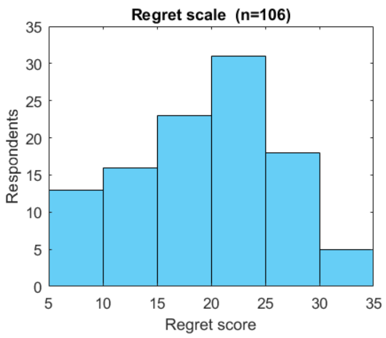


Figure 3.8. Histogram of the regret scores

Table 3.9. Summary statistics of the regret scores

Mean	18,86
Median	20
Mode	20

A slightly larger share of respondents completing the survey were female, about 58%. Compared with the population, this is above average, as the population share of women is around 51% (Statistični Urad Republike Slovenije, n.d.-b). In terms of age, the lowest was 16 and the highest 60. Most likely due to the nature of the survey (online platform) and the diffusion method, older respondents could not be gathered for the survey. The overview of summary statistics is shown in Table 3.10 with the age histogram in Figure 3.9. The biggest age group was 25-to-35-year olds if the data is grouped into ten-year bins. Using 5-year bins, the biggest is 20-25-year olds. Compared to the CSSR population, younger respondents were overrepresented, as in the population the 35-40-year-old group is the biggest, with the groups of those below 30 years old being less populous than those between 40 and 60 (Statistični Urad Republike Slovenije, n.d.-b). The sample did not have any respondent over the age of 60 or under 16, meaning a large unrepresentativeness of youths and the elderly, with respect to the CSSR population. But if considering the working population, this may be somewhat representative, although the age distribution still cannot be considered as representative.

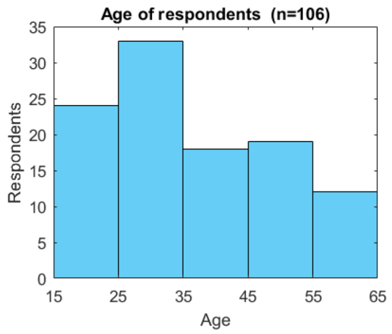


Figure 3.9. Histogram of the regret scores

Table 3.10. Summary statistics of respondents' age

Mean	36,22
Median	32,5
Minimum	16
Maximum	60

The respondents were highly educated, with the most having a university education, either on the bachelor (29%), master (36%) or even PhD level (9%). In third place was the group with a finished high school (20%), but this could also be because of the relatively young respondent base, for which that is their highest completed education level, even though they may still be in education. Respondents with a lower education level were in the minority (7%). The results are depicted in the top chart of Figure 3.10. Compared to the population, the sample vastly overrepresented the highly educated and underrepresented those with a lower education. In the CSSR, half of the population have a finished high school, while higher education is completed by 30%, compared to almost three quarters in the sample (Statistični Urad Republike Slovenije, n.d.-b). However, the sample may again be closer to a commuting population, as higher educated workers are often more likely to commute to cities, whereas lower skilled workers in factories often work in industrial zones outside of cities.

In terms of employment status (middle chart in Figure 3.10), almost all respondents were either full-time employed (67%) or students (23%). The latter proves the statement from the previous paragraph of many respondents with a high-school education. The high number of students also falls in line with the high number of younger respondents. Few were unemployed (6%) or working from home (none). The share of those employed full-time or in education is again overrepresented (49,8% and 11,5% respectively), but if pensioners are omitted from this data, representativeness becomes better, those in the population being employed full-time reaching 69%, those in education 16% and unemployed at 8% (Statistični Urad Republike Slovenije, n.d.-b).

Household income is generally a more sensitive topic, so it is understandable that many did not wish to disclose theirs. From the ones that did, most seem to have an income between 1.000 € and 3.000 € per month. The chart at the bottom of Figure 3.10 does resemble a slightly right-skewed distribution. Data on income is very scarcely available and the only widely available metric is the average net income, which was 1.167€ at the start of year 2018 (Statistični Urad Republike Slovenije, n.d.-b).

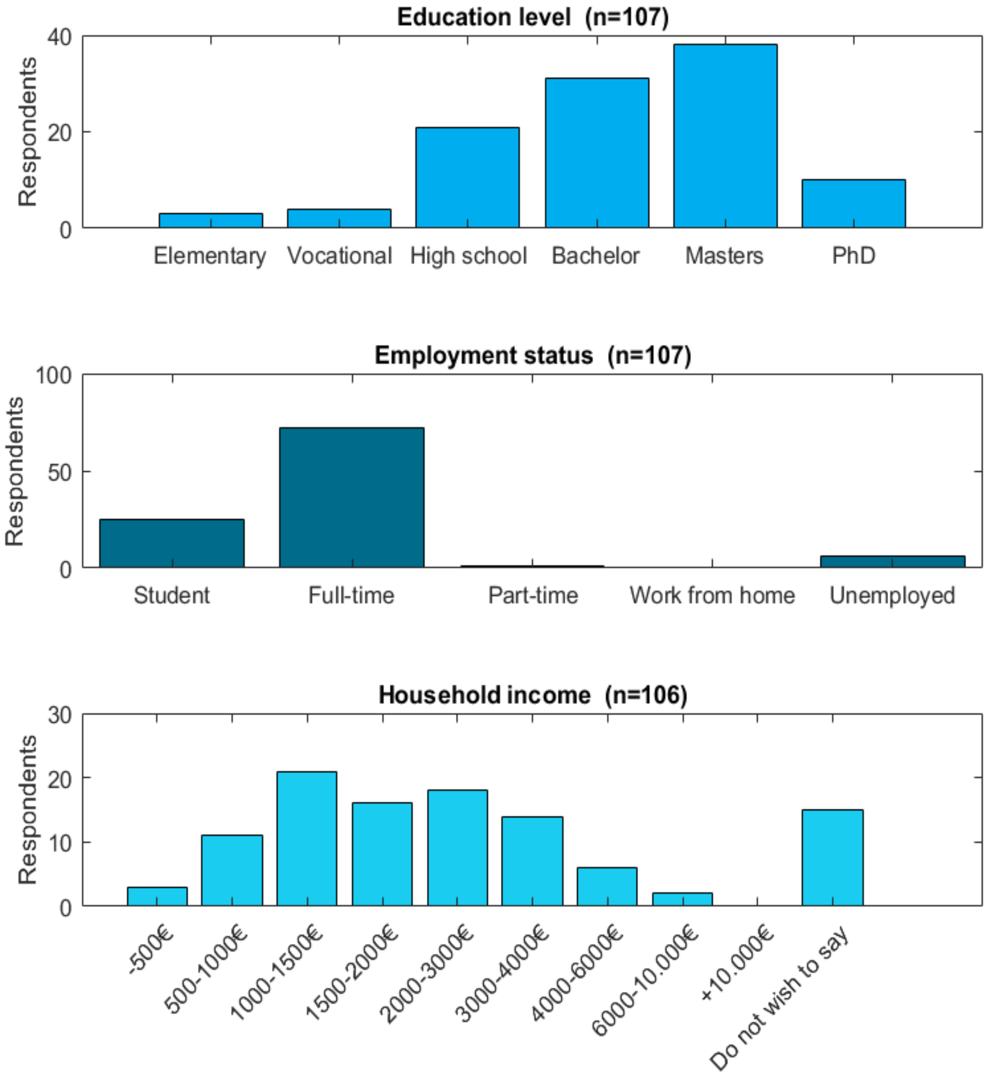


Figure 3.10. Education (top), employment status (middle) and household income (bottom) of respondents

Moving towards travel possibilities and household size (Figure 3.11), the vast majority of respondents have a driver's licence and the same can be said about households having at least one vehicle. The majority have either one or two, with the share of three or more being only slightly above households with no cars. In other words, there were an average 1,5 cars per household in the sample with a motorisation level of 551 cars per 1000 people. This also proves to be fairly representative (slightly above average) of the population, with households in the CSSR having 1,37 cars on average and the motorisation level being 538 cars per 1000 inhabitants (Statistični Urad Republike Slovenije, n.d.-b).

Considering the household size, most are occupied by two people (33%), followed by four (24%), three (21%) and then one (16%). Very few respondents were from a household with more than four people (7%). While the representativeness of larger households is close to the population (9%), single-occupant households are the most common in the CSSR (35%), followed by two, then three people households (Statistični Urad Republike Slovenije, n.d.-b).

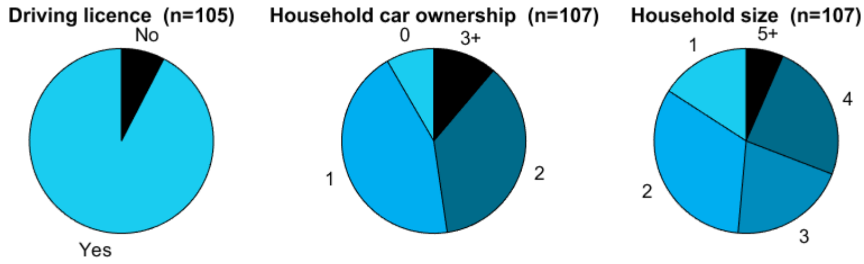


Figure 3.11. Driving licence possession (left), car ownership (middle) and household size (right) of the respondents

Finally, the travel behaviour of the respondents is analysed. Due to the nature of these questions, the respondents are divided into two groups, if they live in the city of Ljubljana or outside of it. 61% live within the city boundaries, with another 23% residing in the Central Slovenia Statistical Region. The remaining few live in other regions of the country. In terms of their working locations, 83% work in the city of Ljubljana, with the rest being split evenly between the CSSR and other regions.

Of those residing outside of the city 42% have used a P+R facility at least once. This number is relatively high, and it suggests that many people try it but do not practice it on a regular basis (only 5%). Considering the modes used for the commute to Ljubljana, almost two thirds travel by car, with another 15% as car passengers. 10% use the city bus, as some lines extend from the city into the region. Comparing this travel behaviour to that of those living in the city, the share of car trips is halved to only 34%. The second mode, standing at 32% is bike, which is a mode that is becoming increasingly popular in the city in recent years. A fair bit of respondents also travel by bus (23%), with other modes being utilised by only a few people. The comparison of travel behaviour modes can be seen in Figure 3.12.

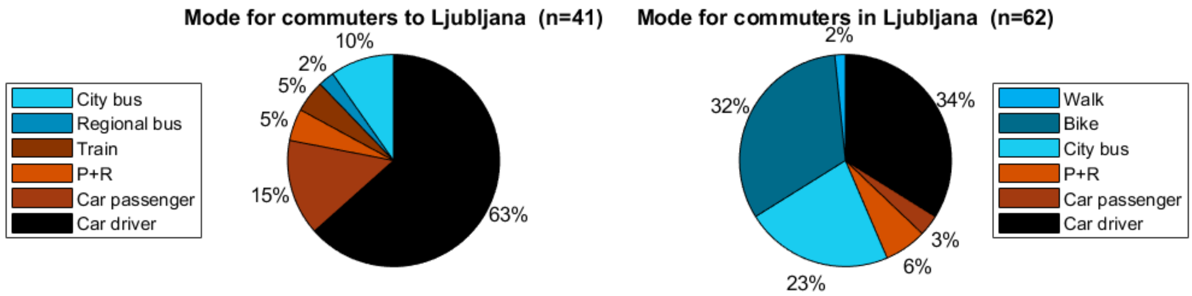


Figure 3.12. Travel behaviour of respondents living outside Ljubljana (left) and of those living in the city (right)

From the socio-demographics of the survey sample and the population of the CSSR, we cannot conclude that the sample accurately represents the population. Although some metrics were fairly representative (car ownership in particular), others were not. In most cases, a qualitative analysis of what a possible population of working adults in the city would be like and comparing it to the sample resulted in a better understanding of the unrepresentativeness of the sample compared to the CSSR population. However, as there is no data to back-up this qualitative analysis, no conclusion on its accuracy can be drawn.

3.6 Conclusion

In the methodology chapter, an outline of how to answer the research question was given. The chapter started by outlining how the model will be constructed to obtain estimation results from the data. The different models that will be compared in this research were outlined and the test used to compare them were elaborated. The result of these models and tests will give an answer to the research questions.

For obtaining the model estimates, the case study and experimental design needed to be selected. The model is based on respondents selecting P+R facilities, described by five attributes. The selected attributes were travel time by car, travel time by public transport, total cost, the frequency of public transport and the mode of public transport. In total, five alternatives will make up each of the twelve choice sets. They were generated using a Bayesian design and the design with the lowest composite (RUM and P-RRM) D-error was selected for the survey. The pilot survey tested three different layouts for the discrete choice questions and assessed if the instructions are clear and understandable. Based on the feedback of the respondents, the necessary changes were implemented into the final survey.

The survey was carried out fully online through various distribution channels. A total of 108 responses were obtained in the 41 days the survey was available online. Unfortunately, the socio-demographics are not representative of the population of the CSSR, however it should be pointed out that it is in fact not the true population that the sample was based on. The true population are daily commuters who drive into the city of Ljubljana, mostly for work or education. The socio-demographic characteristics of the CSSR were used as the best approximation of a population for which the data was available.

The obtained results will also be used in a case study to determine the attractiveness of individual P+R lots by varying the public transport LOS. The focus will be on improving attractiveness of link-and-ride parking lots, which provide more sustainable mobility than traditional P+R lots.


PETROL
11 ZALOG
LJ LPP 221

4 Monte Carlo experiment

An important step before collecting data and estimating the models is to make sure that the models can accurately obtain the parameter estimates from the sample. To test this, synthetic observations were generated with the help of Monte Carlo simulation. On the generated dataset, several developed models were tested for their predictive ability. The chapter starts with an outline of how the data was generated, what were some of the hesitations in the data generation process and the number of respondents to be simulated. The second subchapter then reports the estimation results of different models on the simulated data. The estimated parameter values are evaluated, and the consistency and unbiasedness of the results is tested, along with the information of how many respondents may be needed to obtain an accurate and reliable result in the empirically gathered data.

4.1 Choice data simulation

As with any discrete choice experiment, before a model can be estimated, the choice data needs to be obtained / gathered. For generating stated preference data with Monte Carlo simulation, the structural part of the utility (or regret) is calculated using prior parameter values in the same way as later in the estimation phase. The main component of Monte Carlo simulated choice data is the error term of each alternative. This error term is what is used to mimic the uncertainty and variability in the choices people make. The error is assumed to be extreme value (EV) Type I distributed (Chorus, 2016).

For generating individual choices, the structural utility (regret) must be calculated for every choice situation and then a random draw from the EV distribution function needs to be added to the total utility (regret) of each alternative. The alternatives are then evaluated based on their utilities (regrets) and the one with the highest utility (lowest regret) is chosen as the best alternative and vice-versa for the worst alternative (van Cranenburgh, 2018). As each choice set is evaluated four times (each time with the previously chosen alternative removed), the systematic regret needs to be computed every time, as not the same alternative is always chosen, due to the inclusion of the draw from the EV distribution. The systematic regrets were therefore computed anew for each new simulated respondent. The Monte Carlo simulation was performed using Matlab software and the results of the estimation are presented in chapter 4.2. The following two subchapters provide more detailed information on the way the error term is drawn for the total regret and how many respondents are simulated.

4.1.1 Making draws for the error component

A crucial part of Monte Carlo simulations for discrete choice modelling is generating the error term and adding it to the calculated structural utility (regret). The error term allows for variation in observations, mimicking realistic discrete choice data. In most choice models, each choice set is only evaluated once, with the first-best alternative being selected. The inclusion of the error term in such an experiment is straightforward, by making a new draw from the EV for every choice set. In this study however, alternatives are evaluated up to four times (if not chosen in any of the previous choice sets) and thus the question arises, how to include the error term? In other words, is the error term alternative specific (meaning that for the same respondent, it is always the same for the same alternative) or is it choice set specific (meaning that a new draw from the EV distribution is taken for every choice set, regardless of the present alternatives)? This question was dealt with by Hess & Rose (2009), who looked at how choice models account for inter-person variation (between different people) and for intra-person variation (between the choices the same person makes).

As this section does not deal with behavioural aspects of the choice-making process, but merely testing the developed SBWMNL μRRM model, the data generation process should mirror the estimation process. Based on this, we can conclude that for each evaluation a new error term needs to be drawn from the EV distribution, as choice sets are modelled as independent, with alternatives' regrets being based on other present alternatives in the choice set.

4.1.2 Number of simulated respondents

The model estimation will be performed on a simulated dataset of 100 respondents. This dataset will then be used to estimate the results using a variety of SBWMNL μRRM models with different numbers of scale parameters and choice set size constants.

Data for LC models will be generated by combining two generated subsets of 50 respondents each. The observation subsets will be generated with different parameter values to include heterogeneity in the responses. The taste parameters will remain the same in both subsets, but the scale parameters will be switched, as shown in Table 4.1.

Table 4.1. Scale parameter values used for generating heterogeneous data

	Subset 1	**Subset 2**
μBest	10	0.1
μWorst	0.1	10

On top of this, to determine if the estimation process is consistent and unbiased (which are revealed through increasing the number of respondents), the simulation will be carried out for different numbers of respondents. This also provides insight into how many respondents are needed to answer the choice situations for a reliable result. A total of 33 groups of respondent sizes were used, ranging from one to 200. As each respondent answers 12 choice sets with four choices each, this yields 48 observations per respondents, so the number of observations will range between 48 and 9.600. The exact numbers of respondents used in the simulation are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200.

4.2 Model estimation results

In this subchapter, the observations generated using MC simulation were estimated with five different single class models (E.4-E.8) and one LC model (LC.0) to see whether the models do indeed produce the same outcomes as the input values to generate the observations. The first subchapter analyses the performance of single class models, with the LC model being investigated in subchapter 4.2.2. Finally, subchapter 4.2.3 examines if the models are consistent and unbiased with respect to a growing number of observations between one respondent and 200 respondents or between 48 and 9.600 observations.

4.2.1 Single class models

Five models developed in this research were used to analyse the Monte Carlo simulated dataset. All five models yielded highly significant results, with similar levels of model fits. The model fit of all models is presented in Table 4.2 and the model estimates of all the parameters can be seen in Table 4.3. Some of the models did not immediately converge, so the scale parameter for the first-best choice was fixed to ten, to avoid the model looking for a higher value despite the fact that a higher value gives no additional information on the decision rule: a value of ten or higher indicates fully compensatory behaviour. If the model still did not converge, the choice set size constant of the first-best choice was also fixed as it is rooted in the scale parameter and thus is susceptible to the same problem.

Regarding the model fits (Table 4.2), all five models produced a roughly similar result, with E.4 having a lower fit, as it does not model the data in the same way it was generated. E.4 assumes a single decision rule across choices, while in fact different decision rules were used for best and worst choices. Of the other four models, the model fit is almost identical.

Table 4.2. Model fit outcomes from MC simulated dataset

	E.4	E.5	E.6	E.7	E.8
Null LL	-5744,99	-5744,99	-5744,99	-5744,99	-5744,99
Final LL	-3518,97	-3393,57	-3393,57	-3443,78	-3443,78
Rho-squared	0,3875	0,4093	0,4093	0,4006	0,4006
Param	9	10	12	8	10

The parameter estimates (Table 4.3) are also very similar to the true parameter values used to construct the survey and generate observations with MC simulation. Table 4.4 shows the p-values for the significance of the difference between the true value and the estimate. None of the estimates is significantly different from the true value at the 5% or even the 10% level. The most significant difference can be observed for the mode attribute in models E.7 and E.8, with a p-value of 0,3.

Table 4.3. Model estimation outcomes

	True	E.4	E.5	E.6	E.7	E.8
Car	-0,0490	-0,0486	-0,0490	-0,0490	-0,0486	-0,0486
PT	-0,0535	-0,0527	-0,0530	-0,0530	-0,0518	-0,0518
Cost	-0,2575	-0,2617	-0,2582	-0,2582	-0,2555	-0,2555
Headway	-0,0990	-0,1014	-0,0973	-0,0973	-0,0975	-0,0975
Mode	0,1700	0,1895	0,1793	0,1793	0,1435	0,1435
μ		1,37				
μBest	10		*10,00		*10,00	
μWorst	0,1		0,18		0,21	
μBest 1				*10,00		*10,00
μWorst 1				0,18		0,21
μBest 2				*10,00		*10,00
μWorst 2				*1,00		*1,00
Λ					3,85	3,85
Λ 5		*0,99	*0,99	*0,99		
Λ 4		0,95	1,21	1,21		
Λ 3		0,97	1,01	1,01		

** fixed parameters to achieve model convergence*

Table 4.4. P-values for the difference between model outcomes and true values used for generating data

	E.4	E.5	E.6	E.7	E.8
Car	0,84	1,00	1,00	0,93	0,93
PT	0,69	0,82	0,82	0,72	0,72
Cost	0,63	0,94	0,94	0,93	0,93
Headway	0,47	0,60	0,60	0,86	0,86
Mode	0,42	0,68	0,68	0,30	0,30

4.2.2 LC models

In contrary to the results of single class models, the LC model did not predict the input parameters to the same level of accuracy. The presented model is the third iteration of estimating the model, as the previous models did not converge, so two parameters were fixed: the μWorst and Λ 4, both for the second class. This was done for the same reason as in the previous subchapter, as the scale parameter was initially estimated at over 50 but is in terms of estimation the same as ten and thus does not converge. The choice set size constant was subsequently fixed as well because it is rooted in the scale parameter value and is this prone to the same issue.

The model outcome (Table 4.5) was similar to the single class models with the rho-squared at 0,39. The classes are roughly the same size, matching the input of them being exactly the same size, but due to the addition of variation a slight discrepancy is understandable.

Table 4.5. Model fit of the LC model

Null LL	-5744,99
Final LL	-3489,32
Rho-squared	0,3926
Param	21

Table 4.6. Class sizes in the LC model

Class 1	55%
Class 2	45%

More surprising were the attribute estimates (Table 4.7), which for Class 2 seems to be significantly different from the true values used in the data generation. Table 4.8 shows that all but the mode parameter are significantly different at a 95% level. Estimates of the first class on the other hand cannot be seen as significantly different with all having very high p-values.

Considering the scale parameters, μBest in Class 1 (Table 4.7) has a surprising value of only 3,38. This is already a relatively strong compensatory behaviour, but still indicates some level of semi-compensatory behaviour. The other two non-fixed scale parameters, although with relatively high p-values, cannot be seen as significantly different from the true value.

The difference between the classes – significantly different attribute parameters in Class 2 and scale parameter μBest in Class 1 – may be a consequence of the choice set size constant. All three constants in Class 1 are lower than in Class 2, influencing the scale parameter values and possibly also the attribute parameters. Comparing the parameter ratios (Table 4.9), the significantly different estimates of Class 2 produce very similar ratios to both the true values and those of Class 1.

LC models do not produce the true values as accurately as the single class model, but when interpreting the scale parameters, looking into the parameter ratios and considering a possible influence of choice set size constants on both, do give a result close to what the input values were.

Table 4.7. Parameter estimates of the LC model

	Class 1			**Class 2**		
	Est	SE	p-val	Est	SE	p-val
Car	-0,0510	0,01	0,00	-0,0385	0,00	0,00
PT	-0,0531	0,01	0,00	-0,0456	0,00	0,00
Cost	-0,2562	0,03	0,00	-0,2117	0,01	0,00
Headway	-0,0992	0,01	0,00	-0,0826	0,01	0,00
Mode	0,2151	0,04	0,00	0,1299	0,03	0,00
μBest	3,38	1,38	0,01	0,04	0,05	0,41
μWorst	0,11	0,09	0,21	10	---fixed---	
Λ 5	0,98	0,13	[1] 0,88	1,29	0,12	[1] 0,01
Λ 4	1,18	0,16	[1] 0,26	1,37	---fixed---	
Λ 3	1,08	0,15	[1] 0,61	1,35	0,13	[1] 0,01
δ_2				-0,20	0,25	0,42

[1] the p-value is calculated for the significance of difference from the value of 1

Table 4.8. P-values for the difference between model outcomes and true values used for generating data

	Class 1	Class 2
Car	0,76	0,00
PT	0,95	0,01
Cost	0,97	0,00
Headway	0,99	0,00
Mode	0,30	0,15
μBest	0,00	0,29
μWorst	0,87	/

Table 4.9. Parameter ratios of the true parameter values and both estimated classes in the LC model

	True	Class 1	Class 2
Car travel time value	11,42	11,94	10,93
PT travel time value	12,47	12,43	12,93
Headway value	23,07	23,23	23,42
Mode value	0,66	0,84	0,61

4.2.3 Consistent and unbiased

To test for consistency and unbiasedness of the model, datasets with different numbers of respondents were also generated and solved with the E.5 model. This model was selected as it was the initial SBWMNL μRRM model to be evaluated. The remaining four models estimated in this research (E.4, E.6, E.7 and E.8) evolved from model E.5 with the goal of investigating additional aspects of the respondent behaviour and/or model characteristics (varying the number of scale parameters and/or choice set size constants). It was also selected because it is the model used in the estimation of models with fixed scale parameters (I.1-I.4 models).

The expectation was that the estimation results should get consistent and unbiased towards higher numbers of respondents. Consistent in this case refers to estimates being closer to the true value with higher respondent numbers, while unbiased means that the true value is approached form both sides (positive and negative). The resulting graphs for all five parameters can be seen in Figure 4.1. Apart from the individual estimates, the charts also show the true parameter value and curves approaching the true parameter value from both sides, indicating how the estimates should improve towards higher numbers of respondents. Considering this, it can be said that all five estimates are unbiased and consistent as with the increasing number of respondents, the estimates get closer and closer to the true value parameter value. The largest discrepancies can be observed for the mode parameter, which is most often located outside the desired region. Given the results presented in Table 4.4, it would seem that the mode parameter differs most from the true values used in the data generation step.

