

# VU Research Portal

## Preference uncertainty in stated choice experiments

Dekker, T.

2012

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Dekker, T. (2012). *Preference uncertainty in stated choice experiments*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## Preface and acknowledgements

With great satisfaction I look back upon five years of research that shaped this thesis. It feels good to end this period and celebrate its achievements, while still enjoying the lively environment called academia. Getting to the point of finishing this thesis would not have been possible without the help, insights, support and friendship of various people, and I wish to thank all of them here.

I would like to thank my supervisors Roy Brouwer and Marjan Hofkes for the freedom they gave me in exploring the topics that I considered to be the most interesting. I realise that applied (Bayesian) econometric analysis is not your core expertise, so my apologies for the effort you had to put in to read some of the material. Nevertheless, your comments have been helpful in making the thesis more accessible and putting things into perspective. I am also grateful for the numerous international conferences and courses I was able to attend.

The research visits to the Universities of Sydney, Leeds and Belfast have made working on this thesis a pleasant experience. In particular, I would like to thank David Hensher, John Rose and Stephane Hess for valuable discussions and providing an inspiring environment to work in. Klaus Moeltner has been extremely helpful in teaching me the basics of Bayesian econometrics and showing the importance of writing my own code. The many researchers I've met during these travels are acknowledged for their good company and open-mindedness. A direct word of thanks to Alejandro, Andrew, Janina, Jorge, Patrick and Vikki.

It has been a pleasure to work with all my former colleagues at the Institute for Environmental Studies. Social coffee breaks with Ralph, Laurens and Hans were much appreciated as well as the monthly game of football with many others. Marije, Luke and Wouter are thanked for their valuable comments, discussions and guidance on the thesis and other issues in general. Finally, working together with Paul has been a great experience. All in all good reasons to keep visiting the VU every now and then.

I am grateful to my friends and family for their support and distractions. Academic life is not what you are used to, and you were right that it's foolish to work during most of the weekends (and sometimes nights). Being abroad in Leeds makes me appreciate the value of having good friends and family around more and more. To my parents; Bram and Irene; Benno, Frank, Jeroen, Luuk, Martijn, René, Ruud and Theo: I couldn't have done all this without you.

The research described in this thesis was financially supported by the BSIK-KvR IC05 project under the Dutch National Research Programme 'Climate changes Spatial Planning'.

## Chapter 1: Introduction

### 1.1 Thesis focus

This thesis addresses the extent to which individuals are willing to pay for public safety programs reducing the risks of natural hazards. More specifically, we study individual willingness-to-pay (WTP) for flood risk reductions in the Netherlands using stated choice experiments and develop a set of new discrete choice models to control for the impact of preference uncertainty and related learning and fatigue effects on the welfare measure(s) of interest. The main contributions of this thesis are therefore methodological in nature. Researchers in alternative areas like health and transport economics, also using stated choice experiments, can easily adopt the set of developed methods given the generic properties of the data and econometric models. A wider set of empirical tests for heterogeneity and dynamics in response patterns across and within respondents is therefore provided by this thesis.

The application of stated choice experiments in the context of (public) safety programs is not straightforward. Natural hazards are complex phenomena of which even scientists do not fully understand the dynamics. They are unable to accurately predict when, where and at which intensity they will occur. In a stated choice experiment on flood risk reduction in the Netherlands most respondents are likely to have limited experience with the natural hazard and related safety trade-offs. Many, if not most, respondents can therefore not be treated as experts on this topic. These aspects are likely to result in what we label as preference uncertainty about appropriate risk reducing policies, and variation in risk perceptions across the population. Preference uncertainty may affect response patterns in stated choice experiments. Similarly, heterogeneity in preferences induced by variations in risk perception or other respondent characteristics affects response patterns. The validity and reliability of WTP estimates can be questioned if neither preference uncertainty nor heterogeneity in general is controlled for in the discrete choice model. Within this thesis we will analyze the impact of preference uncertainty and heterogeneity in response patterns across respondents in general on WTP estimates derived by stated choice experiments. The discrete choice models applied are specifically designed for this purpose.