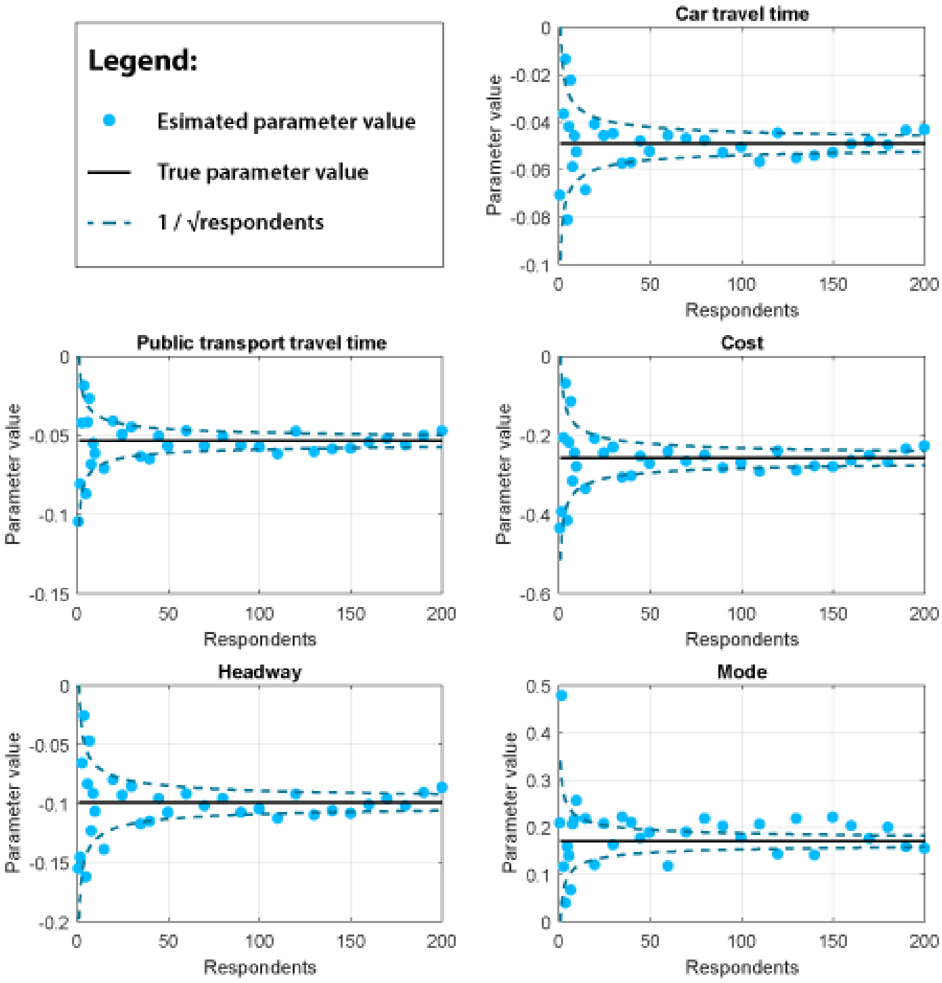


Figure 4.1. Estimation results of all five parameters for different sample sizes

4.3 Conclusion

The model outcomes show that the developed models can accurately reproduce the input values. Single class models in particular were good in this, with all attribute parameters being insignificantly different from the true values. In the LC model, the values of Class 2 were significantly different from the true values, possibly pointing towards an issue in obtaining the same information back from the data, but when considering parameter ratios, the same results emerged as in the true values.

Scale parameters were in most cases very close to the inputs, especially considering their interpretation of what a certain value represents. Some scale parameters reached very high values (upward of 50), which is fundamentally the same as the input value of ten. Some discrepancies were observed in the LC model, but given the values of the choice set size constants and how they differed between classes, they can also provide an explanation for the difference between inputs and outputs.

With respect to consistency and unbiasedness, the results show that having upwards of 50 respondents already yields a relatively accurate overall result. It must be noted however, that this data is simulated, meaning it is almost perfect from the choice modelling perspective. It does provide valuable information that the data can be estimated and that the model behaves in line with expectations and can with confidence be applied on empirical data that will be gathered in the following step of the research.


6
CRNUCE

5 Empirical results

To answer the posed research questions and confirm or deny the stated hypothesis, the gathered observations were estimated with the help of several different SBWMNL μRRM models, as well as latent class SBWMNL μRRM models. The estimation results were compared to results of already established models, namely the first-choice RUM, first-choice μRRM and the SBWMNL RUM model.

In the first subchapter, the models were used by implying a decision rule (fixing the value of the scale parameter) and evaluating the outcomes based on model fit. The models are also compared to each other with the Ben-Akiva and Swait test for non-nested models. Subchapter 5.2 presents an extension of the previous subchapter by allowing the value of μ to vary freely and to converge on its own, without being implied upfront. This subchapter also evaluates different modelling approaches to account for choice set size variation in an RRM model. Finally, subchapter 5.3 is focused on analysing respondent heterogeneity by identifying possible latent classes within the sample with the help of an LC model, which also uses socio-demographic data the respondents provided.

As there are two main topics to comment on regarding the model outcomes, this chapter will focus more on the modelling aspect and the resulting decision rule from the estimation process, meaning the evaluation of model fits, scale parameters (μ) and constants accounting for choice set size variability (Λ). A more detailed analysis of the attribute estimates is given in chapter 6, where the resulting models will also be put to the test in a real-world environment and their predictions evaluated.

5.1 Implied decision rule

Using fixed values of the scale parameter μ, four models covering all four RUM and P-RRM combinations for best and worst choices were estimated. For a RUM decision rule, the scale parameter value was fixed to ten, while for P-RRM the value was 0,1. In all four models there were 5184 observations and eight estimated parameters in addition to the two fixed scale parameter values, bringing the total number of parameters to ten. The results of the estimations are presented in Table 5.1. Based on the final log-likelihoods, models I.1 and I.3 performed almost equally well, with the difference being only 0,08 log-likelihood points. These two models both assume a RUM decision rule for best choices but make different assumptions on the worst decision rule. Because RUM and P-RRM yielded very similar model fits, there is a possibility that the peak of the log-likelihood function lies somewhere in-between.

Models I.2 and I.4 had a much lower final log-likelihood, both assuming the decision rule P-RRM for best choices. These results point towards the fact that in this dataset, RUM was clearly the decision rule used when choosing the best alternatives in a choice set. The decision rule of worst choices is less conclusive, with results pointing towards a decision rule lying between the RUM and P-RRM decision rules, possibly the regular RRM, which is semi-compensatory and is characterised by a scale parameter value of one.

Table 5.1. Modelling outcomes with different fixed values of the scale parameter

	I.1	**I.2**	**I.3**	**I.4**
Decision rule				
Best	RUM	P-RRM	RUM	P-RRM
Worst	RUM	P-RRM	P-RRM	RUM
Null LL	-6204,59	-6204,59	-6204,59	-6204,59
Final LL	-5346,00	-5391,87	-5346,08	-5396,12
Rho-squared	0,1384	0,1310	0,1384	0,1303

To get a better understanding of model comparisons, a Ben-Akiva and Swait test was performed on all combinations of model pairs and the results are presented in Table 5.2. The Ben-Akiva and Swait test "*gives upper bound for the probability that the better fitting model is the incorrect model in the population, although having a better model fit for the particular sample*" (Chorus, 2016, p. 288). The order of model fits is 1-3-2-4 from best fitting to worst fitting model. Based on the final log-likelihoods, only very similar-fitting models would produce a meaningful upper bound probability, which is exactly what can be seen in Table 5.2. Models I.2 and I.4 were relatively similar in their model fit, but much worse than the other two. That is why in their comparison with the two better-fitting models, the probability is essentially zero. When compared to each other, model I.4 has a very small probability (0,18%) of being the true model in the population. The closest model fit was between models I.1 and I.3 and the test proves this, by concluding that model I.3 has a 34% probability of being better in the population than model I.1.

Table 5.2. Ben-Akiva and Swait test comparing the five models with fixed scale parameters

		Worse fitting models			
		I.1	I.2	I.3	I.4
Better fitting models	I.1		4,93E-22	0,34	6,74E-24
	I.2				0,0018
	I.3		5,34E-22		7,31E-24
	I.4				

5.2 Estimated decision rule

Following the testing with fixed scale parameters, models were also estimated where the scale parameters were to be determined in the estimation process. Five developed models were estimated and compared to three established models. Of the five models developed in this research one had a single scale parameter and three constants accounting for the choice set size variation, while the other four models had either two or four scale parameters and one or three constants to account for choice set size variation. Model E.4 estimated to determine if using the SBWMNL RUM is the best option when implying a single decision rule across both the entire sample and the different choices (best / worst) of each respondent. The other four SBWMNL μRRM models were used to explore the decision-making process in different choice tasks (best or worst), which presents the core of this research. The model fit results of all eight models are presented in Table 5.3.

Table 5.3. Model fit outcomes with estimated scale parameter values

	E.1	**E.2**	**E.3**	**E.4**	**E.5**	**E.6**	**E.7**	**E.8**
	First-choice RUM	First-choice μRRM	SBWMNL RUM	SBWMNL μRRM				
# of μ				*1*	*2*	*4*	*2*	*4*
# of Λ				*3*	*3*	*3*	*1*	*1*
Null	-2085,83	-2085,83	-6204,59	-6204,59	-6204,59	-6204,59	-6204,59	-6204,59
Final	-1549,08	-1548,82	-5397,20	-5346,00	-5340,49	-5337,51	-5372,99	-5372,95
Rho-squared	0,2573	0,2575	0,1301	0,1384	0,1393	0,1397	0,1340	0,1340
Observations	1296	1296	5184	5184	5184	5184	5184	5184
Parameters	5	6	5	9	10	12	8	10
BIC	3134,00	3140,64	10837,17	10768,97	10766,51	10777,67	10814,40	10831,43

Starting with the two first-choice only models (E.1 and E.2), the very minimal difference in their model fit indicates that a RUM decision rule can be assumed for first choices. Although model E.2 has a slightly higher rho-squared value, the additional estimate reduces its parsimony. Conducting a likelihood ratio test, LRS = 0,527 with df = 1, resulting in a p-value of 0,47, meaning there is minimal difference between the models. Comparing the first-choice models to the rest, using the final log-likelihood is not an option due to the major difference in sample size. They can be compared based on the rho-squared value, which is much higher for the first-choice only models, almost twice as high as for the SBWMNL models.

Considering the six SBWMNL models, all five developed models achieved better model fits than the SBWMNL RUM model on the given dataset. The best model fit was achieved by model E.6, which makes sense as it has the highest number of estimates and can thus account best for variation in the data. The BIC value places it in the third position, after models E.5 and E.4 respectively, while the Ben-Akiva and Swait test states there is a near zero possibility of E.5 and E.4 being the true models of the population compared to E.6. Comparing the models with respect to the number of scale parameters (either two or four), the differences seem to be minimal. Comparing the change between a single Λ value or three Λ values to account for choice set size variation on the other hand, gives far more significant differences in model fit. Models E.7 and E.8, using a single Λ, score 30 log-likelihood points lower than models E.4, E.5 and E.6 which use Λ for each choice set size. This results in a rho-squared that is 0,5 percentage points lower and a higher BIC value, much closer to the SBWMNL RUM model than the other three models. Using the LRT, only two model pair combinations had a p-value of 0,01 or more. Between models E.5 and E.6, the better fitting E.6 model has a 5% probability of having a better fit due to coincidence, while the Ben-Akiva and Swait test concluded model E.5 only has a 0,241% chance of being the true population model. Comparing models E.7 and E.8 with an LRT, the fact that E.8 achieved a better fit is almost entirely due to coincidence, as the p-value is 0,96. This is echoed in the Ben-Akiva and Swait test as well, as the probability was the highest out of all model pair combinations, at 7,456%.

Based on the model fit results, using multiple lambdas to account for choice set size variation is better than a single estimated constant, while increasing the number of scale parameters from two to four does not provide much additional information. This could indicate a similar decision rule for choices of the same kind (best or worst), regardless of it being the first or third choice. This is explored in Table 5.5. Along with Table 5.4, they show the model estimates of all eight models, including parameter estimates as well as scale parameters and choice set size constants.

Table 5.4. Model estimates of models E.1, E.2 and E.3

	E.1			E.2			E.3		
	Est	**SE**	**p-val**	**Est**	**SE**	**p-val**	**Est**	**SE**	**p-val**
Car	-0,0874	0,00	0,00	-0,0350	0,00	0,00	-0,0581	0,00	0,00
PT	-0,0601	0,00	0,00	-0,0241	0,00	0,00	-0,0453	0,00	0,00
Cost	-0,2918	0,01	0,00	-0,1168	0,00	0,00	-0,2274	0,01	0,00
Headway	-0,0643	0,00	0,00	-0,0257	0,00	0,00	-0,0383	0,00	0,00
Mode	0,1875	0,06	0,00	0,0770	0,00	0,00	0,1864	0,04	0,00
μ				53,50	0,00	0,00			

In all SBWMNL μRRM models, the value of μBest(1) had to be fixed to ten, as the model did not converge otherwise. As the value of the scale parameter increases, the difference in decision rule it represents becomes smaller. After surpassing the value of ten, fully compensatory behaviour can be assumed, and all higher values do not add information. The model however keeps increasing the scale parameter value but finds only miniscule differences in final log-likelihood values and thus does not converge. In all cases of the scale parameter value being fixed to ten, the model was first run without this limitation and did not converge, while the scale parameter value was estimated to be anywhere between 12 and 60. The parameter was fixed to ten and the model re-estimated achieving convergence in all models except for model E.4. As the μ and Λ values are rooted in one another, Λ5 (for the first best choice) was also fixed and the model estimated a third time. The value of the scale parameter accounting for the second worst choice made by respondents was also fixed in both cases (models E.6 and E.8). This is because the final

Table 5.5. Model estimates of models E.4, E.5, E.6, E.7 and E.8

	E.4			E.5			E.6			E.7			E.8		
	Est	SE	p-val	Est	SE	p-val	Est	SE	p-val	Est	SE	p-val	Est	SE	p-val
Car	-0,0371	0,00	0,00	-0,0367	0,00	0,00	-0,0366	0,00	0,00	-0,0368	0,00	0,00	-0,0369	0,00	0,00
PT	-0,0282	0,00	0,00	-0,0282	0,00	0,00	-0,0282	0,00	0,00	-0,0280	0,00	0,00	-0,0281	0,00	0,00
Cost	-0,1412	0,01	0,00	-0,1399	0,02	0,00	-0,1392	0,02	0,00	-0,1367	0,02	0,00	-0,1371	0,02	0,00
Headway	-0,0260	0,00	0,00	-0,0264	0,00	0,00	-0,0265	0,00	0,00	-0,0267	0,00	0,00	-0,0268	0,00	0,00
Mode	0,1004	0,02	0,00	0,1023	0,03	0,00	0,1037	0,03	0,00	0,0954	0,03	0,00	0,0957	0,03	0,00
μ	10,00	---fixed---													
μBest				10,00	---fixed---					10,00	---fixed---				
μWorst				0,43	0,17	0,01				0,36	0,13	0,00			
μBest 1							10,00	---fixed---					10,00	---fixed---	
μWorst 1							0,43	0,17	0,01				0,36	0,13	0,00
μBest 2							0,50	0,25	0,05				12,31	20,54	0,55
μWorst 2							1,00	---fixed---					1,00	---fixed---	
Λ										5,23	0,66	0,00	5,25	0,67	0,00
Λ 5	0,86	---fixed---		0,87	0,11	0,24 [1]	0,88	0,11	0,27 [1]						
Λ 4	0,61	0,05	0,00 [1]	0,70	0,10	0,00 [1]	0,71	0,10	0,00 [1]						
Λ 3	1,00	0,07	0,95 [1]	1,00	0,13	0,99 [1]	1,11	0,16	0,48 [1]						

[1] the p-value is calculated for the significance of difference from the value of 1

choice was made between only two alternatives, meaning that there is no modelling difference between a P-RRM, regular RRM or RUM model. The scale parameter value was arbitrarily fixed to one but could have also been fixed to 0,1 or ten with no difference in modelling outcome.

Considering first the parameter value estimates, all achieved a very low p-value (all 0,00), meaning high significance for all five attributes in describing the decision-making process. Among the five SBWMNL μRRM models, the differences between the estimates are minimal, while the estimates of models E.1, E.2 and E.3 are more than two times higher (lower in the case of negative parameter estimates). E.2 achieved the closest values to the SBWMNL μRRM models, most likely due to the μRRM model itself. The differences in parameter estimates are mostly a result of different model formulations.

Looking into the ratios of the estimates (Table 5.6) reveals that all the intensities of the parameters increase/decrease by about the same level, as the ratios are very similar. Referring to parameter ratios as values of time is not correct for RRM models: the explanation and justification of their use is given below. The results in Table 5.6 show that models E.1 and E.2 identified higher VoT for car travel time and the headway of public transport compared to the other models, with value of mode being lower and travel time by public transport being the most similar to the other models. The six SBWMNL models have very similar results for car and public transport parameter ratios, with E.3 achieving a slightly lower value in both. Models E.7 and E.8 have the highest first three ratios and the lowest mode-cost ratio, which is most likely related to the use of a single constant accounting for the choice set size variation. Interesting to note is that in the public transport and headway ratios, model E.4 lies between model E.3 and the other SBWMNL μRRM models, indicating the fact that it is the most similar to the RUM model, by implying a single decision rule (which turns out to be RUM) across all choices.

It must be mentioned at this point that in RRM models, parameter ratios do not translate to value of time (VoT), as the parameters in RRM models evaluate the difference between alternatives' attribute performance, while in RUM models, parameters evaluate the performance of an attribute's parameters independently from other alternatives. Therefore, RUM model parameter ratios can be interpreted as VoT, while RRM parameter ratios cannot.

However, as the value of scale parameter for the first-best choice is synonymous with a RUM model and also to allow for easier interpretation of the estimates, value of time calculations are made and are referred to as the marginal rate of substitution or simply parameter ratios.

Table 5.6. Parameter ratios of the six estimated models

	E.1	E.2	E.3	E.4	E.5	E.6	E.7	E.8
Car travel time value	17,97	17,97	15,34	15,77	15,74	15,76	16,14	16,14
PT travel time value	12,36	12,36	11,96	11,97	12,09	12,17	12,30	12,30
Headway value	13,21	13,20	10,10	11,05	11,34	11,41	11,73	11,73
Mode value	0,64	0,66	0,82	0,71	0,73	0,75	0,70	0,70

The scale parameter values of the five SBWMNL μRRM models reveal the decision rules used by respondents for making choices in the presented survey. They prove the hypothesis that a RUM decision rule is utilised for best choices and an RRM rule for worst choices. Looking at the scale parameter for worst choices, it takes the value of somewhere around 0,5 in all four models, indicating a combination of regular RRM and P-RRM. This means the behaviour is not non-compensatory but is quite strongly semi-compensatory: better performance on an attribute does help an alternative when the difference is small, while for much better performance, the alternative does not gain any added value. Performing worse on an attribute on the other hand has consequences, that increase linearly with the difference in performance. The performance of the scale parameter value of μ(Worst)=0,43, compared to the typical P-RRM, regular RRM and RUM decision rules, is depicted in Figure 5.1. In addition, Figure 5.2 depicts the estimated scale parameter values in all four estimated models with multiple scale parameters.

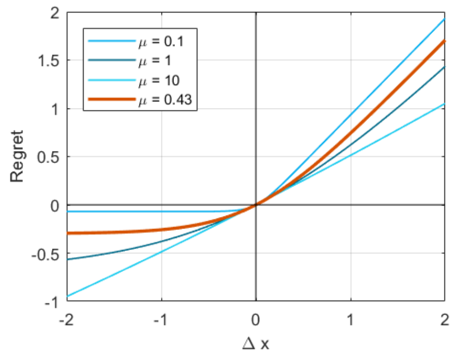


Figure 5.1. Contributions to regret with different scale parameter values, including the resulting 0.43 value

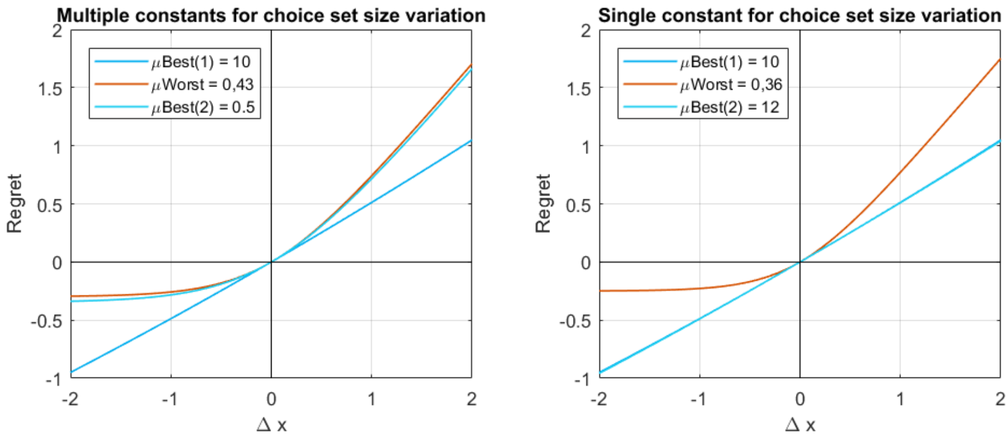


Figure 5.2. Scale parameter values of Models E.5 & E.6 (left) and Models E.7 & E.8 (right)

Examining the values of scale parameters for each individual choice, Figure 5.3 shows the relationship between the μ and the final log-likelihood of the model. The value of the scale parameter for worst choices takes the same value in both a 2- and 4-scale parameter model, as the second-worst choice is, as previously mentioned, between two alternatives and thus any scale parameter value yields the same model fit. For this reason, a separate depiction for the μWorst 2 is not given in Figure 5.3. What this means for the scale parameter for worst choices in models E.5 and E.7, it can simply take the value that offers the best fit for the first-worst choice. For that as well, the estimated μ(Worst) was the same in models E.5 and E.6 as well as in models E.7 and E.8.

The scale parameter for best choices is different however, as even the second-best choice is still among three alternatives. The first-best choice was always fixed to 10, indicating fully compensatory behaviour. The second-best choice however takes completely different values in models with a single Λ compared the model with multiple Λ. When a single constant for choice set size variation is used (model E.8), the scale parameter seems to indicate a RUM decision rule (value of 12,31), but the p-value is substantial (0,55) meaning high insignificance of the estimate. In the model with multiple constants for choice set size variation (E.6), the scale parameter converges to a value of 0,5, indicating semi- to non-compensatory behaviour, as the worst choice scale parameter. In this case, the significance is also substantially higher, with a p-value of 0,05. The clear difference of the μBest 2 can also be seen in Figure 5.3 where the log-likelihood functions for either a single or multiple Λ are completely different.

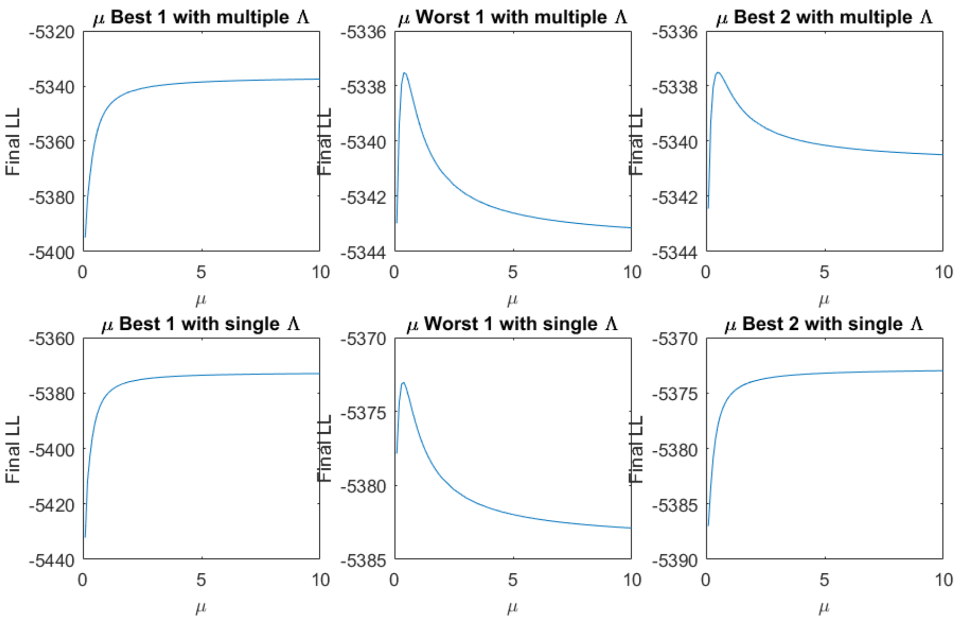


Figure 5.3. Relationship between the μ value and the final log-likelihood in models 4 and 6

To evaluate the rate of compensatory behaviour, profundity is calculated for all the attributes, scale parameters and models (van Cranenburgh, Guevara, et al., 2015). The results prove what has already been stated in the paragraphs above regarding the scale of compensatory behaviour observed by respondents. Profundity goes a step further by classifying the behaviour for each parameter as well. Table 5.7 show the average profundities across the four attributes that can be used for analysing the decision rule (mode has only two attribute levels and thus cannot be used). The averages prove that indeed μ(Best1) is almost fully compensatory in all models, as the value of α approaches zero. The differences between individual attributes were not extensive, with the least compensatory being cost in all models, but still only achieving a value of α =0,037. μ(Worst) also proves that the behaviour is somewhat semi-compensatory, but looking into the attribute levels, cost was again the least compensatory, resulting in values between α=0,67 and α=0,73. Among the other three attributes, PT travel time is most compensatory, but still only going as low as α=0,40. The attributes for headway and car travel time were in the middle, mostly around the 0,5 value which translates as semi-compensatory.

For μ(Best2), the difference between models E.6 and E.8 can be seen here as well. In E.6, the value of the scale parameter was slightly higher than that of μ(Worst) and a similar result can be seen here. The same can be said for E.8, where the scale of compensatory behaviour is similar to that of the first-best choice. In both cases, the relationships between individual attributes stayed the same, with the cost attribute being least compensatory (α=0,61 in E.6 and α=0,03 in E.8) and PT travel time being the most compensatory (α=0,35 in E.6 and α=0,02 in E.8).

Table 5.7. Average profundities for each scale parameter in each model

	E.4	**E.5**	**E.6**	**E.7**	**E.8**
μ	0,0257				
μBest		0,0256		0,0255	
μWorst		0,5072		0,5736	
μBest 1			0,0256		0,0255
μWorst 1			0,5068		0,5749
μBest 2			0,4516		0,0208

The constants accounting for choice set size variation were investigated on the significance of their difference from one (instead of zero as for the other estimates). $\Lambda 4$ proved highly significant in all cases, while $\Lambda 5$ and $\Lambda 3$ did not. The latter especially had very high p-values in models E.4 and E.5, where it is virtually the same value as one. These differences and (in)significances are most likely due to Λ and μ being rooted in each other.

The major difference between the scale parameters in different models seem to be rooted in the way the model accounts for choice set size variation. Looking at the values of the constants (Table 5.8), there seems to be a major difference between the estimated and calculated values in choice sets with three alternatives. Estimated values are from models where multiple constants were estimated, while calculated values are based on the single constant used for choice set size variation, adjusted for choice set size in the same way as in the model. The values indicated with an asterisk show the calculated value of the constant, used in the model, based on the single estimated constant. The results of models E.5 and E.6 show clearly, that there is no linear relationship between the constants, as the middle value is the lowest. Since the scale parameter and choice set size constant are rooted in one another, the values of the constants seem to be related to the best-worst choices and thus a quadratic relationship could also not be appropriate. A linear relationship accounting for seasonality (seasonality representing the best and worst choice as different seasons and choice set size the time) could yield good results, but a greater number of different choice set sizes would be needed to make meaningful conclusions. As the model was constructed linearly, the estimation process had to find the best fit within this formulation. The result indicates that the linear relationship is strongest between Λ 5 and Λ 4 and brings the best improvements in log-likelihood. This is proven by the relatively similar values of these constants across all four models. Figure 5.4 shows the relationship between Λ 5 and Λ 4 where the estimated line (light blue) follows mostly the same path as the line between Λ 5 and Λ 4. In the figure, the three constant values estimated separately for each choice set size are shown in orange with the single constant represented as a light blue line (dependent on choice set size). A linear regression line was estimated based on the estimated Λ values from model E.5 and is shown in dark blue. There seems to be a strong relationship between the scale parameter value and the constant for accounting for choice set size variation, so it is analysed further in subchapter 5.4.1.

Table 5.8. Constants accounting for choice set size variation

	E.5	**E.6**	**E.7**	**E.8**
Λ			5,23	5,23
Λ 5	0,87	0,88	*0,96 **	*0,96 **
Λ 4	0,70	0,71	*0,76 **	*0,76 **
Λ 3	1,00	1,11	*0,57 **	*0,57 **

** values are calculated from the single Λ constant*

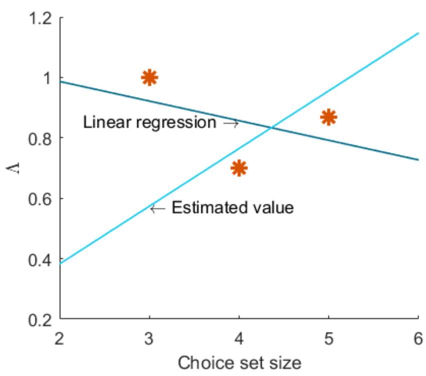


Figure 5.4. Comparison of single (lines) and multiple (dots) constants for choice set size variation and linear regression

5.3 Latent classes of decision rules

The final group of models to be estimated on the sample are the LC models. Unlike previously estimated models which assume a homogeneous sample, LC models can identify possible latent user groups. The key motivation for estimating these models was the notion that different people may utilise different decision rules in everyday life. This idea is also backed by the resulting distribution of respondents on the regret scale, proving that people experience regret differently when presented with a choice situation and thus may utilise different decision rules based on their level of regret-averse behaviour.

LC models are a straightforward and insightful way of identifying and analysing latent user groups in the sample. Its strength also lies in the capability of using socio-demographic data to help predict the class allocation of individuals. These were the key reasons for using LC models to analyse respondent heterogeneity, instead of other models in the field, most notably the mixed logit model.

The model was first estimated solely with a constant that divides the population based on their different values of time and/or their decision rules, and secondly by including a predictor variable in the class allocation function that accounts for respondents' socio-demographic characteristics.

5.3.1 Latent class models without socio-demographic predictor variables

A 2-class LC model was estimated in which all scale parameters were estimated from the data and the outcomes are presented in Table 5.9 (model fit) and Table 5.10 (parameter estimates). Although the model did not converge, a clear improvement in model fit can be observed in LC.0 compared to the single class E.3 and E.5 models, as the final log-likelihood increased substantially. This improvement is also echoed in the higher rho-squared value, over 0,2. Better model fit alone does not automatically mean a better model, as the LC.0 model uses a far greater number of parameters compared to the previously estimated models. It utilises 21 parameters, compared with the five and ten used in the E.3 and E.5 models respectively. Carrying out the LRT between LC.0 and E.5 yields a p-value of almost equal to zero. The χ^2 value for p-val=0,01 and df=11 is 24,725, whereas the LRT is 864,82.

Table 5.9. Results of the latent class model compared with the previously estimated SBWMNL RUM and RRM models

	LC.0 model	E.3 model	E.5 model
Null LL	-6204,5893	-6204,59	-6204,59
Final LL	-4908,0821	-5397,20	-5340,49
Rho-squared	0,2090	0,1301	0,1393
Observations	5184	5184	5184
Parameters	21	5	10
Size of Class 1	52%		
Size of Class 2	48%		

Looking at the scale parameters of both classes (Table 5.10), two different user groups with different decision rules emerge from the estimation. In class 1, users tend to utilise a RUM decision rule (μ value above 10) for best choices, while the μWorst is estimated to be around 0,7, indicating a semi-compensatory behaviour that is leaning more towards non-compensatory (P-RRM). For the second class, the decision rule for both best and worst is semi-compensatory behaviour leaning towards fully compensatory, as it takes the value of around two for both scale parameters. It should be noted that the μWorst parameter for class 1 had a p-value of 0,36, indicating relatively high insignificance, while the one for class 2 was significant at the 5% level, but not the 1% level, unlike other estimates.

Of interest in an LC model is also the class allocation parameter(s), δ in the case of model LC.0. A rather surprising result with a very high p-value, at first indicating high insignificance of the parameter. However, the low value of the constant and the high p-value mean that the classes are mostly equally sized, which is also shown in Table 5.9, where it states that class 1 is only slightly larger than class 2.

Table 5.10. Parameter estimates of a 2-class latent class model

	Class 1			Class 2		
	Est	SE	p-val	Est	SE	p-val
Car	-0,0321	0,01	0,00	-0,0802	0,01	0,00
PT	-0,0272	0,00	0,00	-0,0604	0,01	0,00
Cost	-0,0862	0,01	0,00	-0,3869	0,04	0,00
Headway	-0,0388	0,01	0,00	-0,0300	0,00	0,00
Mode	0,1600	0,04	0,00	0,1350	0,05	0,01
μBest	61,4545	11,45	0,00	2,0290	0,65	0,00
μWorst	0,7233	0,79	0,36	2,0391	0,84	0,02
Λ 5	0,8531	0,14	0,00	0,6025	0,07	0,00
Λ 4	0,4831	0,11	0,00	0,6249	0,07	0,00
Λ 3	0,9192	0,17	0,00	0,8874	0,11	0,00
δ_2				-0,0729	0,20	0,72

From the results in Table 5.10, clear differences can also be seen with respect to the attribute parameters of travel time, cost etc. People in different classes clearly seem to value the attributes differently, but in order to better understand this, parameter ratios (marginal rate of substitution) are very helpful in understanding the differences between both classes (Table 5.11). Members of class 1 appear to value all travel aspects higher than the members of class 2. For both the car and public transport in-vehicle travel times, they value it almost twice as high, with the difference in value of headway and mode being even more pronounced (around five times higher). Whether these differences can be explained by socio-demographic characteristics of respondents (such as income) will be analysed in subchapter 5.3.2.

Table 5.11. Parameter ratios of both classes in the latent class model

	Class 1	Class 2
Car travel time value	22,31	12,44
PT travel time value	18,96	9,36
Headway value	53,99	9,32
Mode value	1,86	0,35

5.3.1.1 Latent class models with implied decision rules

Like the single class models, LC model with fixed values of the scale parameters were also estimated to explore the differences between different implied decision rules. A total of six models with all possible decision rule combinations were estimated, with the results presented in Table 5.12. Due to a substantial increase in combinations of classes and decision rules, only RUM or P-RRM were used.

From the perspective of model fit, the best fitting decision rule combination is RUM (best) – RUM (worst) for one class and RUM (best) - P-RRM (worst) for the other (model LC.2). Conducting the Ben-Akiva and Swait test proves the superiority of model LC.2 over all other models. Interestingly, the closest probabilities in the Ben-Akiva and Swait test were in pairs LC.3-LC.6 (Pr=2,583%) and LC.1-LC.4 (Pr=0,870%). In both pairs, the worse fitting models only change one of the worst choices from RUM to RRM, with the other three scale parameters staying the same. This shows that using RUM or P-RRM for the worst choice makes a smaller difference than changing the decision rule for best choices. The results of single class models with inferred decision rules (I.1 to I.4) proved the same, as the models with a RUM decision rule for best choices always performed better, while having a RUM or P-RRM decision rule for worst choices yielded roughly the same results. However, given the rather small differences in class sizes in all six models, the difference between classes may not be in the decision rule.

Table 5.12. Model outcomes of 2-class latent class models with fixed decision rule scale parameters

	LC.1	LC.2	LC.3	LC.4	LC.5	LC.6
Null LL	-6204,5893	-6204,5893	-6204,5893	-6204,5893	-6204,5893	-6204,5893
Final LL	-4980,2749	-4916,0396	-4974,074	-4983,1026	-4991,7254	-4975,9675
Rho-squared	0,1973	0,2077	0,1983	0,1969	0,1955	0,1980
Observations	5184	5184	5184	5184	5184	5184
Parameters	21	21	21	21	21	21
Size of Class 1	52%	47%	56%	47%	50%	43%
Size of Class 2	48%	53%	44%	53%	50%	57%
Class 1 μBest	10	10	10	0,1	0,1	0,1
Class 1 μWorst	10	10	10	0,1	0,1	10
Class 2 μBest	0,1	10	0,1	10	0,1	10
Class 2 μWorst	0,1	0,1	10	0,1	10	0,1
δ_2	-0,0781	0,1172	-0,2222	0,1033	-0,0112	0,2679
p-val	0,73	0,57	0,33	0,64	0,96	0,20

Looking at possible different explanation for the differences between latent classes, attention is turned to the parameter ratios of attribute parameters within each of the classes (Table 5.13). What emerges is, similar to the previous results, one class with much higher parameter ratios. Interestingly, regardless of the decision rules implied in the models, the ratios are almost the same across the different models with only small variations. Looking at the class sizes, the slightly bigger class is always the one with much higher ratios, possibly indicating that a slight majority of people value their time substantially more. Interestingly, all the larger classes also have a RUM decision rule for best choice. In LC.2 (the only model with both classes having RUM for best choice), the larger class is that with a P-RRM decision rule.

Table 5.13. Parameter ratios of both classes in the latent class models with fixed scale parameters

		LC.1	LC.2	LC.3	LC.4	LC.5	LC.6
Class size	Class 1	52%	47%	56%	47%	50%	43%
	Class 2	48%	53%	44%	53%	50%	57%
Car travel time value	Class 1	23,04	12,71	22,57	11,57	23,01	10,48
	Class 2	11,59	21,57	10,73	22,82	11,41	22,49
PT travel time value	Class 1	19,46	9,29	18,47	9,41	19,88	8,61
	Class 2	9,40	18,43	8,74	19,25	9,10	18,36
Headway value	Class 1	54,22	7,64	50,10	13,20	55,65	10,78
	Class 2	13,23	52,25	10,92	53,29	11,43	49,01
Mode value	Class 1	1,81	0,24	1,61	0,60	1,94	0,48
	Class 2	0,60	1,83	0,47	1,82	0,43	1,59

The analysis of LC models with no additional predictor variables for class allocation proved that slight differences do occur when implying different decision rules, with a clear preference for RUM decision rules utilised in best choices. However, the main distinction between classes seems to be dictated by the attribute parameter values (parameter ratios) that members of both classes utilise. This means that no truly meaningful conclusion can be made regarding the LC models with respect to decision rule heterogeneity in the population.

5.3.2 Latent class models with socio-demographic predictor variables

The true advantage of LC models in uncovering respondent heterogeneity is in their ability to utilise socio-demographic characteristics to predict class membership of individuals. Four socio-demographics were used in four separate models to see if they can assist in predicting class membership: the regret scale (regret-averseness), income, gender and age. Results of the four models can be seen in Table 5.14. As some respondents did not provide their socio-demographic characteristics, their data was excluded from the model estimation, resulting in a lower number of observations.

Considering the class allocation parameters (constant δ_2 and socio-demographic parameter γ_{2D}), they are all highly insignificant, pointing to the fact that none of the socio-demographics can be used to accurately predict class membership. The higher rho-squared, initially pointing towards added value of using socio-demographics, is in fact a consequence of the reduction in sample size. As stated in the paragraph above, the sample was reduced due to missing socio-demographic data. This was most notable in the case of income, seen in Table 5.14, as the number of observations is significantly lower. By excluding those who did not wish to reveal their personal information, model fit improved solely because their stated choice answers appeared to be more random. This was proven by running an LC model with the reduced dataset (reduced by the respondents who did not disclose their income) and not accounting for socio-demographics. The resulting rho-squared of this model was the same as the rho-squared of the LC model accounting for the income parameter.

From the results of latent class models that include socio-demographic data, an insightful conclusion cannot be drawn. What would appear to be a problem is that either a correlation between class allocation and class membership is not there, or the sample may be too small to produce a significant result. Although the best-worst method is praised for its ability to identify significant results in samples with a smaller number of respondents (Lancsar et al., 2013), this does not fully extend to an LC model, as class membership is estimated per respondent and not per respondent's individual choice. Increasing the number of choice situations per respondent and reducing the number of respondents, while good for the budget and efficiency of the choice model, is unfavourable for an LC model.

Table 5.14. Model outcomes from latent class models with different socio-demographics used as predictor variables

	LC.Regret	LC.Income	LC.Gender	LC.Age
Null LL	-6089,69	-5227,94	-6089,69	-6147,14
Final LL	-4802,38	-4053,89	-4781,66	-4841,30
Rho-squared	0,2114	0,2246	0,2148	0,2124
Observations	5088	4368	5088	5136
Size of Class 1	41%	32%	30%	30%
Size of Class 2	59%	68%	70%	70%
μBest	2,0285	55,1344	66,0063	63,1753
μWorst	2,1187	1,1216	0,6068	0,6480
μBest	64,5789	2,1724	2,0065	2,0107
μWorst	0,7583	2,6548	2,0204	1,9714
δ_2	-0,0010	0,0115	0,0009	-0,0081
p-val	0,92	0,77	1,00	0,95
γ_{2D}	0,3530	0,7675	0,8242	0,8242
p-val	1,00	0,99	1,00	0,99

5.4 Model performance analysis

This subchapter looks further at some modelling peculiarities, like the intricate relationship between the μ and Λ values, which seem to influence one another depending on how they are accounted for in the model. Model performance is analysed through a k-fold validation process with respect to hit rate and loglikelihood. Finally, the prior values used in the generation of the experimental design are assessed and the choice of the design generation process is evaluated.

5.4.1 Relationship between μ and Λ

As has been seen in the previous subchapter, the way the Λ value was specified drastically changed the value of the scale parameter for the second-best choice in the survey. The constants also seemed to vary at random and not in relation to the choice set size. Van Cranenburgh, Prato, et al. (2015) provide an explanation for this issue that also explains why the values of μ have performed differently. Van Cranenburgh, Prato, et al. (2015, p. 4) state that the choice of size of the constant (lambda): "*is inconsequential. In the μRRM model the scale parameter μ is estimated, as opposed to being implicitly fixed to 1 as in the classical RRM model. Since Λ is perfectly confounded with μ, setting Λ to a large value will merely result in estimating a small-scale parameter μ, and vice versa.*" This can especially help in explaining why the constant for choice sets with four alternatives is rather small compared to the other two: it is because in this choice set size, the respondents were tasked to choose the worst alternative, meaning the scale parameter value of 0,1, while in choice sets with five and three alternatives, the choice was for the best alternative (μ = 10). In this subchapter, this relationship is investigated further. In all examples, the scale parameter for the second-best choice is used, as it was the parameter that experienced change in decision rule when the Λ value changed.

Multiple log-likelihood functions with respect to the scale parameter values, for different Λ values are presented in Figure 5.5. It shows that with increasing Λ values, the peak of the log-likelihood function moves towards smaller μ values. When Λ reaches 0,8 the decision rule is already a regular RRM and the overall best fit is achieved with the Λ = 1,1, the same value it took when estimated with multiple choice set size constants.

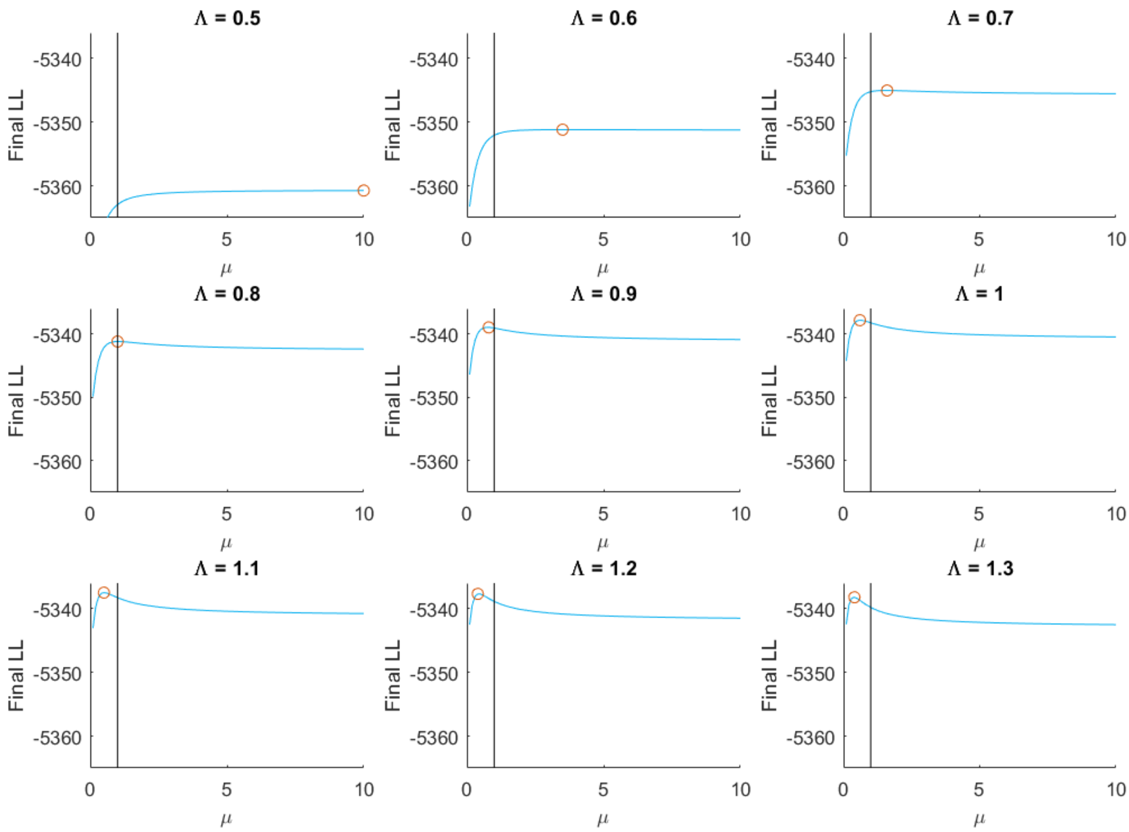


Figure 5.5. Log-likelihood functions with respect to μ for different Λ values

The log-likelihood functions were also combined in a single chart in Figure 5.6, along with a chart depicting the relationship between Λ and μ, showing an inverse relationship: as one increases, the other decreases. The best fitting combination of μ and Λ is also shown in the figure. Given that the relationship seems to be inverse, two separate functions were fitted to this data to see how well they fit. The blue line depicts the typical inverse relationship g(x)=a/x, where a=0,4603 and a model fit R^2=61,73%. As the figure only contains scale parameters of 0,5 or higher, this was the best model fit. Smaller values were also computed but due to non-convergence when larger scale parameter values are estimated (over ten), the scale parameter was fixed. As a μ value of 10 or 100 yields essentially the same result, a slight adjustment to the formulation was made: f(x)=a/x + b. This model produced a very good model fit of R^2=95,45%, where a=0,3054 and b=0,5132. This formulation is somewhat problematic, as the function limit when x→∞ is 0,5132 (the equivalent of b). The limit implies the lowest possible value of Λ, which in fact can be below this level, down to 0. For all mathematical purposes, the typical inverse relationship is the only true representation, but given the characteristics of μ, some ambiguity remains.

This relationship is important to understand, as by implying the Λ value, we may influence the decision rule emerging from the model. It is therefore best practice to estimate both values and allow the model to find the best fit with as few restrictions and limitations as possible.

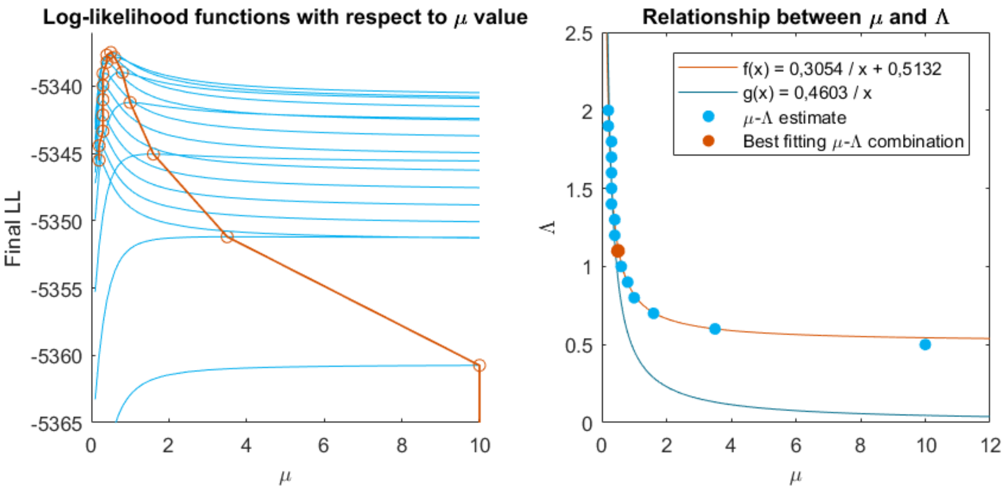


Figure 5.6. Log-likelihood functions with peak values (left) and Λ values (right) with respect to μ values

5.4.2 Model validation

Before the application of a model, it is important to validate it and, in this case, compare different models and their predictive ability on the same dataset. A k-fold validation was performed, meaning that the complete dataset was split into three equal subsets and in three iterations, two of the subsets were used for model estimation, with the third being withheld for validation purposes. In each of the iterations, a different subset was used for validation. The results of the validation process, showing both the log-likelihood of the hold-out sample and the hit rate are presented in Figure 5.7.

The first observation comparing the three subsets is that the third set seemed to be best predicted overall, as it has both the highest final log-likelihood and highest hit rate for all five different models. Considering the individual model performance in terms of final log-likelihood, models E.5 and E.6 performed best on all three sets, with E.6 performing slightly better in two out of three cases. The hit rate of the two models on the other hand was much more varied and not clearly better or worse than the other three models. In set 1 they performed worst, while on set 3 they were the best. Models E.7 and E.8 performed almost exactly the same (with respect to each other) on all three sets in terms of both log-likelihood and hit rate. The E.3 model performed reasonably well, given the much lower number of parameters compared to its RRM model counterparts. It performed worst in terms of log-likelihood on all three datasets, while the hit rate was mostly similar to models E.7 and E.8. On set 2 however, the RUM model performed best among all five models.

Overall, models E.5 and E.6 have a slight upper hand in the validation process by performing best in log-likelihood on all sets and doing very well on hit rate. Their advantage is to be expected as they were the best performing models in terms of model fit as well. The differences in performance however, were not so great and especially with respect to hit rate the performance of all five models was very similar, with each validation set seeing a different model perform best and worst.

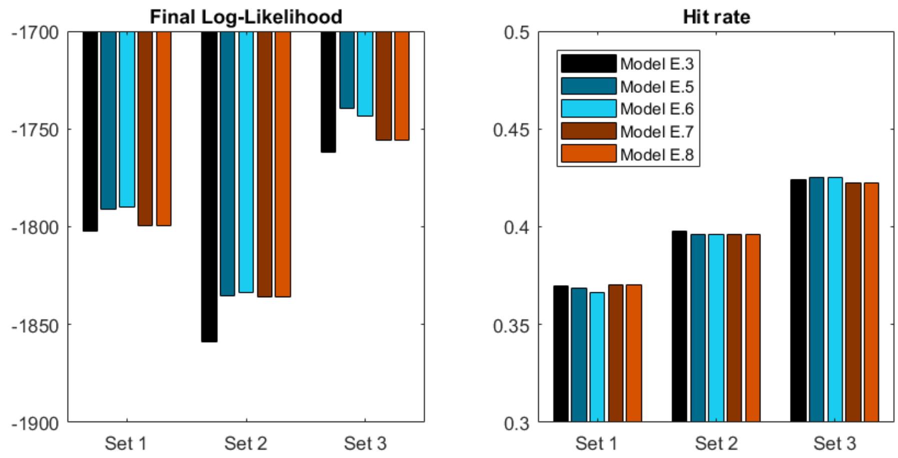


Figure 5.7. Validation results. Left: final log-likelihood, right: hit rate

5.4.3 Evaluation of parameters and comparison to prior values

Given the importance and influence of prior values on the experimental design of a stated choice survey, an evaluation of the priors is carried out, where the values are compared to the estimation outcomes of the five different models. Table 5.15 presents the parameter estimates of five estimated models and the prior values in the second column, while Table 5.16 presents the ratios of these parameters to get a better understanding of their relative meaning (the marginal rate of substitution).