The remainder of this chapter is organized as follows. Section 1.2 provides a brief introduction into flood risks and flood risk reducing policies in the Netherlands. In addition, it makes the reader familiar with Social Cost-Benefit Analysis (SCBA) as a tool for policy evaluation; and the role of non-market valuation techniques in SCBA. In Section 1.3 we

discuss the complexity of the decision environment faced by respondents in non-market valuation studies concerning natural hazards in more detail. Specific attention is paid to the role of preference uncertainty and how it may have an impact on the validity and stability of WTP estimates. Section 1.4 describes the main research questions and aims of this thesis. The outline of this thesis is presented in Section 1.5 and Section 1.6 summarizes its main contributions.

## **1.2 Background**

### **1.2.1 Flood risks in the Netherlands**

A large part (26%) of the Netherlands is situated below sea level and in total 55% of the country is vulnerable to flood risks (PBL, 2010). Flood prone areas are threatened by coastal floods from the North Sea and river floods from major rivers like the Rhine and Meuse. Like most natural hazards, flood risks in the Netherlands can be characterized as low probability – high impact events. Public safety structures and regulations limit the probability of a flood to once every 1250 years along the main rivers and up to once every 10,000 years in densely populated areas. Despite their low probability of occurrence, floods are likely to cause substantial damage due to the large share of the population and significant economic activity located in flood prone areas (Bouwer and Vellinga, 2007). Scientists predict rising sea levels and increasing peak discharge levels in rivers due to climate change. Accordingly, flood risks in the Netherlands are expected to increase over the coming decades. ‘Climate proofing’ the Netherlands requires additional investments in safety structures and alternative adaptation mechanisms (Kabat et al., 2005). In contrast to other European countries, private insurance mechanisms against river and coastal flooding are not (yet) available in the Netherlands (Botzen and Van Den Bergh, 2009). Dutch citizens mainly rely on the provision of public safety programs in their flood risk protection. Therefore, we are interested in the extent to which individuals are willing to pay higher taxes in order to prevent (or limit) the increase in flood risks due to climate change.

### **1.2.2 Social Cost Benefit Analysis, non-market valuation and stated choice experiments**

The benefits of flood risk reducing policies arise in preventing damage to people and property. A substantial part of these damages exceed the area of direct tangible damages, i.e. goods for which a market price exists. Loss of life, injuries and other personal inconveniences, but also damages to cultural heritage and natural parks are examples of impacts of floods for which a market price is lacking (Jonkman et al., 2008). For a complete

account of policy benefits the SCBA should also cover the benefits of preventing damage to these so-called non-market goods (or immaterial damage).

Deriving a monetary value for specific non-market goods or policies has received a lot of attention in different fields of economics, ranging from environmental and resource economics (e.g. Champ et al., 2005; Freeman, 2003) to health economics (e.g. Ryan et al., 2008) and transport economics (e.g. Tseng and Verhoef, 2008). The applied elicitation methods are all based on the analysis of self-expressed or empirically observed trade-offs by individuals between the non-market good and income (or other priced goods). By accepting or rejecting changes in the characteristics or price of the non-market good, the individual reveals to which extent (s)he is willing to pay money, or give up income, for an improvement in the non-market good. As such, non-market valuation techniques are able to link individual decisions to the public decision framework by estimating the (monetary) welfare impacts of policy changes (more details in Chapter 2).

To derive WTP estimates economists prefer the use of empirically observed trade-offs over self-expressed preference statements, since real-world behaviour (i) does not suffer from the well known hypothetical bias (Murphy et al., 2005) and (ii) is more likely to be incentive compatible (Collins and Vossler, 2009). Daniel et al. (2009) provide an example of such a revealed preference (RP) study in which the selling prices of various properties located along the Meuse river in the Netherlands are used to analyze the effect of differences in the spatial distribution of flood risks on housing prices. Though RP studies rely on real-world data, they are also associated with several drawbacks to which the use of hypothetical scenarios in stated preference (SP) studies are believed to offer more flexibility (see Sections 2.2.1 and 2.2.2). Stated choice experiments, as a specific form of SP studies, request the individual to select his (or her) most preferred alternative from a limited set of available alternatives. Each alternative is described as a package of attributes, of which the levels vary across alternatives (Hanley et al., 1998). The varying attribute levels force the individual to make a trade-off between, for example, flood probability, flood damage and taxes. Based on these trade-offs the researcher is able to estimate an individual's marginal WTP for changes in the specific attributes. In this thesis we conduct a stated choice experiment eliciting WTP for a public safety program reducing flood risk exposure in the Netherlands in the face of climate change.