The prior values correspond well with the estimate values of the RUM model in all but the headway parameter. When comparing the parameter ratios, the discrepancy in the headway estimate is clearly seen, with the priors and thus the experimental design assuming a much higher value of headway (much higher disutility for every additional minute between services). Although this could be due to the way this attribute was presented to respondents in the survey. They were not told explicitly how long they would wait at a stop but merely how often services are running. Many respondents also commented on their answers by stating that they would plan their arrival according to the timetable. For the other three parameter ratios, the differences are not so big with the closest matches being the value of mode (train) and the public transport in-vehicle time.

Although not immediately noticeable, the biggest difference may be a different valuation order of in-car travel time compared to PT travel time. The survey from which the priors were taken had the respondents valuing in-car time less than transit travel time, while the opposite was found in this survey. This means for an equal travel time by either bus or car, the results from this research mean that more people would opt for the bus, while based on the research from which the priors were taken, more people would have taken the car. This result is interesting and points to the possibility of respondents in this research not being transit or car commuters on a daily basis. A second option is perhaps the fact that in the Ljubljana Urban Region, transit travel time is always much longer than car travel time (or perceived as such by respondents) and the commuters would already be satisfied with public transportation services if the travel times of buses were equal to those of cars. It is important to note that in both surveys, the travel time by car and public transport was not used as a standalone travel time, but in a chain with a P+R facility in-between.

Given the estimation results and their difference compared to the prior values used, the argumentation for creating a Bayesian efficient design proved correct. Bayesian designs are, as already stated before, much more robust towards possible different values of time in the population (Walker et al., 2018).

Table 5.15. Parameter estimates compared to prior values used in the experimental design

		E.3	E.5	E.6	E.7	E.8
	Priors	SBWMNL RUM	SBWMNL μRRM			
# of μ			*2*	*4*	*2*	*4*
# of Λ			*3*	*3*	*1*	*1*
Car	-0,049	-0,0581	-0,0367	-0,0366	-0,0368	-0,0369
PT	-0,0535	-0,0453	-0,0282	-0,0282	-0,0280	-0,0281
Cost	-0,2575	-0,2274	-0,1399	-0,1392	-0,1367	-0,1371
Headway	-0,099	-0,0383	-0,0264	-0,0265	-0,0267	-0,0268
Mode	0,17	0,1864	0,1023	0,1037	0,0954	0,0957

Table 5.16. Parameter ratios from estimated models compared to ratios of prior values

	Priors	E.3	E.5	E.6	E.7	E.8
Car travel time value	11,42	15,34	15,74	15,76	16,14	16,14
PT travel time value	12,47	11,96	12,09	12,17	12,30	12,30
Headway value	23,07	10,10	11,34	11,41	11,73	11,73
Mode value	0,66	0,82	0,73	0,75	0,70	0,70

5.5 Conclusion

This chapter presented the analysis and model estimation undertaken on the 108 responses to the stated choice survey designed for this research. Although the data may not have been representative of the regional or commuting population, the results still provide a good basis for drawing behavioural conclusion regarding the decision-making process of individuals.

Implying the decision rules for best and worst choices, models achieved a better model fit if a fully compensatory behaviour was inferred for best choices, whereas worst choices did not immediately show a clear preference for either a RUM or P-RRM decision rule. A Ben-Akiva and Swait test carried out on the models proved this as well: comparing the two models with RUM for best choices and either RUM or P-RRM for worst choices, despite the former having a slightly better model fit, the latter still had a possibility of 34% of being the true population model.

Estimating the scale parameter, a high value for the first-best choice in all models proved that people do in fact use utility maximisation when selecting the first best alternative. The first worst alternative was selected with a decision rule between a P-RRM and a regular RRM (μ = 0,4), meaning not completely non-compensatory behaviour but not far from that. An interesting finding emerged when investigating all four (three) choices individually, as for the second-best choice, the decision rule changed depending on the way choice set size variation was accounted for: by either a single Λ value adjusted for choice set size, or multiple Λ values for each choice set size. This finding re-emphasized the relationship between the scale parameter and the choice set size constant, which when investigated turned out to be inverse.

Validation was performed on the five different models to see if differences between their respective predictions can be observed and the result was that E.5 and E.6 performed slightly better on log-likelihood, whereas on the hit rate none of the models could be seen as superior to the others.

Evaluating the resulting attribute estimates and comparing them to the prior values used in the design generation proved that the decision to use a Bayesian design, which allows for the researcher's uncertainty regarding the priors to be captured, proved to be the right decision, as some estimates had a higher value than the priors whereas others ended up lower.

To investigate the heterogeneity of decision rules between respondents, LC models were also estimated, both with and without using socio-demographic data in the class allocation function. The main takeaway from it was the model did see substantial improvement in terms of model fit compared to previously investigated models. The rho-squared increased from 0,13 to over 0,2. The decision rule however did not seem to play a role in this increased model fit, as the differences in model fits when implying different decision rule combinations were relatively minor. Moreover, the classes in all estimated models were very similar; one having high parameter ratios and the other (being slightly smaller in size) having low ratios. The class allocation constant was also found to be insignificant in many cases, indicating roughly equally sized classes. When four socio-demographics were added separately, model fit did seem to improve, but it was later determined to be caused by a reduction in sample size, rather than the inclusion of additional data. The result of the LC models was that no meaningful conclusion can be drawn regarding respondent heterogeneity when it comes to different decision rules.

The chapter gave valuable and detailed results on the model estimations and their performance. In the next chapter, a few of them will be put to the test on a realistic case study to see how they compare in making travel behaviour predictions.


LPP-245

6 Model application

This chapter puts the previously developed and analysed models to the test by applying them on a case study to determine the effects of a P+R facility development scheme in the CSSR (RDA LUR, 2014) and to analyse proposed alternative policy measures based on the link-and-ride concept, that has been identified as more sustainable compared to the traditional P+R scheme.

Two sets of scenarios have been constructed to test the evolution of applying different policy measures and understanding the effect of each one individually, as well as collectively. From the understanding gathered by the analysis, policy recommendations for future measures in the region are presented.

Three choice models have been selected to forecast travel behaviour for the proposed scenarios: the traditional first-choice RUM model, the SBWMNL RUM model (Lancsar et al., 2013) and the E.7 SBWMNL μRRM model developed in this research. The model performances are evaluated, and the differences of their predictions are commented upon.

The chapter begins with a detailed description of how the case study was constructed, the individual scenarios developed and where the data was obtained. Subchapter 6.2 describes the models that were selected for the application in the analysis. Finally, subchapter 6.3 presents the results and the discussion of the corresponding travel behaviour and model behaviour. With the understanding of travel behaviour obtained through the various scenarios, the subchapter concludes with policy recommendations.

6.1 Preparation of case study

For the model application to take place, a case needs to be carefully selected, with the goals of what the outcomes should show in mind, which in this research means (1) showing how different models make different predictions of travel behaviour and (2) forecasting market share of P+R facilities after the implementation of different policy measures.

Once selected, the case study must be properly assembled, including setting up representative, realistic and meaningful scenarios which help understand how commuters behave and will give an indication of what kind of travel behaviour to expect, given a certain policy measure. The data for the case must also be sourced carefully and thoroughly.

The following subchapter provides insight into what the exact case study is and why (subchapter 6.1.1). Subchapter 6.1.2 gives an overview of the area and the current situation that the case study is covering. It is followed by a detailed subchapter (6.1.3) where the construction of the case is described, including the detailed selection of scenarios, their corresponding characteristics and what their main goal and purpose are. Finally, subchapter 6.1.4 gives an explanation for the data gathering process; how travel times, frequencies and costs were obtained and why certain assumptions were made.

6.1.1 Selection of case study

For the model application, a case study from the CSSR was selected to evaluate the model performance and to analyse the current P+R plan of the RDA. A corridor to the southwest of Ljubljana was selected. On this corridor a total of three P+R facilities are already in operation, and the plan anticipates more being built in the future, giving the possibility of showing how these changes will affect mobility patterns. The corridor also gives the opportunity to make an alternative plan which could explore a different P+R concept and a comparison with the current plan could be made. This corridor also offers a unique opportunity of analysing the potential of (re)introducing railway services.

Along this corridor, the town of Bevke was selected as the origin. Although a small settlement, it can be seen as exemplary for the settlement structure of the CSSR and Slovenia, where the population density is low, with many villages located near transit corridors, but still too far to be accessible by either walking or cycling. Such villages are often very poorly connected themselves in terms of transit, despite often being close to higher quality transit corridors. Bevke itself has only two buses per day ('Vozni redi', n.d.), while being located only 2-3 km away from the Vrhnika-Ljubljana corridor, which sees a bus operating every fifteen minutes in peak periods ('Vozni redi', n.d.). A similar spatial pattern, with smaller towns and villages can be observed across the region and country and understanding how different measures and policies affect commuters' travel behaviour is crucial when creating transport policies.

6.1.2 Introduction on the case study area

The research area in this case study is a transport corridor to the southwest of the city of Ljubljana. It is one of the four highway corridors into the capital, meaning it is among the most congested. There are currently three P+R facilities already in operation, with more planned to be built in the near future. Two of the operational lots are of the satellite type and one is a remote P+R. There are two planned parking lots, located slightly further away from the city, but still fall mostly under the category of satellite P+R. P+R 3 could be classified as either satellite or remote, but given that a new highway exit is planned next to it, the satellite characteristics appear to be stronger. In addition to these five lots, this research proposes a sixth, remote P+R, located between the currently planned and operating satellite and remote parking lots. The complete list of parking lots along the SW corridor is summarised in Table 6.1 and a visualisation on a map is presented in Figure 6.1, along with car access routes, bus lines and a possible (not operating or officially planned) new railway line.

Table 6.1. List of P+R parking lots with their classification and status

P+R parking lot	Number	Type	Status
Sinja Gorica	P+R 1	Remote	In operation
Log pri Brezovici	P+R 2	Remote	Proposed in this research
Brezovica	P+R 3	Satellite / Remote	Planned
Pri Gorjancu	P+R 4	Satellite	Planned
Dolgi most	P+R 5	Satellite	In operation
Barje	P+R 6	Satellite	In operation

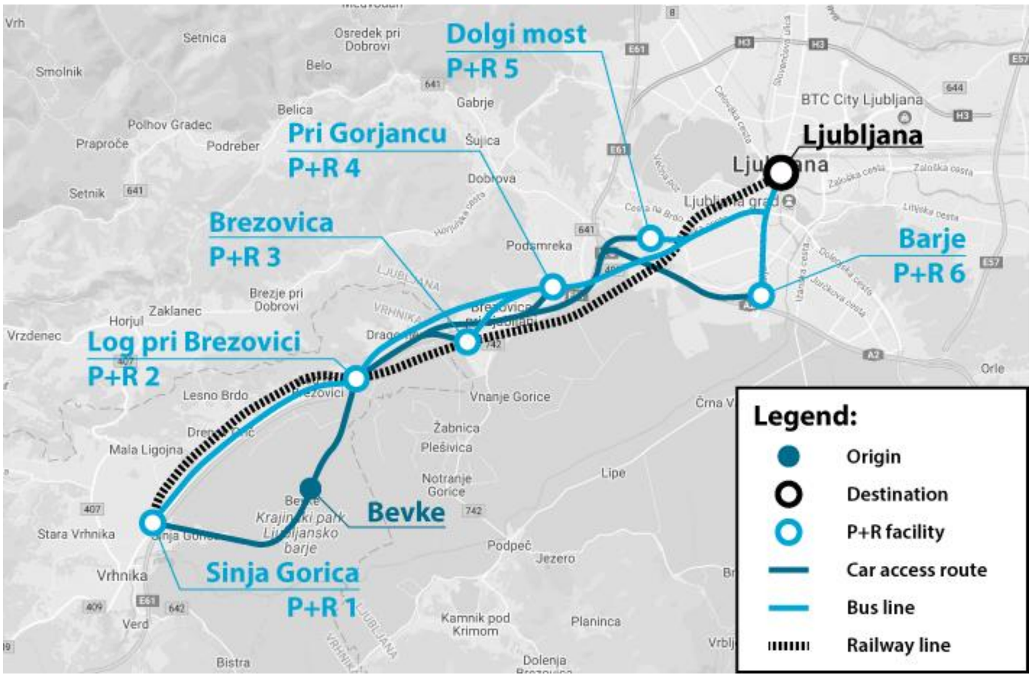


Figure 6.1. South west corridor with public transport routes and P+R facilities

6.1.3 Scenarios and attribute levels

Scenarios for the application of the model can roughly be divided into two chains of scenarios. The first follows the current plan of the Regional Development Agency of the Ljubljana Urban Regio (RDA LUR) (scenarios 1 and 2), where P+R 3 and 4 are to be constructed in the near future and focus is given on improving links to satellite P+R facilities (RDA LUR, 2014). As has been shown in the literature overview in subchapter 3.2.1, focusing on satellite P+R, although better for the city centre, does not lead to sustainable mobility, but instead often results in increased traffic and VMT. That is why in line with research, this study proposes a way to move towards a more sustainable mobility plan for the SW corridor. Firstly, building on the already existing plan, the scenarios look at what would happen if sustainable measures are implemented after the currently proposed P+R plan is already in operation (scenarios 3, 4 and 5). In addition to that, a second chain of scenarios (scenarios 6, 7, 8 and 9) looks at what impact the same policy measures would have, if the current plan of the RDA was not executed.

In both chains, the scenarios aiming for more sustainable mobility and with a focus on remote P+R facilities are the same, but in a different sequence and with different P+R lots included (the first chain also includes satellite P+R 3 and 4, while the second chain only considers four P+R facilities). These three scenarios are **Additional remote P+R** (scenarios 3 and 7), **Introducing rail services** (scenarios 4 and 9) and **Pricing policy** (scenarios 5 and 8). A complete overview of scenarios and the two chains can be seen in Figure 6.2. Each individual scenario is explained in more detail in the following paragraphs, with subchapter 6.1.4 giving a detailed explanation of how and where the characteristics of each P+R (travel times, costs etc.) were obtained.

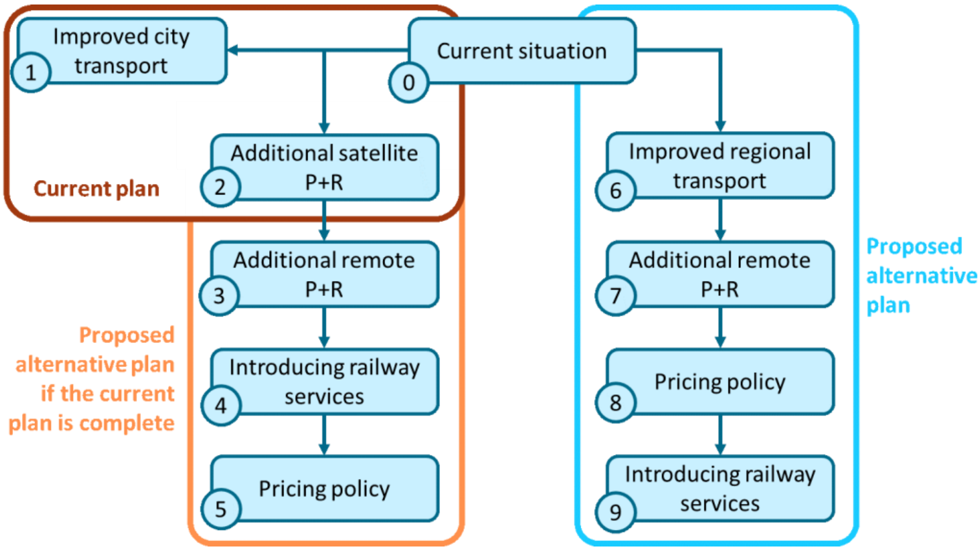


Figure 6.2. Scenarios for the model application

Starting with the current situation (**Scenario 0**), it consists of the three currently operating P+R facilities on the SW corridor: Sinja Gorica (1) Dolgi most (5) and Barje (6). The current access times, travel times, costs and frequencies are presented in Table 6.2.

Table 6.2. Characteristics of existing P+R facilities along the SW corridor (scenario 0)

	Car access time	Bus travel time	Price	Interval in AM peak	Mode
P+R 1	8 min	33 min	5,40 €	15 min	bus
P+R 5	18 min	15 min	1,20 €	6 min	bus
P+R 6	18 min	11 min	1,20 €	10 min	bus

Improved city transport (**Scenario 1**) shows what happens when accessibility of satellite P+R facilities is improved by increasing public transport frequency and/or reducing travel time by adding dedicated shuttle lines that run non-stop from the parking lot to the city centre. To show how this affects the choice probabilities, the interval of buses at P+R 6 is increased to every 5 min and the travel time is varied to see how this affects usage of all three P+R facilities.

The current plan calls for the construction of two additional satellite P+R lots (**Scenario 2**): P+R 3 and 4, located slightly upstream from P+R 5, acting as a secondary and tertiary P+R, once the closer one fills up. The characteristics of both parking lots are presented in Table 6.3.

Table 6.3. Characteristics of Pri Gorjancu and Brezovica parking lots in scenario 2

	Car access time	Bus travel time	Price	Interval in AM peak	Mode
P+R 3	10 min	23 min	1,20 €	15 min	bus
P+R 4	14 min	17 min	1,20 €	10 min	bus

That situation could then be remedied by opening an additional remote P+R lot (**Scenario 3 and 7**), located farther from the city of Ljubljana but still along the main regional public transport corridor. The characteristics of P+R 2 are presented in Table 6.4. This P+R will be added to the case study with different characteristics. To the current plan as scenario 3, the P+R is added with the existing frequency of public transport services, while in the proposed alternative plan, the P+R is added as scenario 7, after an increase in operating frequency has been instated, so that increased frequency is used here as well.

Table 6.4. Characteristics of Log pri Brezovici parking lot in scenarios 3 and 7

Scenario		Car access time	Bus travel time	Price	Interval in AM peak	Mode
3	**P+R 2**	5 min	24 min	4,60 €	15 min	bus
7	**P+R 2**	5 min	24 min	4,60 €	10 min	bus

(Re)introducing a railway line (or a similar form of high LOS public transport) on the SW corridor (**Scenarios 4 and 9**) between Ljubljana and Vrhnika could also be an improvement for the accessibility of remote P+R facilities along the corridor. As no plans have been put forward, the route of the discontinued narrow-gauge railway is used as a proxy for determining the travel times. Given the route characteristics, a tram-train type service is assumed. Three of the P+R lots would be located along this railway link. The changes to travel time, interval and mode are presented in Table 6.5.

Table 6.5. Changed characteristics of parking facilities due to the introduction of railway services in scenarios 4 and 9

	Train travel time	Interval in AM peak	Mode
P+R 1	25 min	15 min	train
P+R 2	16 min	15 min	train
P+R 3	11 min	15 min	train

A different pricing policy (**Scenario 5 and 8**) should also be considered as a possible measure for encouraging more sustainable mobility in the region. The price at P+R facilities covers daily parking and a return ticket (two single tickets) from the parking lot to the city centre (90 min of travel from the first tap-in). Satellite parking lots cost 1,20€, as much as a single ticket to the city centre, while remote parking lots cost either 4,60€ (P+R 2) or 5,40€ (P+R 1) which is the equivalent of a return ticket. To better stimulate the usage of remote P+R lots, the pricing policy could be reversed, meaning that remote P+R would cost as much as a single ticket, while using a satellite P+R would double in price (Table 6.6).

Table 6.6. Changed characteristics due to a new pricing policy in scenarios 5 and 8

	Price
P+R 1	2,70 €
P+R 2	2,30 €
P+R 3	2,40 €
P+R 4	2,40 €
P+R 5	2,40 €
P+R 6	2,40 €

Improving regional transport (**Scenario 6**) can be achieved through higher operating frequencies during peak times and/or lower travel times, achieved through various operational and infrastructural measures (express services, priority at traffic lights, bus lanes etc.). Scenario 6 therefore lowers the headway of the suburban line serving P+R to 10 min and varies the travel time.

6.1.4 Attribute level selection

This subchapter presents the argumentation for how the attribute levels for each of the P+R facilities were determined. Starting with the car access times to parking lots, this was done in a straightforward way with the help of Google maps, where the origin was Bevke and the destination the corresponding P+R facility. For the not yet built lots, an approximate location was used. The travel time was calculated for Wednesday, October 3rd 2018, with a departure time at 5:00 in the morning ('Google Maps', n.d.), meaning that a travel time with no road congestion is assumed.

Public transport travel times and frequencies were determined from the published autumn 2018 timetables of the bus operating company LPP ('Vozni redi', n.d.). Parking facilities 1, 2, 3 and 6 only have a single bus line (either urban or regional), while P+R 5 has two urban lines and P+R 4 one urban and two regional lines. For P+R 4, the urban line was chosen, as the travel time is not much longer, but it offers much better access to the city centre. For P+R 5, urban bus line 6 was chosen as it offers much faster travel time to the city centre and has higher frequencies. For travel times, two possible destination stops in the city centre were taken: stop Ajdovščina for urban bus lines and the Ljubljana bus station (main coach station) for regional bus lines and the tram-train. The frequencies of public transport services were also obtained from the timetable. For the lines running in regular intervals, the interval of arrivals to the city between 7:00 and 8:00 was selected. For lines with less regular intervals, the same time period was taken, and an average interval was obtained and rounded.

The prices for regional buses and use of P+R facilities were gathered from the published price list of the bus operator LPP (Ljubljanski potniški promet, n.d.).

The travel times for a possible train or tram-train service was determined from similar systems in other cities. In particular, the extensive network of tram-train services in and around the city of Mannheim in Germany was taken as a benchmark for travel times. The exact routing proposed in this research follows the path of the former narrow-gauge railway, that joins the existing railway in the town of Brezovica. From there to Ljubljana, the tram-train would use the heavy rail line with limited stops to allow for shorter travel times to the city centre.

6.2 Model selection and attribute analysis

Three different models will be used to determine travel behaviour in the selected case study: the traditional first-choice RUM model (E.1), the SBWMNL RUM model (E.3) and one of the SBWMNL μRRM model developed in this research. The specific μRRM model selected for the case study was the single-Λ and 2-μ mode (model E.7). As many of the choice sets do not contain exactly five alternatives, it was important to have flexibility, so a model with a single Λ value was needed. The decision to use the model with two rather than four scale parameters is due to the results of the BIC value, which was lower for the former and also because the scale parameter for the first best choice was the same in both E.7 and E.8. Given that the μBest scale parameter in the μRRM model was estimated at a value over ten, it indicates a utility maximisation decision rule for first-best choices, meaning that with respect to decision rule all three models are fully compensatory. As such, the differences in model outcomes will mostly be due to different parameter estimates rather than decision rule.

Given that attribute parameter values will be responsible for most differences between the models, the analysis of said parameters and their ratios is crucial. Although one of the models is a μRRM model and thus the parameter ratios do not directly translate to values of time, they can be seen as a such given that the decision rule is utility maximisation and hence implies fully compensatory behaviour. The parameter ratios of all three models are presented in Figure 6.3. The figure shows the substitution rate between two parameters. For example, respondents are willing to pay 0,27 € (according to SBWMNL μRRM) for a one-minute reduction in travel time by car. Or, they are willing to travel 4:07 min longer by PT if it is a train rather than a bus (according to the SBWMNL RUM model).

Starting with the ratio of time and mode parameters with the cost parameter, the differences between the models do not seem to be too great. For all three time-related parameters (TT by car, TT by PT and headway), the RUM model has the highest WtP, followed by the SBWMNL μRRM, with the SBWMNL RUM having the lowest. A high parameter ratio with respect to cost means that commuters are willing to pay more if it gets them to their destination faster and vice-versa for a lower ratio. When looking at the ratio with respect to mode (using a train rather than a bus), the SBWMNL μRRM is again in the middle, with the SBWMNL RUM having the highest ratio this time. Here, a high ratio can be translated into how much of their travel time are commuters willing to 'sacrifice' to travel with a better mode of PT.

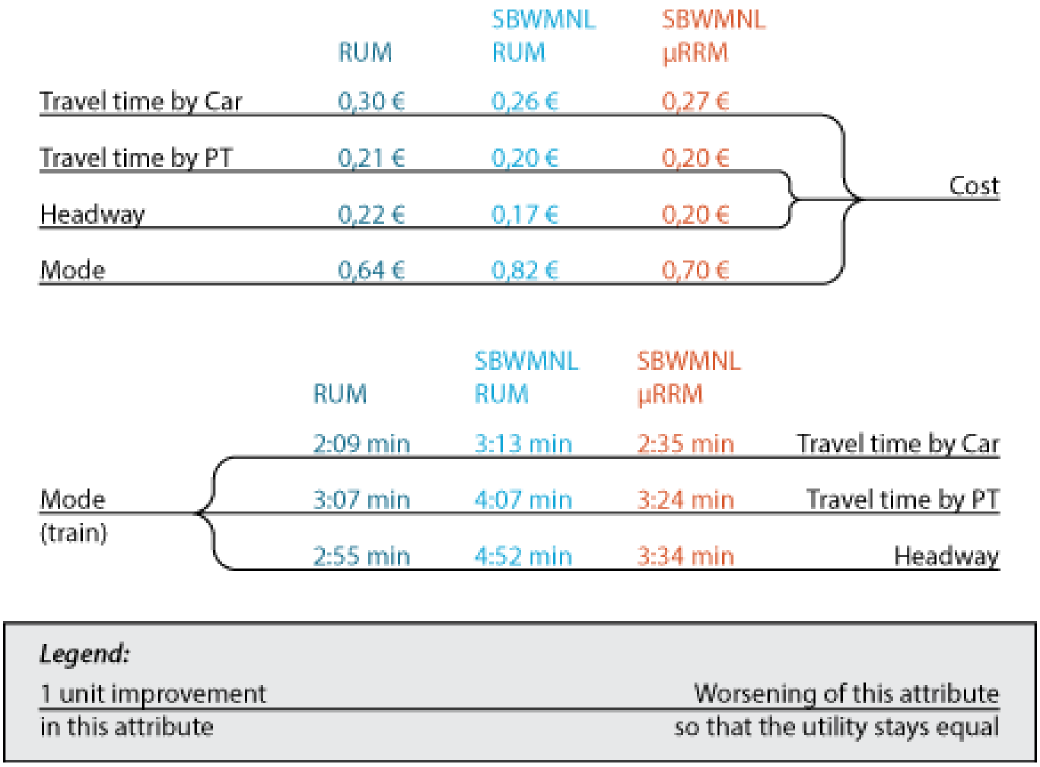


Figure 6.3. Parameter ratios / Marginal rate of substitution

Given the parameter values, the SBWMNL RUM model can be interpreted as prioritising alternatives with lower prices, rather than lower travel times, when compared to the other two models, as the ratios indicate that respondents are willing to pay less for travel time improvements. This means that the SBWMNL RUM model will give higher choice probabilities to cheaper alternatives with longer travel times and vice-versa. However, the same model also values an improvement in mode from bus to train more than the other two models and is therefore willing to 'sacrifice' more travel time or cost if an alternative by train is available. The SBWMNL RUM model therefore puts less value on any travel time aspect (either car, PT or headway time) and more value (compared to the other two models) to a better mode of PT and lower cost.

Considering the three time-related estimates of travel time by car, by public transport and public transport service headway, the corresponding parameter ratios can be seen in Figure 6.4. The figure is interpreted by looking at the direction of the arrows and the values next to the arrows. For example: a 1-minute change in travel time by car is compensated by 1:27 minutes of travel time by public transport (opposite direction) for the utility to remain the same (choice probability), according to the RUM model. Similar does a 1-minute change in PT travel time require a 1:03 minute change in headways according to the SBWMNL µRRM model. For the relationship in the other direction, an inverse value is needed. The figure again shows that a change in car travel time requires a larger change of travel time by public transport to balance it out, indicating a higher valuation of in-car travel time. Headway is also less valuable in all cases compared to car travel time. What is interesting to observe is the headway and travel time by PT ratio. For the RUM model, a minute of headway is more valuable than a minute of PT travel time, while the opposite is true for both SBWMNL models. This means the RUM model puts more emphasis on the frequency, while the SBWMNL models are more concerned with the actual travel time.

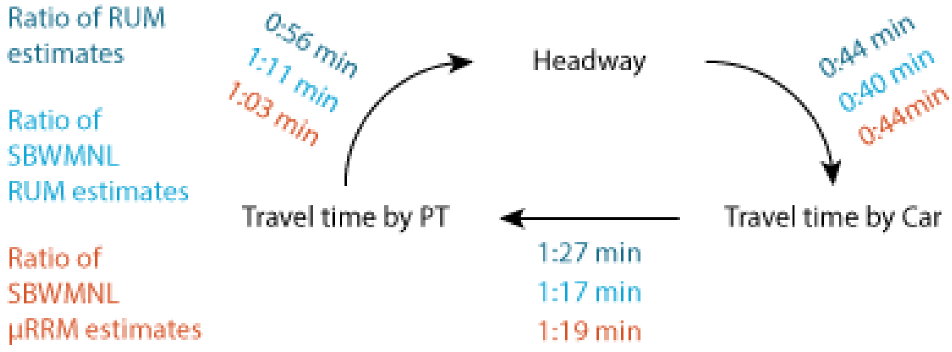


Figure 6.4. Parameter ratios of car and PT travel time and PT headway

6.3 Model application outcomes

This chapter reviews and interprets the outcomes of scenarios constructed in the previous subchapter. The results are interpreted from the perspective of different models and their respective performance, as well as looking at the travel behaviour outcomes and policy measure implications. Based on these outcomes and interpretations, a conclusion regarding how the policy plan on P+R facilities should be adapted (if at all), based on the vision of mobility for the municipality and region.

The subchapter first looks at the scenarios representing the current plan of the RDA, followed by the proposed extension to the existing plan and thirdly covers the proposed alternative plan for the CSSR and Ljubljana. The subchapter concludes with an overview of policy implications pf different measures.

6.3.1 Current plan and proposed evolution

The outcomes of the current plan scenarios and proposed extensions to this plan can be seen in Figure 6.7. Travel behaviour in the current scenario (Scenario 0) is a clear indication that Bevke, as an origin, has no real remote P+R lot for commuters to use on their way to Ljubljana. The only remote P+R currently in operation is located further away from the city than the origin and it would most likely only be used by a handful of people, who prefer to avoid congestion around Ljubljana entirely. All three models (as seen in Figure 6.7) predict a market share between three and ten percent, with the SBWMNL μRRM being the most pessimistic with respect to P+R 1. When comparing the two existing satellite parking lots, all models point to a slight advantage of P+R 5, that despite a longer travel time by PT (and total travel time as well), has a superior service frequency, operating every six minutes, compared to every twelve at P+R 6, giving it a competitive advantage.

What is often present in traditional P+R schemes is increasing the accessibility of satellite parking facilities to reduce congestion in inner cities. This has been proven to reduce traffic in urban cores, but also induces more trips to the edge of the city. Figure 6.5 depicts the resulting market shares of all three parking facilities by all three models, when travel time by bus from P+R 6 is varied, and the headway reduced from twelve minutes to five. It shows that even keeping with the same travel time, the more than doubled frequency results in P+R 6 getting a majority stake in all three models. Travel times could realistically be lowered down to five minutes, if bus lanes, priority at traffic lights and an express service would be introduced. With such a high level of service, P+R 6 would have a market share of 60%-70%, depending on the model, with the SBWMNL μRRM predicting the largest increase in market share per minute saved. The point at which P+R 5 and 6 would have roughly the same share of the commuter market is around the 12-to-13-minute travel time marker, which is logical as the travel time from P+R 5 is 15 min by bus, but the car access time is two minutes lower than for P+R 6. P+R 1 on the other hand sees little to no change in market share, as bus travel times from P+R 6 higher than the current are unrealistic.

When comparing the different model predictions, it is interesting to note that for P+R 6 predictions are equal (intersect) around a travel time of 18 min. Interestingly, this is the car access time from the origin to P+R 6. From that point the SBWMNL μRRM is either the most optimistic, if the travel time is decreasing, or the most pessimistic with travel time increasing, while the SBWMNL RUM is sees the smallest change in market share for higher or lower travel time.

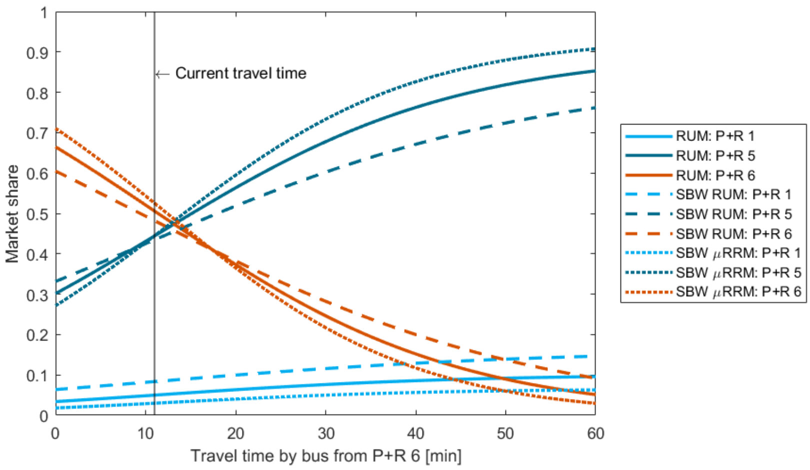


Figure 6.5. Variation of travel time by bus to P+R 6 with increased frequency

Following the current P+R implementation plan in the CSSR, two additional P+R facilities (3 and 4) are added. With their addition to the SW corridor, the market share is redistributed and the four satellite P+R each have a similar stake, with P+R 5 and 6 retaining their top two positions, followed by P+R 4 and then P+R 3. Particularly the parking lots 6 and 4 perform very similarly to each other. P+R 6 has a slightly better frequency and travel time by bus, while P+R has an access time by car that is a third shorter. The SBWMNL RUM model is still the one that predicts a lower market share for P+R 5 and giving more to other facilities, in particular P+R 1 and 3, which have a combined market share that is five percentage points higher compared to the other two models.

Increasing the service quality of satellite P+R would likely attract new users (in addition to existing users form other P+R facilities), which would result in more car traffic and worsening of congestion (Meek et al., 2011). To address this issue, a new plan based on the link-and-ride concept is proposed here. It is investigated both as a continuation of the current regional plan, should it be completed or as a standalone plan from the current situation (in subchapter 6.3.2).

As the population along the SW corridor has no remote P+R facility option, this could be addressed by building one that would cater to the inhabitants of settlements located between Vrhnika and Ljubljana. P+R 1, although a remote parking facility, is better suited for those living in Vrhnika, its immediate vicinity and farther away from the city, and not so much for those living closer to Ljubljana. In line with this, P+R 2 is proposed near Log pri Brezovici, some six kilometres closer to Ljubljana than P+R 1, yet still almost four kilometres farther than P+R 3 or almost eight kilometres farther than P+R 5. This parking facility can act as a true remote P+R, unlike P+R 3 and 4, whose main role is acting as a backup for P+R 5.

The effect of adding a sixth P+R facility as a remote parking lot can be seen in scenario 3. P+R 2 would have a market share of around 10%, positioning itself as relatively unimportant on the SW corridor, but still reducing the pressure on the satellite parking facilities and giving the residents along the corridor a true alternative to driving to the city edge. Regarding the different model predictions, the SBWMNL RUM continues to be the most optimistic for remote facilities and the SBWMNL μRRM being the least, as the former predicted a market share of over 10% for P+R 2, while the latter only around 7%.

A (re)introduction of railway services along the SW corridor would also be an option to improve accessibility and attract additional users, both solely PT users and P+R users. Given the route of the railway, it would have a significant impact on P+R 1, 2 and 3, with the most notable improvement on the latter. The travel time from P+R 3 would reduce by more than halve and the additional bonus of having a railway connection means that this facility would become the most attractive along the entire SW corridor. While P+R 1 and 2 would see less improvement, as the travel time reduction is lower, their market share would still increase by roughly 25%, with P+R 2 having a similar share to P+R 4 and 6. With respect to the differences in model predictions, the SBWMNL μRRM seems to be most sensitive to change, as the changes are most significant both in the increase and decrease of market share. The RUM model emphasises more the growth of P+R 3 compared to the SBWMNL RUM, which is most conservative when it comes to change.

The final policy measure proposed and tested within this study is a new pricing policy, where using remote P+R would be encouraged, while using satellite P+R is discouraged, by halving and doubling the prices respectively. This single measure has had the largest impact on parking facility choice in terms of promoting the use of remote P+R. Market shares of P+R 1 and 2 doubled or even tripled in all three models, with again the SBWMNL μRRM emphasising the growth the most: by almost three-fold from the previous scenario for P+R 1, and 2,5 times for P+R 2, which would become the most attractive P+R facility on the SW corridor. The SBWMNL RUM also remains the least sensitive to change and the RUM model staying as a middle ground between both SBWMNL models. Combined, both remote P+R would have a market share of 30%-40%, with P+R 3 being a hybrid also adding over 20%. The three remaining parking lots would lose about 30% of their market share each and holding onto some 30%-40% of the market. P+R 3 would see its market share reduced due to the new pricing policy but would still be more attractive than before the railway introduction, meaning that the benefits of a railway connection at P+R 3 cannot be balanced by the proposed pricing policy.

6.3.2 Proposed alternative plan

Results of the proposed alternative P+R development plan are shown in Figure 6.8. The sustainability-focused policy measures (Scenarios 3, 4 and 5) implemented on top of the completed current plan of the RDA could also be implemented on their own, before completing the official plan, possibly reducing some of the negative effects of focusing too much on satellite P+R facilities.

Similar to what has been shown with varying the travel time of buses linking P+R 6, Figure 6.6 shows the market share as a result of travel time changes of the bus line connecting P+R 1, with an already increased frequency from four to six buses per hour during peak times. The initial market share of P+R 1 is so small that the increase in frequency alone adds little to its market share; only around two percentage points. But should the travel time also be reduced, the stake could be higher and potentially reach around 20% if larger investments and operational changes are made, like priorities at traffic lights, express services and physically separated bus lanes. Comparing the different model performances, a similar observation can be made as in Figure 6.5, but in this case less clearly: the different models produce the same market share at around the value of the car access time, which in the case of P+R 1 is eight minutes. Unlike in Figure 6.5, all three model predictions do not cross in the same point, but both RUM models intersect around the 10-minute mark, while both SBWMNL models cross slightly below the 8-min travel time. Again, the SBWMNL μRRM model is the most pessimistic at higher travel times, while the SBWMNL RUM is the least pessimistic. The predictions of all three models can also be investigated for the two satellite parking lots. P+R 5 is valued most by the SBW μRRM model, while the prediction for it is lowest by the SBWMNL RUM model. P+R 6 on the other hand sees the SBWMNL RUM model predicting it the highest market share.

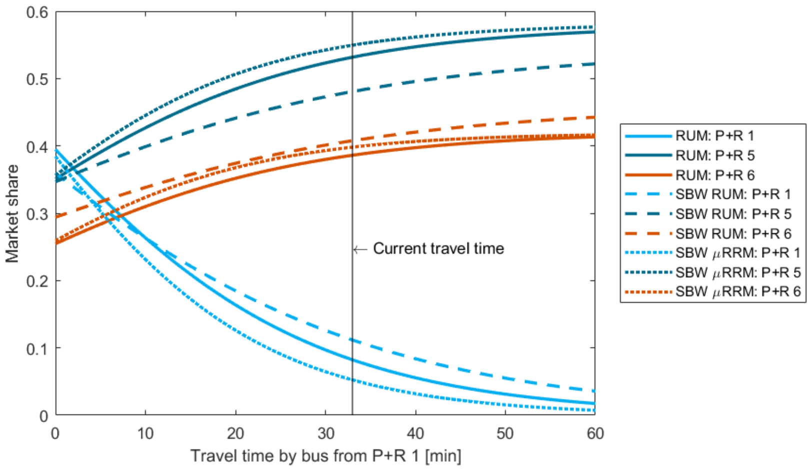


Figure 6.6. Variation of travel time by bus to P+R 1 with increased frequency

Increasing the frequency alone, without travel time improvements would already mean an improvement in the market share of P+R 1, most notably in SBWMNL μRRM. Adding P+R 2 without constructing P+R 3 and 4 would mean a market share for remote parking facilities of 25%, rather than 15%-20% if the two extra P+R were built. The number of users at each individual P+R may not be different, as the current plan, with the proposed extension would have a higher combined capacity of all P+R, but these models only consider market share, rather than an absolute number of users. Just as in the current plan, P+R 5 would retain its market dominance even with the introduction of P+R 2, with P+R 6 also holding on to second place in all three models.

In the proposed alternative development plan of P+R facilities, the pricing policy (scenario 8) is implemented before the railway is reintroduced (scenario 9). It shows that a pricing policy has a higher impact on the market share of remote P+R facilities than adding a railway connection. The price change is quite significant and the reduced disutility of remote P+R and extra disutility of satellite P+R has a more radical impact on parking lot market share than a railway connection, which only reduces the disutility of remote P+R, while leaving the utility of satellite P+R unchanged. This may have led, to a different result, should the decision rule have been RRM rather than RUM. In scenario 8, the most significant increase for remote P+R is again predicted by the SBWMNL μRRM, increasing 2,6-times for P+R 1 and 2,3-times for P+R 2. The lowest relative increase is predicted by the SBWMNL RUM, which forecasts less than 2-times the market share for both. However, the same model had the highest share out of all three models for both P+R 1 and 2 before the pricing measure was tested.

Adding a railway on top of the pricing policy has still helped the performance of remote P+R, but not nearly to the same extent as the pricing policy. And having been implemented afterwards, the increase was lower than in scenario 4, where it was tested beforehand. In scenario 9, most remote P+R only saw an increase of around 20%, which is still good, but nowhere near the 200% increase from the pricing policy. In the end, the remote P+R have a majority of the market with slightly less than 60%. Of the two satellite facilities, P+R 5 remains above P+R 6, while P+R 1 is last but only by a small margin. This is logical, as it is in the opposite travel direction when starting from Bevke and has the longest combined travel time and highest costs, even after the pricing policy implementation.

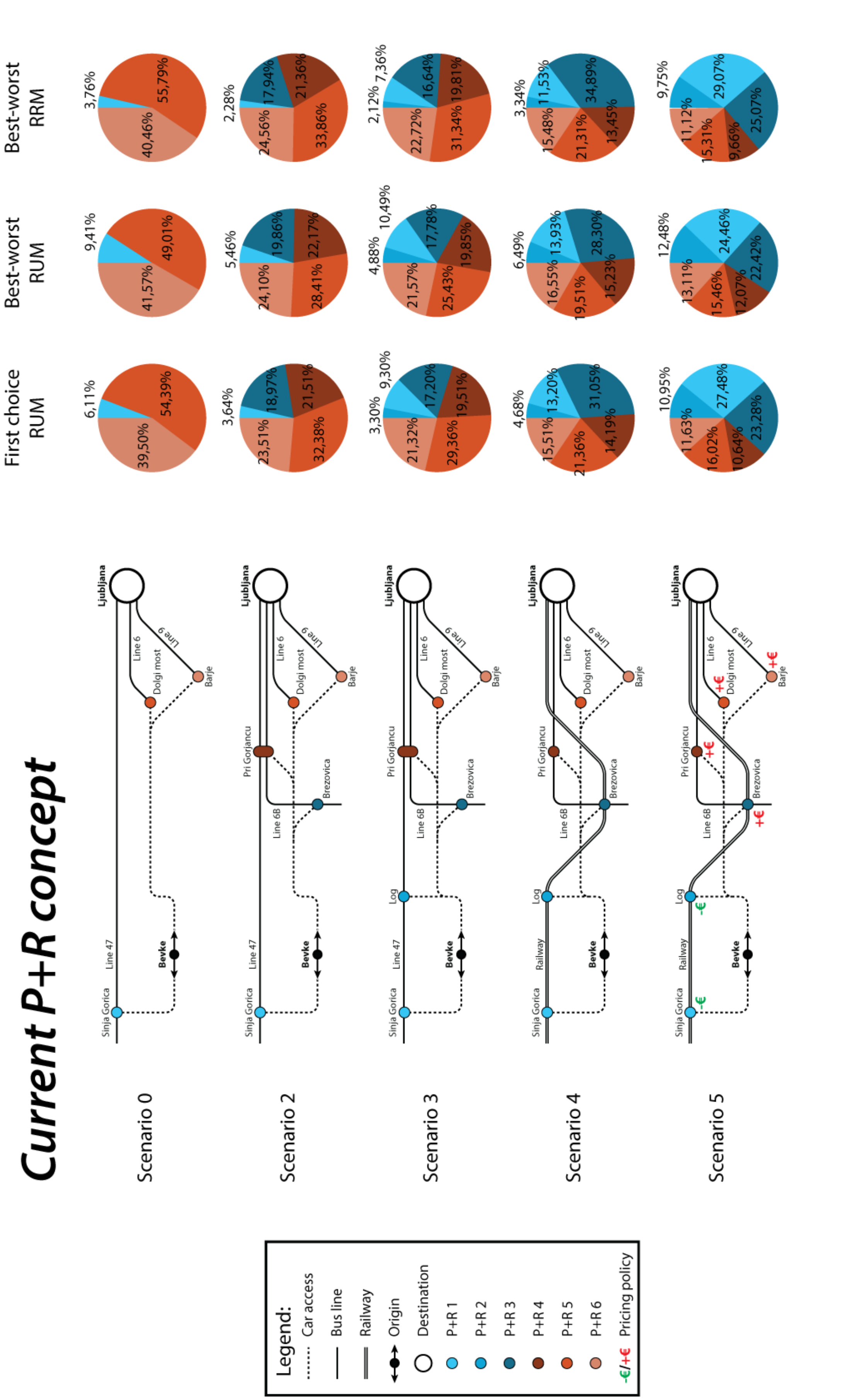


Figure 6.7. Results from the chain of scenarios covering the currently existing plans in the Ljubljana Urban Region

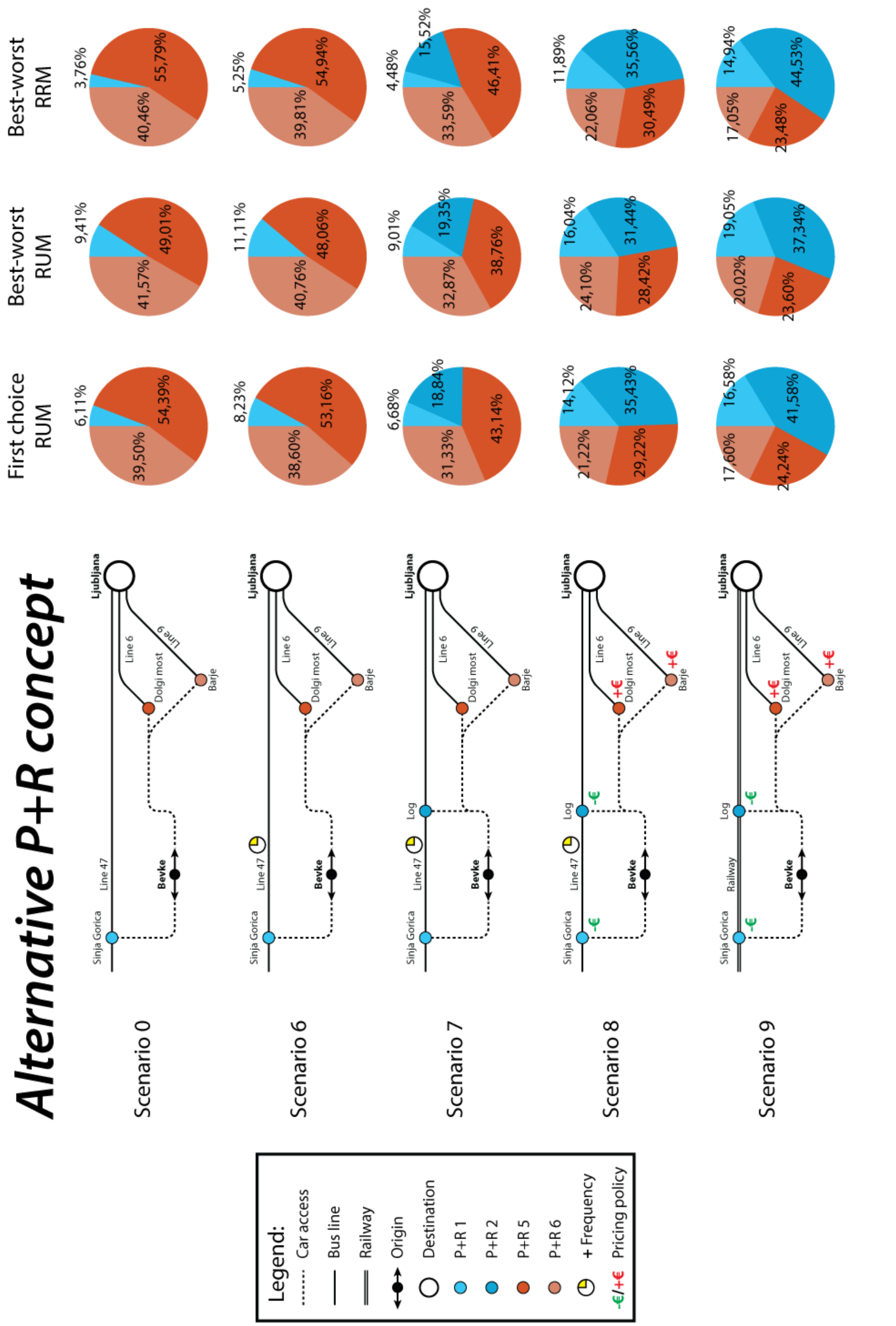


Figure 6.8. Results from the chain of scenarios covering the alternative concept of P+R, proposed in this research

6.4 Conclusion, policy implications and recommendations

The analysis of accepted and proposed scenarios carried out in this research provides valuable insight into the workings of commuter behaviour following the implementation of certain policy measures on the SW corridor and can also be used to predict behaviour on other corridors in the CSSR. What can immediately be said about the current plan of the RDA LUR is that it will almost certainly attract more commuters to use P+R instead of the modes they are currently utilising. From research, we can assume that these new users will come both from car drivers and PT users. This aspect of travel behaviour was not, however, covered within this research, as it focused solely on the market share among P+R facilities.

By introducing new satellite parking facilities close to the city along highway corridors, usage is likely to increase and cause additional VMT and congestion near the P+R. The analysis here showed a decrease in market share of the remote parking facility, which does not really cater well to the sizeable population along the corridor, as commuters must travel in the opposite direction to use it. Because of that, people are forced to use satellite P+R and travel longer distances by car if they wish to use a P+R facility. If a plan would go forward to increase the level of service at a satellite P+R, it could have detrimental effects on all other parking facilities. The experiment showed that a PT connection operating with a very high frequency could capture up to two thirds of the market, resulting in high demand for that particular facility and without doubt inducing even more car travel. Although a plan like this can prove very beneficial for urban cores by pushing daily commuter car traffic to the outskirts, the overall situation in the region does not necessarily improve and may even worsen from the perspective of sustainable mobility.

While the introduction of new satellite P+R facilities had a minimal impact on remote P+R market share, the added capacity of all satellite P+R facilities combined would most likely also induce additional car trips. Should those be exclusively from commuter who previously drove downtown, such a measure would prove beneficial, but research has proven that many users would be former PT users or non-commuters.

Focusing on the link-and-ride concept of P+R, an additional parking facility is suggested along the SW corridor, further away from Ljubljana, that is designed to better cater to the needs of locals. Potentially every bus/train stop along that corridor could have a small P+R facility (with around ten parking spaces) to allow the nearby population, living too far from the station to walk, to drive there by car. As many parts of the corridor and region are hilly, cycling is also often not a viable alternative, even if the necessary infrastructure would be in place. The location in Log pri Brezovici was selected as it may have the highest potential along the corridor. It was placed at the first bus stop following the junction from the village of Bevke, which has a notable population and extremely limited public transport services.

Simply adding P+R 2 already showed that it is much more attractive to users compared to P+R 1, yet still not being as popular as the other satellite P+R, mostly due to the much higher cost of parking and public transport. To increase the use of P+R 2 and similar parking lots along other corridors, a stimulating pricing policy was found to have the biggest influence. As mentioned, the current prices are made to stimulate the use of satellite P+R, possibly as an alternative to parking in the urban core. This is a good incentive, but it also results in reduced attractiveness of remote P+R facilities. To allow for a slower transition compared to the modelled scenario, the price of using remote parking lots could first be lowered and the price of satellite P+R increased later, perhaps together with an increase of parking prices in the city centre as well, to avoid traffic returning to the city. Compared with other measures, a changed pricing policy is also relatively easy to implement in financial and infrastructural terms. There might however be strong political opposition towards a measure that increases parking prices.

Other factors found to influence the attractiveness of remote P+R are various improvements in the level of service. While railway services are popular and have a large carrying capacity, the investment costs are significant, so the business case for such a measure needs to be strong. Instead of a railway, funds could be better used for new buses to increase the operating frequency and/or for infrastructural changes such as bus lanes, priorities at traffic lights etc. Figure 6.3 shows that the same benefit a train provides can also be gained from increasing the frequency by five minutes (assuming travel time does not change). Although the headway parameter was not tested for non-linearity (even though it is almost definitely not linear), this could be done by reducing the headway from 15 minutes to 12, 10 or even 8 minutes in the morning peak. As trains have a much higher capacity, the headways to transport the same number of passengers are lower, meaning that the added utility of a railway service is countered by a disutility of a lower operating frequency.

An exception to the pricing measure being superior to introducing railway services is P+R 3, which sees a very big improvement in service characteristics thanks to the new railway connection. So much so that even the pricing policy could not lower its market share back to the initial level. This shows that if done right, a new/improved public transport service can still have a higher impact on the attractiveness than a pricing policy. If the improvement would have occurred at a remote P+R facility, the pricing policy would have boosted the attractiveness of the parking lot even more. That was not what resulted from the study however, as for the parking lots that had both policies working in their favour (P+R 1 and 2), the improvement due to the railway was marginal and thus always offered much lower growth rates in market shares. The models did however show that both the railway and pricing policy have a bigger impact on the corridor on their own rather than on top of one another in relative terms. Take the case of P+R 2 in the SBWMNL μRRM model: Introducing a railway increases its share by 57% and then adding the pricing policy measure on top adds a further 152%. If the order is reversed however, the pricing policy initially brings an increase of 181%, with the railway on top adding only 41%. The combined result is the same in both cases: an increase of 295%. The same can be observed for both remote P+R in all three models.

The models themselves made mostly similar predictions throughout the modelled scenarios, with respect to the order of alternatives. What was notable was that the SBWMNL RUM model started out with the highest market share prediction for remote P+R facilities compared to the other two models. When policy measures aimed at improving attractiveness of these facilities, the SBWMNL RUM model predicted the smallest growth, very explicitly shown in the travel time sensitivity analyses. The SBWMNL μRRM on the other hand predicted the most drastic changes in market shares, showing the largest in-/de-creases in market share if travel time to remote P+R facilities de-/in-creased. The first-choice RUM model stayed in the middle in most cases, as a sort of middle ground between both SBWMNL models.

This research was done on a simplified model, considering a single corridor with a single origin point. It also looked only at the market shares and not the full demand in the number of users. This could not have been done in this research, as the model was based solely on P+R usage and thus including car commuters and PT commuters would have required a different survey to be constructed. In addition, the number of commuters was not considered and so the induced demand with changing P+R characteristics can only be assumed based on previous research. This has shown that better satellite parking facilities do attract new users from public transport and those who previously would not have travelled, but the scale of this differs among researches and an exact effect can therefore not be assumed. The model application has still added valuable insight into travel behaviour with respect to P+R facilities and how different policy measures impact commuter behaviour.


INTERCONTINENTAL
310-006

7 Conclusion

This research was undertaken with the goal of combining sequential best worst discrete choice experiments (a more efficient data gathering technique) and estimating the obtained data with a μRRM model, which has higher behavioural realism with respect to decision rule utilisation. It was hypothesised that best choices would be selected using a utility maximisation decision rule, while the worst alternatives would be chosen by utilising the principles of regret minimisation. An SBWDCE was setup to gather the necessary data and the survey was carefully assembled with the principle workings of potential choice models in mind. The method was applied in the field of transportation and a Bayesian efficient design was generated and tested using Monte Carlo simulation, to confirm that parameters can be obtained from the data. Several different models were developed and estimated using the gathered data and their results were compared to those of established choice models using an array of statistical tests. To explore the performance of the models on a real-world case, a simple park-and-ride facility choice case study was designed for the Central Slovenia Statistical Region.

This chapter summarises the outcomes, major findings as well as limitations and recommendations for future research. Subchapter 7.1 answers the questions posed at the start of the research. Subchapter 7.2 then discusses the model outcomes and links the findings to theory and literature. Subchapter 7.3 lists the limitations of this research and its potential shortcomings due to a limited time and budget. Finally, recommendations for future research, as well as policy recommendations for transport policy-makers in the CSSR and elsewhere are given in subchapter 7.4.

7.1 Research conclusion

This subchapter uses the obtained modelling results from the developed SBWMNL μRRM, other established models and the results of the model application to answer the research questions posed in subchapter 1.2. The questions are answered one by one, followed by an evaluation and the acceptance/rejection of the stated hypothesis.

RQ1: *Can a μRRM model extract accurate taste and scale parameters from data that was obtained in a sequential best-worst discrete choice experiment setup?*

The fist research question has a short and straightforward answer: yes. To test if the developed SBWMNL μRRM model can properly extract taste and scale parameters, synthetic discrete choices were generated using Monte Carlo simulation and those observations were then used to re-estimate the parameters. The taste parameter estimates of the single class models were tested against the true values and all scored a high p-value (ranging from 0,30 to 1,00), meaning that the estimates were not significantly different from the true values used in the data generation process. Scale parameters estimates were not as similar, but their evaluation needs to be based on the interpretation of their value. The parameters for best choices had to be fixed in order for the model to converge, as all achieved a higher value than ten. Although numerically different, values above ten all indicate a fully compensatory behaviour and the differences are minimal. Scale parameters for worst choices were also higher than the true value and although their value is still close to non-compensatory, it is not as similar as with the best-choice scale parameters.

RQ2: *To what extent does accounting for intra-person decision rule variability between making best and worst choices improve the model fit (if at all) when compared to already established discrete choice models?*

To account for intra-person variability of decision rules, several different models were put forward, each accounting for the decision rule variability and choice set size variability in a different way. Nevertheless, all five models that were put forward did achieve a better model fit. Although the differences were not great, between 24 and 60 log-likelihood points, translating to 0,0039 or 0,0096 p.p. of rho-squared, the Ben-Akiva and Swait test results proved that the probability of the model not accounting for different decision rules being the true population model is virtually zero when compared to all five developed models. The largest model fit improvement was achieved by model E.6 in which four (three) scale parameters and three choice set size constants were estimated, the most out of any model. This model achieved a rho-squared of 0,1397, compared to 0,1301 of the E.3 model (the RUM-only model, not accounting for decision rule variability).

RQ3: *How significant is the presence of inter-personal variability of decision rule utilisation in the setting of best-worst discrete choice experiments?*

Testing of inter-personal variability (differences between people) was done with the help of latent class models. Unlike with the single class models, the LC models did not result in such a clear-cut answer. By implying the decision rule combinations of two latent classes, the resulting model fit was much higher compared to the single class models, reaching a rho-squared of around 0,2. The highest was achieved by model LC.2 (RUM-RUM decision rules for one class and RUM-P-RRM for the other). Estimating the decision rule of both classes, one class had a RUM for best choices and a semi- to non-compensatory decision rule for worst choices (μ=0,72), while the other class had both best and worst choice scale parameter around a value of two, indicating semi- to fully compensatory behaviour. What all the models had in common however was - regardless of the decision rules - one class with much higher parameter ratios compared to the other, indicating this may have been the main driving force of class allocation. Despite models with RUM decision rules for best choices performing better, implying a P-RRM decision for best choices resulted in minimal change in class size, further indicating that class allocation is not driven by decision rules but rather taste parameters.

Including socio-demographic characteristics of respondents did not help in class allocation as all four (regret scale, age, gender, income) were insignificant and while they produced better model fits, this was due to the respondents who did report their socio-demographics being removed, and thus their choice observations as well. Given that LC models rely on personal characteristics and not discrete choice observations, SBWDCEs do not offer the same benefit of allowing model estimation will smaller sample sizes.

RQ4: *In what way does accounting for different decision rules in a best-worst experimental setting differ in predicting (future) market share when compared to already established discrete choice models?*

The developed model that accounted for decision rule variability did produce different results when compared to the first-choice only model and the SBWMNL RUM model. However, given that for best choices RUM was found to be the decision rule used in the best-worst observations, the differences in market share predictions were only due to different taste parameters. Nevertheless, differences in model predictions were still found, with the SBWMNL μRRM model being the least conservative when a certain parameter was varied. With respect to the market shares, the model accounting for decision rule variation was predicting the highest share to satellite P+R and lowest to remote P+R. The SBWMNL RUM model was the opposite, giving the most favourable predictions to remote parking facilities, while the first-choice only model was in the middle.

Hypothesis: *In the context of best worst choice experiments, people choose in line with RUM when selecting the best alternative and in line with (P-)RRM when choosing the worst alternative.*

In the estimated models, the decision rules are obtained from the value of the scale parameter. In all single class models, the scale parameter for selecting the best alternatives has taken a value of ten or higher, indicating strong compensatory behaviour of the respondents. In many cases, the value was much higher than ten and the model did not converge, so the scale parameter had to be fixed to ten. With respect to the selection of worst alternatives, the scale parameter settled at a value of 0,43 (or 0,36 in the case of models E.7 and E.8), representing behaviour that is somewhere between semi-compensatory and non-compensatory. This means that performing better on an attribute gives an alternative a very small benefit, compared to a worse performance of the same scale, which produces much more regret. Based on that, the notion of worst choices being made in a completely non-compensatory manner (according to image theory) cannot be upheld.

The case of second-best choices is an interesting finding to consider and will also be discussed further in subchapter 7.2. The model outcomes show that the scale parameter switched from 0,5 (semi- to non-compensatory) to over ten (fully compensatory) when the choice set size variation specification was altered from multiple constants to a single constant.

7.2 Discussion

Having summarised the results in the previous subchapter, a discussion of those results is presented here. The subchapter starts by looking into the reasoning behind consumers' use of different decision rules for making best and worst choices. In the second part, the appropriateness of such a survey method is discussed with respect to the field of transportation. The third section looks at the different ways of accounting for choice set size variation and the benefits and drawbacks of either option.

7.2.1 Discussion of decision rule utilisation

The model estimation results have proven the hypothesis that people tend to behave in a compensatory manner when selecting the best alternative and in a non-compensatory manner (to a large extent) when selecting the worst. As theorised in subchapter 1.1, such behaviour is expected within the decision-making concept of image theory. Beach & Mitchell (1987) have stated that consumers use a two-stage process to facilitate their decision-making: first exclude the worst alternatives from the choice set in a non-compensatory manner and then select the best in the reduced choice set using a compensatory decision process. Ordóñez et al. (1999) brought the exclusion stage even closer to the definition of the P-RRM model formulation, by stating that for each alternative, consumers evaluate them based on a subjective expectation and sum up all the violation an alternative makes compared to the expectation. If the total sum of violations is above a subjective threshold, the alternative is rejected.

This raises the question of why this was the expectation, when numerous studies (many of which were summaries by Chorus et al. (2014)) have found that semi- and non-compensatory models result in a better model fit in traditional first-best choice surveys. The answer to this may be hidden in the way the choice tasks were presented to the respondents. Several studies have shown that respondents can be influenced in their decision-making with the way the instructions and/or question are phrased (Simonson, 1992) (Yu et al., 2015). In this research, respondents were specifically tasked with selecting the best and the worst choices. While still being told to select based on their own preferences and not what is generally accepted to be best, the given instruction on how to pick alternatives may have influenced the decision-making process. In traditional DCEs, respondents are simply asked to select the alternative they would most likely choose. That way they are given more freedom to select an alternative in whichever way they want. Beach & Mitchell (1987) have shown that the main driver in decision-making is compatibility, whereas profitability is a backup, when multiple alternatives are similarly good. They also state that consumers quickly settle on what is acceptable and do not care so much about maximising their 'utility'. Meloy & Russo (2004) provide an alternative explanation, saying that people prefer to choose by selection, not rejection. They also state that compensatory behaviour may even be used in the first stage to narrow down the choice set, depending on the complexity of the choice task and the topic. A more subjective, personal decision may result in narrowing by inclusion rather than exclusion. The decision process was also found to be influenced by the presentation of alternatives: a negative presentation is more likely to result in choice by exclusion, whereas a positive valence is more likely to induce alternative selection by selection (Meloy & Russo, 2004). These propositions prove that when no specific instructions on how to make a choice are given, people will make decisions in different ways and it can be influenced by many things. In this research, tasking the respondents to select the best and then the worst (and then best and worst again) may have contributed to them following image theory more closely.

A different view on utilising regret avoidance decision-making strategies is the inaction effect (Zeelenberg, van den Bos, van Dijk, & Pieters, 2002). Consumers are typically tasked with everyday decision on which option they will carry out. Transportation is a good example of this, as the options are usually different routes, modes, departure times etc. Asking respondents to select the worst alternative in a choice set is comparable to asking which they would least like to execute. This is what Zeelenberg et al. (2002) call the inaction effect: failing to act in a situation where action is expected induces more regret in respondents, providing an explanation why an a regret minimisation decision rule was prevalent for selecting worst choices.

The correctness of image theory may also be supported by the outcomes of the LC models. Although differences among models were found they seemed to be driven by different taste parameters rather than scale parameters. Models assuming a RUM decision rule for best choices performed best whereas for worst choices, the difference between models implying RUM or P-RRM was minimal. As the worst choice scale parameter was between RUM and P-RRM, proven also in the single class implied decision rule models (models I.1 to I.4), switching the implied decision rule did not result in a drastic change in model fit. Given these findings, the absence of different decision rule combinations may be an indication that, at least within our sample, image theory is consistently applied by all respondents.

Image theory however, does not explain why the second-best choice was estimated to have been chosen using RRM. Although the difference in final log-likelihood when implying RUM decision-making was rather small (less than ten log-likelihood points), this may still be a violation of image theory. One possible explanation is given by Inman & Zeelenberg (2002), who state that experiencing regret is dependent on whether a consumer decides to remain with a certain product or service, or switch to a new one. A bad previous experience will result in more regret if the consumer remains with the same service and vice-versa. Such experiments are usually done by providing respondents with the outcome of their choice and offering them the same choice set again. In our research, this was not the case, but respondents still had to make additional selections in the same choice set. Perhaps by not knowing the outcome of their first-best choice, they perceive making the second-best choice as switching and this may induce regret, as they do not know if their first choice turned out well or not.

The fact that a change from RRM to RUM for the second-best choice proves the finding of Ben-Akiva et al. (1992), that successive choices are less reliable in comparison with previous choices. The findings of Dyachenko et al. (2014) also echo such a statement, as in their research the intensity of taste parameters was lower in successive choices made by respondents.

7.2.2 Discussion of using best-worst choice tasks in transportation

Another point of discussion is the suitability of using best-worst style DCEs in the field of transportation. Their development was largely concentrated in the healthcare sector, where best-worst scaling experiments proved to have greater value for researchers as opposed to DCEs. Flynn, Louviere, Peters, & Coast (2007) and Zhang, Reed Johnson, Mohamed, & Hauber (2015) compared the use of BWS and DCE in healthcare and concluded that BWS style surveys are preferred thanks to their ability to determine the overall importance of an attribute itself on the utility of a product or service, not only the different attribute levels, as is the case in DCEs. Van Dijk, Groothuis-Oudshoorn, Marshall, & IJzerman (2016) assessed the different model outcomes of DCE and BWS and concluded they produce very similar results. When asked which of the tasks was more difficult, respondents said that BWS was, going against what some previous researches have postulated. DCE are often used in healthcare also due to the difficulty of using RP data, as the sector is a typical example of non-market behaviour: people are willing to pay almost any price if they get cured (Flynn et al., 2007). Using both methods in healthcare, it seemed logical to try and combine them: the result is the SBWDCE. It is however much more a DCE style method rather than a BWS style method. By incorporating the benefits of multiple choices within the same choice set, the method offers benefits in situations were smaller samples/populations are available, a particularly common occurrence in healthcare. If a certain drug applicable for only a handful of patients is to be evaluated, an SBWDCE can do so with much higher accuracy than a regular first-choice DCE.

In transportation on the other hand, populations to be investigated are rarely that limited. If one is to be evaluated, SBWDCEs can surely provide benefits, but on topics that concern larger populations, asking respondents to perform a best-worst style experiment may be counterintuitive. Dyachenko et al. (2014) explicitly warn in their research that in fields where typical marketplace condition apply (transport certainly falls into that category), use of BW tasks should be avoided if possible. They argue that because in reality people rarely have to decide for a worst alternative, such a choice task seem unnatural to respondents and may cause problems in model estimation. Ghijben, Lancsar, & Zavarsek (2014) propose a sequential best-best (SBB) approach to gathering discrete choices. They found that BW style choice tasks are more confusing and difficult for respondents, as they require them '*swap to a new mental task*' when they switch between best and worst (Lancsar, Fiebig, & Hole, 2017, p. 705). This fits well with the statement of a few respondents in our survey as well, saying they felt it strange to select the worst alternative, stating how they would least want to commute.

However, using SBWDCEs may prove very beneficial, in transportation as well, for surveys that aim to obtain prior values for the construction of efficient experimental designs. Obtaining priors can often be difficult as since it is not the main survey from which stated choice data is obtained, limited resources are often allocated to it. To avoid problems with less reliable parameters due to follow-up choices, a Bayesian efficient design can be generated with the obtained priors, maximising robustness of the final experimental design.

7.2.3 Discussion of accounting for choice set size variation

The flexibility of the developed SBWMNL μRRM models also warrants further discussion. The μRRM model can be used in a simple and straightforward way for modelling different decision rules with a single variable taking the value between zero and infinity. The difficulty of applying it in a SBWMNL model is the variation in choice set size. RRM models operate differently to conventional RUM models, as regret is based on the bilateral differences in attribute levels of two alternatives and the total regret of every alternative is based on the comparison with all other alternatives. A larger number of alternatives present in a choice set will therefore result in more comparisons and a larger total regret. That is why this choice set size variation is accounted for with the help of an additional constant or constants. While a simple way to tackle the given problem, it presents a new difficulty when applying the model to a case study where the number of alternatives can differ. If each choice set size is accounted for with its own constant, given the relationship between it and the scale parameter, there are limitations in the application. For example, in the survey respondents had to choose the best alternative among five, then the worst of the remaining four, the best of the remaining three and the worst of the two. Each choice is associated with its own scale parameter and a choice set size constant. Such a model can then only be used for determining the best alternative if the choice set is made up of five alternatives. In a choice set of four, the scale parameter and choice set size constant are associated with a worst alternative decision, so they cannot be used for determining the best alternative.

A potential solution for this problem, that was also used in this research is using a single constant that is corrected by the choice set size. This gives more flexibility to the model, but still leaves doubt if the constant can be extrapolated to choice set sizes beyond those used in the estimation phase. There is also the danger when using a single value for choice set size variation that the constant will influence the scale parameters, as happened for the second-best choice in this research, where the scale parameter changed from 0,5 to 12,31 which is a drastic decision rule change. It also casts doubt if the choice set size in which respondents had to select the worst alternative can be used for making predictions for best alternatives, as only worst choices were used in the model estimation process.

7.3 Limitations

This research provides new insight into the decision-making process of respondents in a best-worst choice task, proving that on this dataset different decision rules were used for best and worst choices. By including this variability in models, respondent behaviour can be represented more realistically, and the results can be used to make more accurate and confident forecasts. Such a modelling approach does however come with downsides and limitations.

To account for decision rule variability more complex modelling procedures with a larger number of estimated parameters are required, which leads to longer estimation times and more difficult outcome interpretation. In this research, the number of parameters increased from five in the SBWMNL RUM model to between eight and twelve parameters, resulting in estimation times increasing from a few seconds to up to ten minutes. Estimation time of LC models, due to the necessity of using different starting values to avoid getting stuck in local maxima, was very substantial, ranging anywhere between three and five hours if all 22 parameters were being estimated. Long estimation time is not bad per se, but it needs to be justified with significant model improvements.

Accounting for choice set size variation is also a limitation of this method. Although two different approaches were investigated, one negatively influences the model's applicability while the other causes doubt about its reliability. By using multiple choice set size constants for each choice set size, the applicability to any other choice set size is questionable and not recommended. As the scale parameter and choice set size are related, even using the choice set size where a worst alternative had to be selected for a best alternative prediction is uncertain. Using a single choice-set-size-adjusted constant dismisses this issue but reduces the reliability of the outcomes, as it influences the scale parameter values.

A significant downside of the developed model is the fact that it is based on a random regret minimisation model, which has the benefit of providing valuable information on consumer behaviour and the ability to estimate decision rules using a scale parameter, ranging from fully- to non-compensatory behaviour. The problem with RRM models, as has been shown in literature already, is that they are difficult if not impossible to use for valuation and appraisal of transport-related projects (and in other fields as well), most notably due to their inability to determine the net welfare effect (the value of the whole choice set). This issue of regret models was already discussed in research many times and proposals for metrics have been put forward by Dekker (2014) and more recently by Dekker & Chorus (2018). They have given RRM models more applicability for use in transport policy evaluation, although they still acknowledge the limitations of RRM models compared to RUM models, especially for carrying out welfare analyses. That is why RUM models remain the models of choice for policy and infrastructure project appraisals.

7.4 Recommendations

Following from the discussion and limitations, as well as from the research as a whole, a number of topics were found where there is need for further research to be carried out.

Starting with survey design, this research used a Bayesian efficient design due to the uncertainty of the accuracy of prior values used in the design generation. In addition, the D-error for all generated designs was also computed by considering a P-RRM decision rule and the composite D-error was considered when selecting the best design. We did not however consider the implications of carrying out a best-worst choice task on the generated design. Given that van Cranenburgh et al. (2018) have proven that traditional efficient designs were biased towards a RUM decision rule, there may also be benefits in accounting for the best-worst nature of the choice tasks carried out in the survey and thus a different D-error may emerge, or perhaps a new measure of efficiency could be developed to better account for such a different preference elicitation technique.

Generating synthetic data for a SBWDCE, the issue of how the error component is present in decision-making procedure came up. Given the nature of the developed model, new error terms were drawn for every new choice being made, even in the same choice set. The error terms could however, also stay with the same alternative throughout the choice set evaluation process (through all the best-worst choice tasks until being eliminated) or perhaps the error term is made up of two components: one that is alternative specific, meaning it is present with the alternative in all choice set combinations and one that is choice set specific, that is drawn for anew for every choice set and every alternative, regardless of the alternatives present in the choice set. Investigating how error components are assigned to alternatives can prove very valuable for future synthetic data generation processes.

In the modelling process itself, the uncovered inverse relationship between the scale parameter and the choice set size constant also warrants further investigation to better understand how one influences the other. As there are different ways of accounting for choice set size variation (van Cranenburgh, Prato, et al., 2015), knowing how this influences the scale parameter is vital for understanding the resulting scale parameter, which is then used to determine the decision rule utilised by respondents.

The models developed in this research and the theory used to underpin the outcomes can also benefit greatly from further applications in different fields. Given the findings of Dyachenko et al. (2014) and Flynn et al. (2007), it may be particularly interesting to see this model and theory being tested in healthcare. Applying it on more datasets can also prove if the theory for consumers' decision processes suggested in this research is correct or if it was merely due to the peculiarities of the particular sample used in this research. Considering the LC models, an application on a larger sample could show if classes with sufficiently different decision rule combinations do in fact exist. However, given the fact that the main benefit of the SBWDCEs is smaller sample size requirements, this may be counterproductive.

Policy recommendations

Although this research was largely based on examining the behavioural characteristics of individuals in an SBWDCE, the results from the case study where the model was applied can also be used to make conclusions and recommendations on what future transportation policies concerning park-and-ride facilities should include.

Perhaps the most important result the model application showed is that in the current situation and plan, the satellite P+R are both a focus of the regional policy and more attractive to commuters: as many as 95% of P+R commuters would be using satellite facilities. As has been shown in literature many times, implementing P+R policies incorrectly can cause more congestion and more vehicle miles travelled (Meek et al., 2011) (Mingardo, 2013) (RPS, 2009). Even if city centres are effectively freed from excessive car traffic by these policies, research has shown that satellite P+R facilities attract new users: those that would not have travelled otherwise and public transport users. In the model application, we also show just how difficult increasing the appeal of remote P+R and decreasing the appeal of satellite P+R is. It took a substantial change in pricing policy and better public transport services or even a new railway service just to get the share down from 95% to around 60%, meaning they would still hold the majority.

Knowing this, a clear recommendation for transport policy in the Central Slovenia Statistical Region is to reduce the focus on satellite P+R and rather put more emphasis on local, remote parking facilities. By providing small parking facilities at every stop along a regional transit corridor, local residents can be intercepted as close to their home as possible, reducing road congestion in and around cities. By offering higher quality public transport, many users would likely also switch to using only PT for their commute (by walking or cycling to the stop), reducing the requirements for parking spaces.

An interesting finding was also the valuation of travel time by train compared to bus. While respondents showed a preference for railway services, a bus service with a slightly better frequency proved just as attractive. That is why new railways should only be considered when bus services are overcrowded even in high-frequency operations. The model shows that a train service running every half hour is less attractive than a bus service running every fifteen minutes. Policymakers should therefore focus on improving bus services first.

A pricing policy similar to the one tested in the case study should also be considered at some point, or at the very least, the price of using remote P+R should be halved to bring it on the same level as satellite parking lots. Increasing the price of the latter should only come about when / after a stringent parking policy is adopted in the city.


Jamie Oliver
MAN
LJ LPP-429

8 References

Amaya-Amaya, M., Kalleh, M., Ryan, M., & Odejar, M. (2004). Complexity effects in Stated Preferences Choice Experiments: a simulation exercise. (p. 19). Presented at the Health Economics Study Group Workshop, Paris: Health Economics Research Unit Institute of Applied Health Sciences.

Arentze, T. A., & Molin, E. J. E. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. *Transportation Research Part A: Policy and Practice*, *58*, 15–28. https://doi.org/10.1016/j.tra.2013.10.005

Arentze, T., Borgers, A., Timmermans, H., & DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review*, *39*(3), 229–244. https://doi.org/10.1016/S1366-5545(02)00047-9

Arriva - avtobusni prevozi. (n.d.). Retrieved 26 March 2018, from http://arriva.si/

Beach, L. R., & Mitchell, T. R. (1987). Image theory: Principles, goals, and plans in decision making. *Acta Psychologica*, *66*(3), 201–220. https://doi.org/10.1016/0001-6918(87)90034-5

Bekhor, S., & Shiftan, Y. (2010). Specification and Estimation of Mode Choice Model Capturing Similarity between Mixed Auto and Transit Alternatives. *Journal of Choice Modelling*, *3*(2), 29–49. https://doi.org/10.1016/S1755-5345(13)70034-4

Ben-Akiva, M., & Lerman, S. R. (1994). *Discrete Choice Analysis: Theory and Application to Travel Demand* (6th ed.). Cambridge, Massachusetts: The MIT Press.

Ben-Akiva, M., Morikawa, T., & Shiroishi, F. (1992). Analysis of the reliability of preference ranking data. *Journal of Business Research*, *24*(2), 149–164. https://doi.org/10.1016/0148-2963(92)90058-J

Ben-Akiva, M., & Swait, J. (1986). The Akaike Likelihood Ratio Index. *Transportation Science*, *20*(2), 133–136. https://doi.org/10.1287/trsc.20.2.133

Bennett, J. (2005, April). Choice modelling: A step-by-step guide. Environemntal Protection Agency, Queensland Government. Retrieved from https://www.gbcma.vic.gov.au/downloads/sircs_review_2005_env_2/Choice_modelling_A_stepbystep_guide_by_Jeff_Bennett.pdf

Bergman, A., Gliebe, J., & Strathman, J. (2011). Modeling Access Mode Choice for Inter-Suburban Commuter Rail. *Journal of Public Transportation*, *14*(4), 23–42.

Bos, I. D. M., Heijden, R. E. C. M. V. der, Molin, E. J. E., & Timmermans, H. J. P. (2004). The Choice of Park and Ride Facilities: An Analysis Using a Context-Dependent Hierarchical Choice Experiment. *Environment and Planning A: Economy and Space*, *36*(9), 1673–1686. https://doi.org/10.1068/a36138

Chen, C., Xia, J. (Cecilia), Smith, B., Olaru, D., Taplin, J., & Han, R. (2015). Influence of Parking on Train Station Choice under Uncertainty for park-and-ride Users. *Procedia Manufacturing*, *3*, 5126–5133. https://doi.org/10.1016/j.promfg.2015.07.537

Choice Metrics. (2018, January 30). Ngene 1.2: User Manual & Reference Guide. Retrieved from http://www.choice-metrics.com/NgeneManual120.pdf

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. (2012). *Random Regret-based discrete choice modeling: A tutorial*. Heidelberg, Germany: Springer.

Chorus, C. G. (2014a). A Generalized Random Regret Minimization model. *Transportation Research Part B: Methodological*, *68*(Supplement C), 224–238. https://doi.org/10.1016/j.trb.2014.06.009

Chorus, C. G. (2014b). Benefit of adding an alternative to one's choice set: A regret minimization perspective. *Journal of Choice Modelling*, *13*(Supplement C), 49–59. https://doi.org/10.1016/j.jocm.2014.04.001

Chorus, C. G. (2016, November). *Choice Behaviour Modelling Methods*. Delft.

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. https://doi.org/10.1016/j.trb.2007.05.004

Chorus, C. G., Koetse, M. J., & Hoen, A. (2013). Consumer preferences for alternative fuel vehicles: Comparing a utility maximization and a regret minimization model. *Energy Policy*, *61*, 901–908. https://doi.org/10.1016/j.enpol.2013.06.064

Chorus, C. G., Rose, J. M., & Hensher, D. A. (2012). Hybrid Models of Random Utility Maximisation and Random Regret Minimization: Results from two Empirical Studies (p. 22). Presented at the Transport research Board 92nd Annual Meeting, Washington, D.C. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:8241cf18-66b0-4cf0-9441-a9372d3628cf/datastream/OBJ

Chorus, C. G., van Cranenburgh, S., & Dekker, T. (2014). Random regret minimization for consumer choice modeling: Assessment of empirical evidence. *Journal of Business Research*, *67*(11), 2428–2436. https://doi.org/10.1016/j.jbusres.2014.02.010

Chung, C., Boyer, T., & Han, S. (2011). How many choice sets and alternatives are optimal? Consistency in choice experiments. *Agribusiness*, *27*(1), 114–125. https://doi.org/10.1002/agr.20252

Dekker, T. (2014). Indifference based value of time measures for Random Regret Minimisation models. *Journal of Choice Modelling*, *12*(Supplement C), 10–20. https://doi.org/10.1016/j.jocm.2014.09.001

Dekker, T., & Chorus, C. G. (2018). Consumer surplus for random regret minimisation models. *Journal of Environmental Economics and Policy*, *7*(3), 269–286. https://doi.org/10.1080/21606544.2018.1424039

Direkcija RS za Infrastrukturo. (2017). Traffic Load on Roads in Slovenia in 2016. Retrieved from http://www.di.gov.si/fileadmin/di.gov.si/pageuploads/Prometni_podatki/Prometne_obremenitve_2016_karta.pdf

Dyachenko, T., Reczek, R. W., & Allenby, G. M. (2014). Models of sequential evaluation in best-worst choice tasks. *Marketing Science*, *33*(6), 828–848. https://doi.org/10.1287/mksc.2014.0870

Eurostat. (2017, April). Passenger cars in the EU. Retrieved 13 February 2018, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Passenger_cars_in_the_EU

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, *26*(1), 171–189. https://doi.org/10.1016/j.jhealeco.2006.04.002

Ghijben, P., Lancsar, E., & Zavarsek, S. (2014). Preferences for oral anticoagulants in atrial fibrillation: a best-best discrete choice experiment. *PharmacoEconomics*, *32*(11), 1115–1127. https://doi.org/10.1007/s40273-014-0188-0

Google Earth. (2016). Barje P+R.

Google Maps. (n.d.). Retrieved 27 October 2017, from https://www.google.si/maps/@39.9018207,116.5143522,12.81z

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, *37*(8), 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2

Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice modelling approaches: A superior alternative for environmental valuation? *Journal of Economic Surveys*, *15*(3), 435–462.

Hauser, J. R. (1978). Testing the Accuracy, Usefulness, and Significance of Probabilistic Choice Models: An Information-Theoretic Approach. *Operations Research*, *26*(3), 406–421. https://doi.org/10.1287/opre.26.3.406

Hess, S., Beck, M. J., & Chorus, C. G. (2014). Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives. *Transportation Research Part A: Policy and Practice*, *66*(Supplement C), 1–12. https://doi.org/10.1016/j.tra.2014.04.004

Hess, S., Ben-Akiva, M., Gopinath, D., & Walker, J. (2008). Advantages of latent class models over continuous mixture models in capturing heterogeneity. *EUROPEAN TRANSPORT CONFERENCE 2008; PROCEEDINGS*. Retrieved from https://trid.trb.org/view/923959

Hess, S., Hensher, D. A., & Daly, A. (2012). Not bored yet – Revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice*, *46*(3), 626–644. https://doi.org/10.1016/j.tra.2011.11.008

Hess, S., & Rose, J. M. (2009). Allowing for intra-respondent variations in coefficients estimated on repeated choice data. *Transportation Research Part B: Methodological*, *43*(6), 708–719. https://doi.org/10.1016/j.trb.2009.01.007

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*, 27–38. https://doi.org/10.1016/j.jocm.2013.12.005

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. https://doi.org/10.1007/s11116-011-9365-6

Inman, J. J., & Zeelenberg, M. (2002). Regret in Repeat Purchase versus Switching Decisions: The Attenuating Role of Decision Justifiability. *Journal of Consumer Research*, *29*(1), 116–128. https://doi.org/10.1086/339925

Kam Bus. (n.d.). Retrieved 26 March 2018, from https://www.kam-bus.si/

Klemenčič, M., Lep, M., Mesarec, B., & Žnuderl, B. (2014). *Potovalne navade prebivalcev v Mestni občini Ljubljana in Ljubljanski urbani regiji* (p. 96). Ljubljana: Mestna občina Ljubljana.

Lancsar, E., Fiebig, D. G., & Hole, A. R. (2017). Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. *PharmacoEconomics*, *35*(7), 697–716. https://doi.org/10.1007/s40273-017-0506-4

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., & Burgess, L. (2013). Best worst discrete choice experiments in health: Methods and an application. *Social Science & Medicine*, *76*(Supplement C), 74–82. https://doi.org/10.1016/j.socscimed.2012.10.007

Ljubljanski potniški promet. (n.d.). Ceniki. Retrieved 6 October 2018, from http://www.lpp.si/javni-prevoz/ceniki

Louviere, J. J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. A. J. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling*, *1*(1), 128–164. https://doi.org/10.1016/S1755-5345(13)70025-3

Mahmoud, M., Habib, K. N., & Shalaby, A. (2014). Park-and-Ride Access Station Choice Model for Cross-Regional Commuter Trips in the Greater Toronto and Hamilton Area (GTHA). Presented at the TRB 2014 Annual Meeting, Washington, D.C.

Mai, T., Frejinger, E., & Bastin, F. (2015). Comparing Regret Minimization and Utility Maximization for Route Choice using the Recursive Logit Model. *CIRRELT*, *2015*(38), 1–20.

Malhotra, N. K. (1982). Information Load and Consumer Decision Making. *Journal of Consumer Research*, *8*(4), 419–430.

McFadden, D. (2000, March). Disaggregate Behavioral Travel Demand's RUM Side - A 30-Year Retrospective. Department of Economics, University of California, Berkley. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.4872&rep=rep1&type=pdf

Meek, S., Ison, S., & Enoch, M. (2011). Evaluating alternative concepts of bus-based park and ride. *Transport Policy*, *18*(2), 456–467. https://doi.org/10.1016/j.tranpol.2010.09.006

Meloy, M. G., & Russo, J. E. (2004). Binary choice under instructions to select versus reject. *Organizational Behavior and Human Decision Processes*, *93*(2), 114–128. https://doi.org/10.1016/j.obhdp.2003.12.002

Milanovič, K., Vertelj Nared, P., Kranjc, U., Longar, U., Blaž, T., Jankovič, K., … Butina, K. (2017). *Celostna prometna strategija Mestne občine Ljubljana* (p. 54). Ljubljana: Mestna občina Ljubljana. Retrieved from https://cpsmol.projekti.si/Data/Sites/1/media/Prometna%20strategija.pdf

Mingardo, G. (2013). Transport and environmental effects of rail-based Park and Ride: evidence from the Netherlands. *Journal of Transport Geography*, *30*, 7–16. https://doi.org/10.1016/j.jtrangeo.2013.02.004

Ministrstvo za Infrastrukturo. (2016). Subvencionirana vozovnica. Retrieved 13 February 2018, from http://www.mzi.gov.si/si/delovna_podrocja/promet/subvencionirana_vozovnica/

Molin, E. (2016a, December). *Lecture 1: An introduction to stated choice experiments*. Delft.

Molin, E. (2016b, December). *Lecture 3: Efficient designs*. Delft.

Olaru, D., Smith, B., Xia, J. (Cecilia), & Lin, T. (Grace). (2014). Travellers' Attitudes Towards Park-and-Ride (PnR) and Choice of PnR Station: Evidence from Perth, Western Australia. *Procedia - Social and Behavioral Sciences*, *162*, 101–110. https://doi.org/10.1016/j.sbspro.2014.12.190

Ordóñez, L. D., Benson III, L., & Beach, L. R. (1999). Testing the Compatibility Test: How Instructions, Accountability, and Anticipated Regret Affect Prechoice Screening of Options. *Organizational Behaviour and Human Decision Processes*, *78*(1), 63–80.

Pang, H., & Khani, A. (2018). Modeling park-and-ride location choice of heterogeneous commuters. *Transportation*, *45*(1), 71–87. https://doi.org/10.1007/s11116-016-9723-5

Potniški promet - Slovenske železnice. (n.d.). Retrieved 23 March 2018, from http://www.slo-zeleznice.si/sl/

Potovalna agencija. (n.d.). Retrieved 26 March 2018, from https://www.alpetour.si/

Qin, H., Guan, H., & Wu, Y.-J. (2013). Analysis of park-and-ride decision behavior based on Decision Field Theory. *Transportation Research Part F: Traffic Psychology and Behaviour*, *18*, 199–212. https://doi.org/10.1016/j.trf.2013.02.001

RDA LUR. (2014). *Park and ride: for sustainable mobility in the Ljubljana Urban Region* (1st ed.). Ljubljana: RDA LUR. Retrieved from http://www.rralur.si/sites/default/files/rralur/P_R_brosura-pdf%5B1%5D.pdf

Reed Johnson, F., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., … Bridges, J. F. P. (2013). Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*, *16*(1), 3–13. https://doi.org/10.1016/j.jval.2012.08.2223

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In *Encyclopedia of Database Systems* (pp. 532–538). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_565

Rose, J. M., & Bliemer, M. C. (2005). Constructing efficient choice experiments. *Institute of Transport Studies and Logistics Working Paper*, (ITLS-WP-05-07). Retrieved from https://trid.trb.org/view/771345

RPS. (2009). *The effectiveness and sustainability of Park and ride*.

Ryan, M., & Skåtun, D. (2004). Modelling non-demanders in choice experiments. *Health Economics*, *13*(4), 397–402. https://doi.org/10.1002/hec.821

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing Versus Satisficing: Happiness is a Matter of Choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197.

Sharma, B., Hickman, M., & Nassir, N. (2017). Park-and-ride lot choice model using random utility maximization and random regret minimization. *Transportation*, 1–16. https://doi.org/10.1007/s11116-017-9804-0

Simonson, I. (1992). The Influence of Anticipating Regret and Responsibility on Purchase Decisions. *Journal of Consumer Research*, *19*(1), 105–118. https://doi.org/10.1086/209290

Spillar, R. J. (1997). *Park-and-Ride Planning and Design Guidelines*. New York, New York: Parsons Brinckerhoff Inc.

STAGE. (n.d.). Geographic information system. Slovenia. Retrieved from http://gis.stat.si/

Statistični Urad Republike Slovenije. (n.d.-a). Delovne migracije, letni podatki. Retrieved 13 February 2018, from http://pxweb.stat.si/pxweb/Database/Dem_soc/07_trg_dela/05_akt_preb_po_regis_virih/10_07234_delovne_migracije/10_07234_delovne_migracije.asp

Statistični Urad Republike Slovenije. (n.d.-b). SURS. Retrieved 9 October 2018, from https://www.stat.si/StatWeb/

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)*, 276–278.

Strniša, E. (2009, October 18). Slovenija, dežela avtomobilov? *Prvi Interaktivni Multimedijski Portal, MMC RTV Slovenija*. Retrieved from http://www.rtvslo.si/okolje/novice/slovenija-dezela-avtomobilov/214862

The World Bank. (n.d.). Urban population. Retrieved 13 February 2018, from https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2016&locations=SI-AT-BG-HR-CZ-DK-FI-FR-ES-PT-UK-SK-PL-HU-IE-BE-IT-DE-RO-GR-SE-LT-LV-LU-EE-MT-CY&start=1960&view=chart

Timmermans, H., & Molin, E. (2009). Dealing with Increased Complexity in Conjoint Experiments: Background and Overview of Alternate Approaches. *Transport Reviews*, *29*(5), 12.

Train, K. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambrdige University Press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Uršič, M. (2012). Urbanisation History of Slovenia in the Context of Eastern and South-Eastern European Countries. *European Quarterly of Political Attitudes and Mentalities*, *1*(1), 38–47.

van Cranenburgh, S. (2017, November). *Latent class discrete choice models for travel behaviour research*. Delft.

van Cranenburgh, S. (2018, February). *Discrete keuzemodellen en het Landelijk modelsysteem (LMS)*. Delft. Retrieved from https://brightspace.tudelft.nl/d2l/le/content/67295/Home?itemIdentifier=D2L.LE.Content.ContentObject.ModuleCO-675317 van Cranenburgh, S., & Chorus, C. G. (2017). Does The Decision Rule Matter For Large-Scale Transport Models? *Transportation Research Procedia*, *23*(Supplement C), 848–867. https://doi.org/10.1016/j.trpro.2017.05.047

van Cranenburgh, S., & Collins, A. T. (2019). New software tools for creating stated choice experimental designs efficient for regret minimisation and utility maximisation decision rules.

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*(Supplement C), 91–109. https://doi.org/10.1016/j.tra.2015.01.008

van Cranenburgh, S., & Prato, C. G. (2016). On the robustness of random regret minimization modelling outcomes towards omitted attributes. *Journal of Choice Modelling*, *18*(Supplement C), 51–70. https://doi.org/10.1016/j.jocm.2016.04.004

van Cranenburgh, S., Prato, C. G., & Chorus, C. G. (2015). Accounting for variation in choice set size in Random Regret Minimization models. Retrieved from https://docs.wixstatic.com/ugd/4e8049_201cb98598824c959cf7108599a6b1da.pdf

van Cranenburgh, S., Rose, J. M., & Chorus, C. G. (2018). On the robustness of efficient experimental designs towards the underlying decision rule. *Transportation Research Part A: Policy and Practice*, *109*, 50–64. https://doi.org/10.1016/j.tra.2018.01.001

van Dijk, J. D., Groothuis-Oudshoorn, C. G. M., Marshall, D. A., & IJzerman, M. J. (2016). An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery. *Value in Health*, *19*(4), 316–322. https://doi.org/10.1016/j.jval.2015.12.020

Vanniyasingam, T., Daly, C., Jin, X., Zhang, Y., Foster, G., Cunningham, C., & Thabane, L. (2018). Investigating the impact of design characteristics on statistical efficiency within discrete choice experiments: A systematic survey. *Contemporary Clinical Trials Communications*, *10*, 17–28. https://doi.org/10.1016/j.conctc.2018.01.002

Vozni red. (n.d.). Retrieved 23 March 2018, from https://www.ap-ljubljana.si/vozni-red/

Vozni redi. (n.d.). Retrieved 23 March 2018, from http://www.lpp.si/javni-prevoz/vozni-redi

Walker, J. L., Wang, Y., Thorhauge, M., & Ben-Akiva, M. (2018). D-efficient or deficient? A robustness analysis of stated choice experimental designs. *Theory and Decision*, *84*(2), 215–238. https://doi.org/10.1007/s11238-017-9647-3

Yu, E. C., Fricker, S., & Kopp, B. (2015). Can Survey Intructions Relieve Respondent Burden? (p. 10). Presented at the AAPOR 2015, Hollywood, Florida: Bureau of Labor Statistics.

Zeelenberg, M., van den Bos, K., van Dijk, E., & Pieters, R. (2002). The inaction effect in the psychology of regret. *Journal of Personality and Social Psychology*, *82*(3), 314–327.

Zhang, J., Reed Johnson, F., Mohamed, A. F., & Hauber, A. B. (2015). Too many attributes: A test of the validity of combining discrete-choice and best–worst scaling data. *Journal of Choice Modelling*, *15*, 1–13. https://doi.org/10.1016/j.jocm.2014.12.001

Zhao, X., Li, Y., & Xia, H. (2017). Behavior decision model for park-and-ride facilities utilization. *Advances in Mechanical Engineering*, *9*(7), 1–9.

9 Appendix

In the appendix, additional information and more detailed explanations of certain processes are shown, as they were deemed not as important as the final result in the main report but are still necessary to show how they were obtained.

A List of current and planned official P+R lots in the CSSR

Table 9.1 below lists the 28 officially recognised P+R lots in the Central Slovenia Statistical Region. The P+R are also represented on a map in Figure 9.1. Out of all of them, ten are already open, some are in the final stages of planning, while others are further away from opening. The parking lots are classified as either satellite or local P+R and while for most, the distinction can be made quite clearly, for Brezovica (13) and Medvode (22), this is less clear. While they are not located exactly on the edge of the city, next to the highway ring road, they are still situated along major roads that carry several thousand commuters to work every day. From the southeast, Brezovica (13) can be seen as the third satellite P+R, for use after Dolgi most (4) and Pri Gorjancu (7) fill up. A similar argument could be made for Medvode (22) acting as a backup in case Stanežiče (9) fills up. It is also worth noting that Ježica (5) and Ig (18) are not identified in the P+R development scheme but are officially recognised by the urban bus operator: LPP.

Table 9.1. Official P+R parking lots in the CSSR

ID	Parking lot	Satellite or Local	Status
1	Barje	Satellite	Open
2	Bizovik	Satellite	Planned
3	Črnuče	Satellite	Planned
4	Dolgi most	Satellite	Open
5	Ježica	Satellite	Open
6	Polje	Local	Planned
7	Pri Gorjancu	Satellite	Planned
8	Rudnik	Satellite	Planned
9	Stanežiče	Satellite	Planned
10	Stožice	Satellite	Open
11	Studenec	Satellite	Open
12	Zadobrova	Satellite	Planned
13	Brezovica	Satellite + Local	Planned
14	Dobrepolje	Local	Planned
15	Dobrova	Local	Planned
16	Domžale	Local	Open
17	Grosuplje	Local	Planned
18	Ig	Local	Open
19	Ivančna Gorica	Local	Planned
20	Kamnik	Local	Planned
21	Litija	Local	Planned
22	Medvode	Satellite + Local	Planned
23	Moravče	Local	Planned
24	Rodica	Local	Open
25	Škofljica	Local	Open
26	Trzin	Local	Planned
27	Vodice	Local	Planned
28	Vrhnika	Local	Open

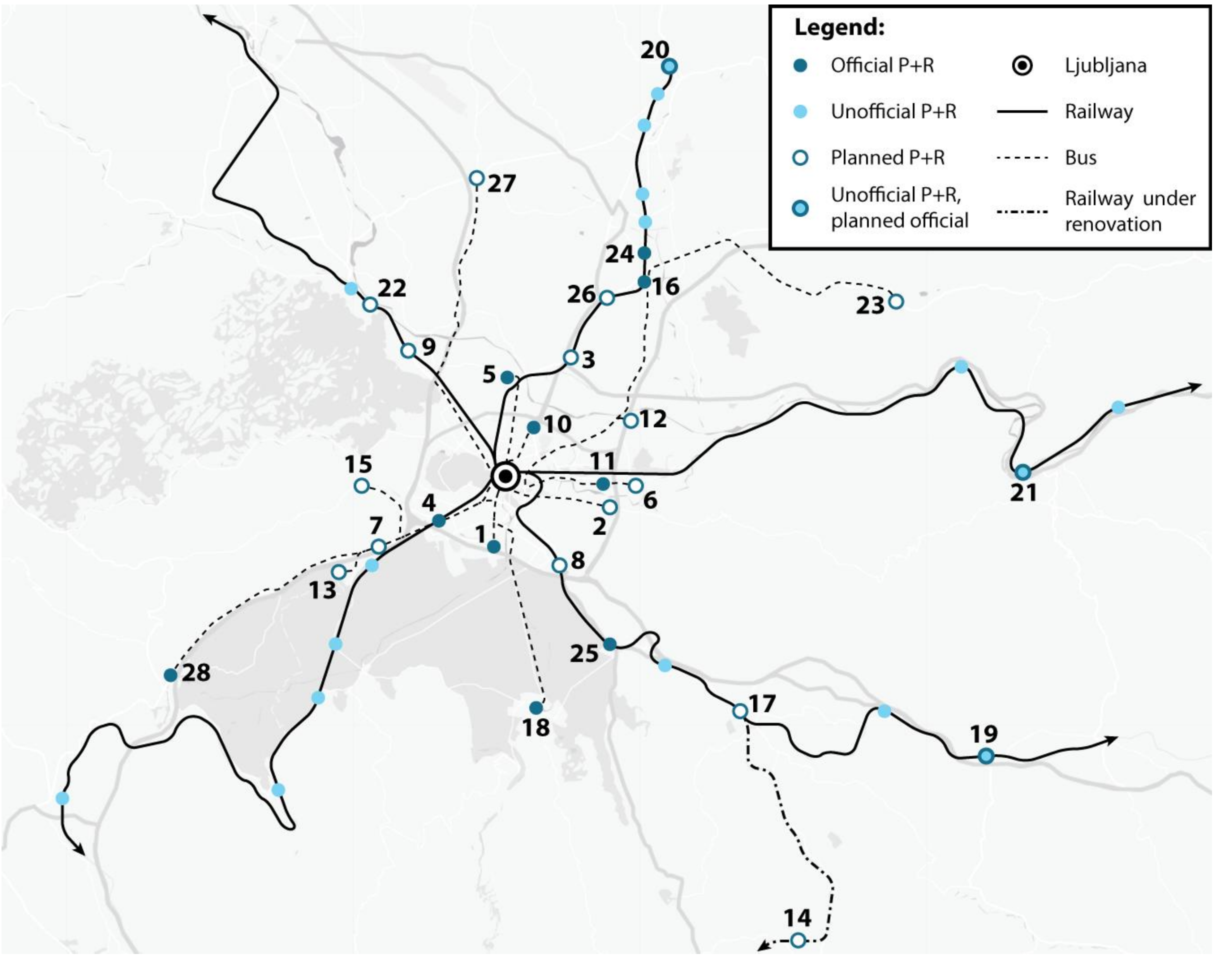


Figure 9.1. Map of official, unofficial and planned P+R facilities in the Central Slovenia Statistical Region

B Attribute level selection

In this subchapter of the appendix, more detailed information is given on how the attribute levels were derived. A total of five attributes are present in the choice experiment and here, the selection of levels for each of them is explained. The attributes are:

- Travel time by car
- Cost of parking and public transport ticket
- Travel time by public transport
- Public transport mode
- Public transport frequency / headway

For car travel times, the minimal travel time of 5 min was chosen. This is a realistic or perhaps even slightly exaggerated for situations where commuters would drive to a P+R facility from a nearby village or from a town centre to just outside the town. A lower travel time would also have been realistic but was deemed too low.

For determining the highest attribute level of car travel time, the access times from settlements on the edge of the region, to a satellite P+R were considered. The highest travel time is 55 min in uncongested situations (from Gabrovka (1) to Stožice P+R). It should be noted that most of the remote settlements have a relatively small population and are not representative, rather the extreme. For this reason, towns with a somewhat substantial population, yet still close to the regional border, were also considered. In Table 9.2, settlements are divided into three groups, based on their populations: below 500 inhabitants, between 500 and 1.000 and above 1.000 inhabitants. They are then ordered in each group based on car access time from highest to lowest. A map indicating the location of each settlement and shortest path is also presented in Figure 9.2. The source for settlement population was (STAGE, n.d.) and the source for travel times ('Google Maps', n.d.), with uncongested times obtained by setting the departure at 05:00.

In Table 9.2, all but one settlement are within 45 minutes driving from a satellite P+R, so this was chosen as the highest attribute level. It is also convenient, as with five attribute levels, they can be spaced at ten-minute intervals (5-15-25-35-45), preserving attribute level equidistance.

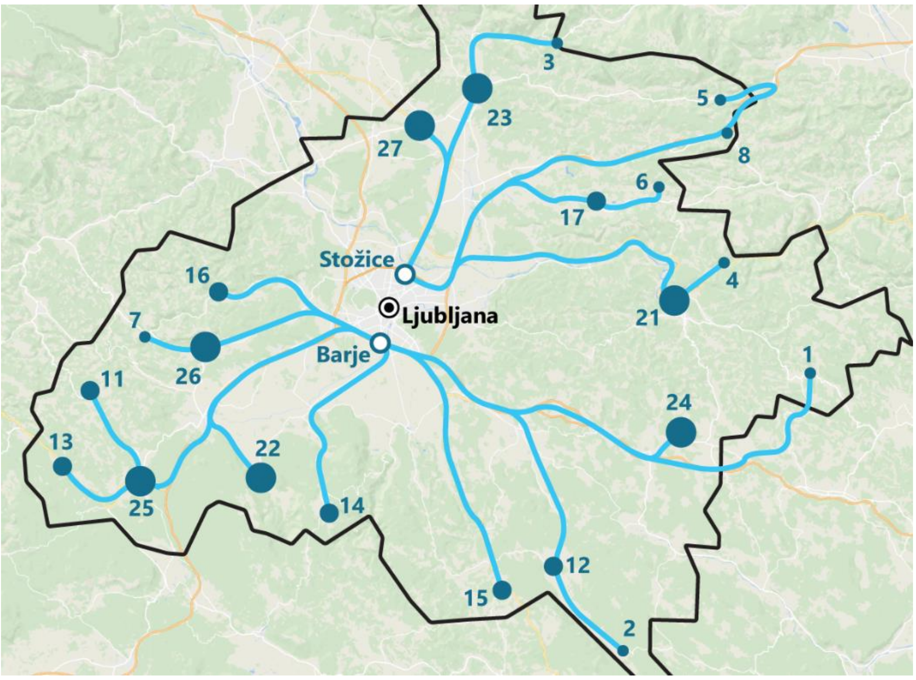


Figure 9.2. Map of settlements on the edge of the region with the indicated path to a satellite P+R facility

Table 9.2. Settlements in the Central Slovenia Statistical Region with the access time to a satellite P+R facility

	ID	Settlement	Population	Closest satellite P+R	P+R access time [min]
Population < 500	**1**	Gabrovka	232	Barje	50
	2	Rapljevo	68	Barje	45
	3	Podlom	45	Stožice	45
	4	Sava	276	Stožice	40
	5	Motnik	178	Stožice	35
	6	Peče	207	Stožice	35
	7	Šentjošt	388	Barje	35
	8	Trojane	120	Stožice	28
Population < 1.000	**11**	Rovte	959	Barje	40
	12	Videm	573	Barje	35
	13	Hoterdšica	599	Barje	35
	14	Rakitna	706	Barje	30
	15	Velike Lašče	700	Barje	28
	16	Polhov Gradec	655	Barje	26
	17	Moravče	908	Stožice	20
Population > 1.000	**21**	Litija	6.467	Stožice	35
	22	Borovnica	2.373	Barje	24
	23	Kamnik	13.803	Stožice	24
	24	Šentvid pri Stični	1.034	Barje	22
	25	Logatec	9.338	Barje	22
	26	Horjul	1.411	Barje	22
	27	Moste pri Komendi	1.338	Stožice	20

Public transport attribute levels were obtained from service characteristics in the Central Slovenia Statistical Region. For this, bus and train stations with an official, unofficial or planned P+R facility were used. For some stations, travel times had to be extrapolated from the timetable as they do not yet exist (in the case of many planned P+R facilities). Travel time, frequency and fare information were obtained from 'Vozni redi' (n.d.) for city buses, 'Potniški promet - Slovenske železnice' (n.d.) was used for train data and regional bus data was gathered from 'Arriva - avtobusni prevozi' (n.d.), 'Kam Bus' (n.d.), 'Potovalna agencija' (n.d.) and 'Vozni red' (n.d.). Since there is no single tool where all the information is available, data had to be gathered from each individual operator.

With respect to travel time to the city centre of Ljubljana, the lowest was five minutes, with all others being ten minutes or more. The five-minute travel time was also extrapolated, as the train station is currently under construction and the travel time is not yet known. For this reason, 10 min was taken as the lowest level. The highest travel times were 28 min for city bus, 55 min for regional bus and 48 min for train. The highest level was chosen to be 50 min, again to have the levels spaced at ten-minute intervals, preserving attribute level equidistance. Making a contingency table of occurrences for all travel times, the result is presented in Table 9.3. Most travel times fall within the first three groups.

Among public transport frequencies, the lowest was two vehicles per hour for city bus and train, with one per hour for regional bus which occurred in two instances. Both are rather irrelevant, as a rail service is also present at both locations (at one, the railway is currently being upgraded). The lowest level for the survey is therefore two vehicles per hour or a 30 min interval. The highest frequencies are offered for city buses at 15 vehicles per hour, ten regional buses per hour and three trains per hour. Since the highest city bus frequency is a combination of frequencies at two nearby stops, the highest frequency at a single stop is taken as the highest level: 12 vehicles per hour or a vehicle every 5 minutes. The results from the CSSR are present in Table 9.4. Unlike all other attributes, a combination of attribute levels that preserved equidistance and retained realistic frequencies could not be constructed. The decision was taken to rather preserve a realistic service pattern, rather than equidistance.

Table 9.3. Contingency table of public transport travel times

Rounded travel time [min]	City bus	Regional bus	Train
10	2	1	7
20	9	6	5
30	5	7	4
40	0	2	2
50	0	3	2

Table 9.4. Contingency table of public transport frequencies

Frequencies	Intervals [min]	City bus	Regional bus	Train
2	**30**	1	5	16
3	**20**	5	7	4
4	**15**	2	0	0
6	**10**	5	6	0
12	**5**	3	1	0

Prices of official P+R facilities are often combined with a return public transport ticket, meaning that when paying for parking, the user automatically receives a return ticket from the P+R to the centre of Ljubljana. Unofficial P+R are often free with some costing only a few euros for the whole day. The lowest combined P+R and return ticket fare costs 1,20€ at satellite P+R facilities, while the most expensive is at remote P+R and comes up to 5,40€. For the unofficial P+R facilities, only the return public transport ticket price was considered in the analysis. This yielded the highest price of 9,60€ which is for a regional bus. This price is not realistic, as a faster train service from the same station is available for only 6,88€. The highest price is therefore 8,56€. Rounding the prices for the survey, the chosen minimum and maximum levels are 1€ and 9€. The five attribute levels, with equidistance are therefore 1€, 3€, 5€, 7€ and 9€.

The attribute describing public transport mode contains only the currently available modes: bus and train. No distinction is made between city and regional bus since the travel time is also given. No other modes, such as tram or metro, are included since there is no firm plan for these modes to be implemented.

C Lead text in survey

At the beginning of the survey, respondents were given a full description of what the research is investigating, a description of the choice situations, instructions how to answer the questions and the context they should consider when making decisions. Both the original text (left column) and the translated text (right column) are presented in Table 9.5.

Table 9.5. Survey lead text in Slovene and English

Spoštovani	Dear survey participant
Najlepša hvala za vaše sodelovanje v anketi. Izpolnjevanje ankete vam bo vzelo približno 15-20 minut.	Thank you for participating in this survey. Taking part in the survey will take you approximately 15-20 minutes.
Sem študent na Tehnični Univerzi v Delftu, na Nizozemskem, kjer izdelujem magistrsko nalogo na smeri Promet, Infrastruktura in Logistika. V magistrski nalogi bom raziskoval, kako ljudje sprejemajo odločitve za vsakodnevna potovanja. Pri tem bom uporabil primer izbire parkirišča Parkiraj in se pelji (P+R) za pot v Ljubljano.	I am a graduating student at the Delft University of Technology, in the programme of Transport, Infrastructure and Logistics. My graduation topic is looking into how people make choices, focused on the field of transportation and with an example in the choice of park-and-ride facilities for travellers going to Ljubljana.
Anketa je sestavljena iz dveh delov: • 12 situacij/primerov izbire parkirišča P+R • Splošne informacije o anketirancu in potovalnih navadah	The survey is made up of two parts: • 12 Choice situations on park-and-ride facility choice • Additional respondent information
Situacije:	**Choice situations:**
Spodaj se nahaja 12 izbirnih situacij. V vsaki izmed njih je 5 različnih P+R parkirišč. Parkirišča so opisana s 5 značilnostmi: • Čas vožnje z avtomobilom od doma do parkirišča • Skupna cena parkiranja in vozovnice za javni prevoz • Čas vožnje z javnim prevozom od parkirišča do cilja • Vrsta javno prevoznega sredstva (avtobus ali vlak) • Pogostost voženj javnega prevoza	Below, you will find 10 choice situations, containing 5 alternative P+R facilities each. The P+R facilities are described by 5 attributes: • Travel time by car from your home to the P+R • Total price, containing both the parking cost and public transport ticket • Travel time from the P+R to your final destination by public transportation • The mode of public transportation available at the P+R (bus or train) • The frequency of PT services
Navodila:	**Instructions:**
V vsaki izbirni situaciji vas prosim, da izberete sledeče: • **Najboljšo** alternativo med vsemi • **Najslabšo** od preostalih 4 alternativ • **Najboljšo** od preostalih 3 alternativ • **Najslabšo** od preostalih 2 alternativ	In each choice situation, you will be asked to make the following choices. • Choose the **best** alternative • From the remaining 4, choose the **worst** • From the remaining 3, choose the **best** • From the remaining 2, choose the **worst**

Po vsaki izbrani alternativi se le-ta odstrani, ostale pa ostanejo za naslednje vprašanje, dokler ne zmanjka alternativ.

Pri izbiri najboljše / najslabše alternative vas prosim, da izberete tisto, ki bi jo vi najbolj / najmanj verjetno izbrali, če bi potovanje morali opraviti in bi bile to vaše možnosti.

Kontekst:

Pri odgovarjanju zanemarite, kje živite trenutno, kje ste zaposleni in s kakšnim namenom potujete.

Predstavljajte si, da živite izven občine Ljubljana in ste namenjeni v Ljubljano na delo oz. izobraževanje. Odločili ste se, da boste šli z avtom do parkirišča P+R in od tam z javnim prevozom v mesto. Alternative predstavljajo 5 različnih parkirišč P+R. Nekaj dodatnih informacij glede potovanja:

- Pri nobeni od alternativ ni nevarnosti, da bi z avtomobilom ali avtobusom obstali v gneči na cesti. Navedeni potovalni časi so zagotovljeni.
- Na parkirišču je zagotovo prosto parkirno mesto.
- Na poti z javnim prevozom vam ni potrebno prestopati.
- Od izstopne postaje do vašega cilja je manj kot 5 min hoje.
- Vsa parkirišča in javno prevozna sredstva so plačljiva z enotno kartico (npr. Urbana), ki jo posedujete.

After every question, the selected alternative is removed and only the non-chosen ones remain, until there are no more alternatives left.

When choosing the best and worst alternatives, please select the ones which you would choose most or least likely, given that you have to make the trip, and these are your only options.

Context:

For this experiment, please disregard your current household location, occupation and reason for travelling.

In these choice situations, the context is that you are living outside of Ljubljana and travelling to Ljubljana to work / school. You have decided to drive from home, park your car at a P+R facility and take public transport to the city. The five alternatives represent five different P+R parking lots. Some additional information is:

- There is no danger of getting stuck in traffic. The car and PT travel times are guaranteed.
- A free parking space is available at all the P+R facility.
- No transfers are required to reach your destination.
- From your PT exit stop, there is less than 5 min walking (egress) time to your destination
- All P+R facilities and PT are paid by a single smart card (Urbana), which you possess

D Choice experiment choice set layout alternatives

This appendix provides the visual representation of all three question layout types. Layouts 1 (Figure 9.3) and 2 (Figure 9.4) have the same graphic but differ in how many questions are asked at the same time. Asking only one question at a time in Layout 2 avoids respondents (accidentally or on purpose) choosing the same alternative multiple times. In Layout 3 (Figure 9.5) the images for each alternative are individual and thus when an alternative is removed after being chosen, so too is the image. This is more reminiscent of the modelling procedure than the other two proposed layouts.

	Alternativa 1	Alternativa 2	Alternativa 3	Alternativa 4	Alternativa 5
Čas vožnje z avtomobilom	15 min	10 min	5 min	25 min	10 min
Potovalni čas z javnim prevozom	25 min	25 min	15 min	10 min	30 min
Cena parkiranja in javnega prevoza	7 €	9 €	3 €	3 €	3 €
Vrsta javnega prevoza	avtobus	avtobus	avtobus	vlak	avtobus
Pogostost javnega prevoza	*vsakih* 15 min	*vsakih* 10 min	*vsakih* 30 min	*vsakih* 10 min	*vsakih* 20 min

Izbirna situacija **All four questions**

	Alternativa 1	Alternativa 2	Alternativa 3	Alternativa 4	Alternativa 5
Katera alternativa je za vas **najboljša**?	○	○	○	○	○
Katera alternativa **od preostalih 4** je za vas **najslabša**?	○	○	○	○	○
Katera alternativa **od preostalih 3** je za vas **najboljša**?	○	○	○	○	○
Katera alternativa **od preostalih 2** je za vas **najslabša**?	○	○	○	○	○

Figure 9.3. Layout 1, with all four questions relating to the same choice set being asked at once

	Alternativa 1	Alternativa 2	Alternativa 3	Alternativa 4	Alternativa 5
Čas vožnje z avtomobilom	15 min	10 min	5 min	25 min	10 min
Potovalni čas z javnim prevozom	25 min	25 min	15 min	10 min	30 min
Cena parkiranja in javnega prevoza	7 €	9 €	3 €	3 €	3 €
Vrsta javnega prevoza	avtobus	avtobus	avtobus	vlak	avtobus
Pogostost javnega prevoza	*vsakih* 15 min	*vsakih* 10 min	*vsakih* 30 min	*vsakih* 10 min	*vsakih* 20 min

One question at a time

Katera alternative je za vas najboljša *

Alternativa 1	Alternativa 2	Alternativa 3	Alternativa 4	Alternativa 5
○	○	○	○	○

Figure 9.4. Layout 2, with the same image as Layout 1 but asking questions one at a time

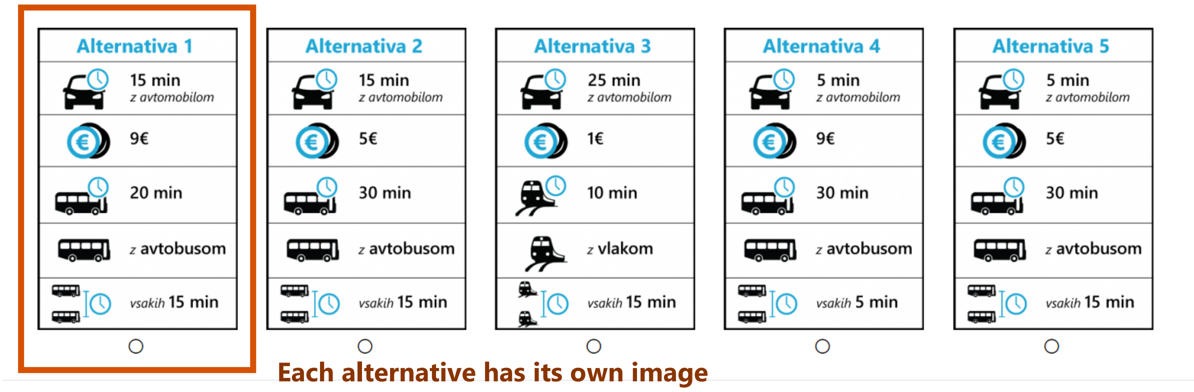


Figure 9.5. Layout 3 has individual images for each alternative, that disappear after being chosen

E Definition of prior parameter values

In their research Bos et al. (2004) investigated many attributes connected with P+R facility choice. Among them are also all five of the attributes to be used in this survey. The opportunity was taken to use these values as priors for constructing an efficient design. In order to do so, the parameters needed to be transformed from effects coded parameters, as done by Bos et al. (2004), to continuous linear parameters that will be used in this research.

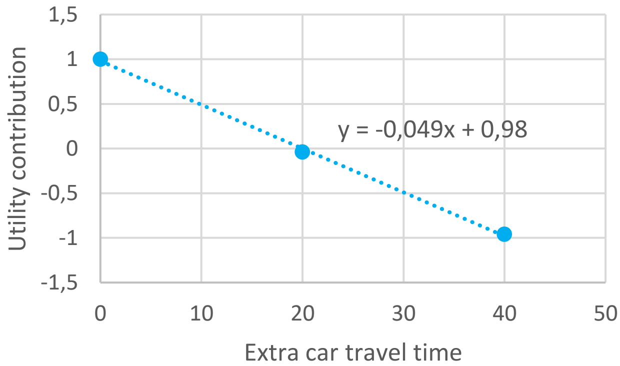
For each of the parameters, a chart was made with the attribute value on the x-axis and the utility contribution on the y-axis. A linear trendline was then added and the coefficient of that trendline taken as the linear continuous parameter. The calculations for each of the four attributes, with the attribute levels from literature, are presented in Table 9.6. An exception was made for the attribute of public transportation mode, as it will be coded as a dummy variable in this research as well. Bos et al. (2004) investigated the effect of three separate modes on consumer choice, namely bus, tram and train, but this research will only distinguish between bus and train. Obtaining the dummy parameter for the difference between train and bus is done in a very straightforward way, by simply looking at the difference in utility contribution: 0,17 (train has a contribution of 0,04 and bus has -0,13). Depending on the coding, the parameter is therefore either 0,17 or -0,17. Since the mode is coded as bus (0) and train (1), the prior parameter value is 0,17, indicating added utility if the public transport service is a train.

Table 9.6. Determination of parameter values

Car travel time

Extra car travel time	Utility contribution
0 min	1
20 min	-0,04
40 min	-0,96

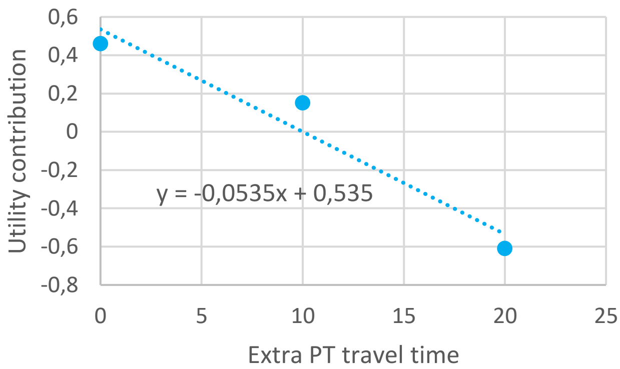
Parameter value: -0,049



Public transport travel time

Extra PT travel time	Utility contribution
0 min	0,46
10 min	0,15
20 min	-0,61

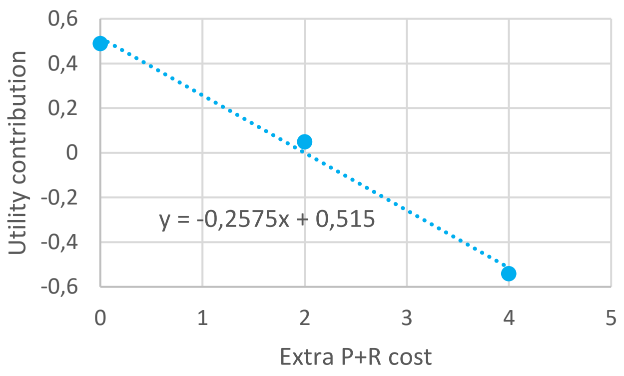
Parameter value: -0,0535



Cost

Extra P+R cost	Utility contribution
0€	0,49
2€	0,05
4€	-0,54

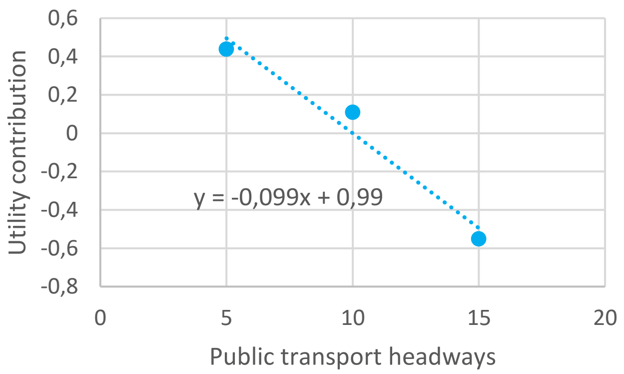
Parameter value: -0,2575



Public transport headway

PT headways	Utility contribution
5 min	0,44
10 min	0,11
15 min	-0,55

Parameter value: -0,099



F Ngene syntaxes

For constructing the experimental designs in Ngene, three separate syntaxes were used, one for each type of the design generated. The syntax used for the D-efficient design is presented in Figure 9.6, the Bayesian design with small standard deviations in Figure 9.7 and the Bayesian with large standard deviations in Figure 9.8. For further information on the used experimental designs, the reader is referred to chapters 2.4 and 3.3.2.6, as well as the research paper by Walker et al. (2018).

```
design
;alts = alt1, alt2, alt3, alt4, alt5
;rows = 12
;eff = (mnl,d)
;model:
U(alt1)= b1[-0.049] * Car[5,15,25] + b2[-0.0535] * PT[10,20,30] + b3[-
0.2575] * C[1,5,9] + b4[-0.099] * H[5,15,30] + b5[0.17] * Mode[0,1]/
U(alt2)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt3)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt4)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt5)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode
$
```

Figure 9.6. Ngene syntax for a D-efficient design

```
design
;alts = alt1, alt2, alt3, alt4, alt5
;rows = 12
;eff = (mnl,d,mean)
;model:
U(alt1)= b1[(n,-0.049,0.0163)] * Car[5,15,25] + b2[(n,-0.0535,0.01783)]
* PT[10,20,30] + b3[(n,-0.2575,0.0858)] * C[1,5,9] + b4[(n,-
0.099,0.033)] * H[5,15,30] + b5[(n,0.17,0.0567)] * Mode[0,1]/
U(alt2)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt3)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt4)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt5)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode
$
```

Figure 9.7. Ngene syntax for a Bayesian design with small standard deviations

```
design
;alts = alt1, alt2, alt3, alt4, alt5
;rows = 12
;eff = (mnl,d,mean)
;model:
U(alt1)= b1[(n,-0.049,0.0245)] * Car[5,15,25] + b2[(n,-0.0535,0.02675)]
* PT[10,20,30] + b3[(n,-0.2575,0.12875)] * C[1,5,9] + b4[(n,-
0.099,0.0495)] * H[5,15,30] + b5[(n,0.17,0.085)] * Mode[0,1]/
U(alt2)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt3)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt4)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode/
U(alt5)= b1 * Car + b2 * PT + b3 * C + b4 * H + b5 * Mode
$
```

Figure 9.8. Ngene syntax for a Bayesian design with large standard deviations

G Experimental design

Table 9.7 presents the selected design for the stated choice survey. There are a total of 12 choice sets, each containing five alternatives. These are characterised by five generic attributes, namely travel time by car (indicated as Car), travel time by public transport (PT), the cost of parking and public transport ticket (Cost), the headways of the public transport service (H) and the public transport mode (Mode).

Table 9.7. Experimental design used in the survey

Choice set	Alt.	Car [min]	PT [min]	Cost [€]	Headway [min]	Mode	Choice set	Alt.	Car [min]	PT [min]	Cost [€]	Headway [min]	Mode
1	1	15	20	9	15	0	7	1	25	30	1	5	0
	2	15	30	5	15	0		2	5	10	1	30	0
	3	25	10	1	15	1		3	25	20	5	15	1
	4	5	30	9	5	0		4	5	10	9	5	1
	5	5	30	5	15	0		5	15	20	5	30	1
2	1	15	20	9	15	1	8	1	5	30	1	30	1
	2	25	30	1	5	1		2	15	20	5	30	0
	3	5	10	5	15	0		3	5	30	9	5	1
	4	25	20	5	15	0		4	15	20	5	30	1
	5	15	20	5	30	1		5	25	10	1	15	0
3	1	5	10	9	5	0	9	1	25	30	5	5	0
	2	5	30	1	15	1		2	25	20	5	15	0
	3	25	20	5	15	0		3	5	30	1	30	0
	4	25	10	1	30	1		4	15	20	5	30	0
	5	25	20	9	5	1		5	5	10	9	5	1
4	1	15	10	1	30	0	10	1	5	10	1	30	1
	2	15	20	5	30	1		2	25	10	9	5	0
	3	15	20	9	5	1		3	15	30	9	5	0
	4	25	10	1	15	1		4	15	20	5	30	0
	5	5	30	1	5	0		5	15	30	5	30	1
5	1	15	20	5	15	0	11	1	25	30	5	5	1
	2	25	10	1	15	0		2	15	20	9	30	1
	3	15	20	9	5	0		3	25	10	5	30	1
	4	5	30	1	5	1		4	5	10	9	15	0
	5	25	10	9	5	1		5	5	30	1	30	0
6	1	25	20	5	15	1	12	1	5	10	9	30	1
	2	5	10	9	5	1		2	5	30	9	5	1
	3	5	10	1	30	0		3	15	30	1	30	1
	4	25	30	1	5	0		4	15	30	9	15	1
	5	15	20	9	15	0		5	25	10	1	15	0