### **1.3 Problem description**

Microeconomic theory assumes that individuals can identify their most preferred alternative in each choice task within the stated choice experiment. Section 1.1 already described

respondents are unfamiliar with the concept of (changes in) flood risk exposure and the impacts of a flood. On top of that, they lack experience with making trade-offs regarding their flood risk safety. Unfamiliarity and lack of experience are likely to undermine the core assumptions of the microeconomic framework. Not taking these complexities into account can influence the reliability of the WTP estimates and consequently the outcome of the SCBA and related policy decisions.

### **1.3.1 Complexities in applying stated choice experiments to risk valuation**

Natural hazards are characterized as low probability – high impact events. Individuals seldom encounter a flood, earthquake or hurricane (Kunreuther and Pauly, 2004). As a result, they experience difficulties in comprehending the associated small probabilities and lack a sensible estimate of the potential damages of such an event<sup>1</sup>. The lack of experience with natural hazards in combination with the public provision of safety programs is a cause of unfamiliarity with the presented trade-offs in stated choice experiments (Brown et al., 2008). Thereby, a heavy cognitive load is put on individuals when being asked to make a trade-off between alternative safety programs. Inexperience with the presented (hypothetical) market, payment vehicle, presented trade-off, specific good (i.e. flood risk) characteristics are all potential sources behind ill-defined individual preferences over the concept being studied in a stated choice experiment (Braga and Starmer, 2005; Lichtenstein and Slovic, 2006). Often this is labelled as respondent or preference uncertainty, which is the main topic of this thesis. Closely related is the impact of risk perceptions on stated preferences. Safety decisions are guided by perceptions of the probability and consequences of the threat to which individuals are exposed (Slovic, 1987). Heterogeneity in risk perceptions across the population of interest is likely to affect the extent to which individuals believe measures to reduce flood risk exposure are necessary and the extent they are willing to pay for it.<sup>2</sup> For example, specific individuals may believe policies should be directed towards reducing the probability of flooding, while others think the consequences of a flood should be minimized or both. In this thesis we do not specifically focus on heterogeneity in risk perceptions, but adopt a broader perspective in which preferences in general vary across individuals. Risk perception is one of

---

<sup>1</sup> Also timing is an element adding to the cognitive load, since it is unclear when floods will occur and most policies have a long time horizon. Note that in this thesis we will not discuss the issue of discounting (Dasgupta, 2008).

<sup>2</sup> Risk perception differs from an individual's attitude towards risks. Risk averse people experience disutility from being exposed to uncertain events. Perceptions are not so much about the impacts of being exposed to a particular risk, but rather they reflect the individual's personal (or perceived) image of the specific event (and its impacts) (Huber and Huber, 2008; Sjöberg, 2000).

many factors that may explain such preference heterogeneity. Consequently, the extent to which researchers are able to derive informative and reliable WTP estimates depends on their ability to reduce and control for the complexities highlighted above, i.e. preference uncertainty and heterogeneity.

### **1.3.2 The challenges raised by the complexities**

The complexity of the non-market good studied in a stated choice experiment affects the extent to which respondents are able to make informed decisions. If respondents do not fully understand (changes in) the non-market good, in our case flood risk exposure, WTP estimates provide little information about the real value of flood risk reductions (Carson and Groves, 2007). During the development of the SP survey the researcher can already take into account some of these complexities. By gradually informing the respondent about the concept of flood risk, the researcher can make the respondent aware of the risks (s)he is exposed to and take away (some of) the uncertainties related to the proposed policies themselves. Adequate pre-testing allows developing effective communication tools. For example, Corso et al. (2001) explored various graphical tools for communicating small mortality risk probabilities to respondents. Hence, properly explaining the concept of flood risks and the hypothetical market constructed in the survey may improve the understanding of respondents and the extent to which they are able to make informed decisions. Despite these efforts, respondents may still face a considerable cognitive load due to the inherent complexity of flood risks. Consequently, respondents may still perceive flood risks differently and they may be left with a remaining degree of preference uncertainty. Biased WTP estimates are likely to be obtained if preference heterogeneity and the remaining degree of preference uncertainty are neglected in the analysis of observed responses in the stated choice experiment.

The micro-econometric models applied in the discrete choice modelling literature rely on the specification of a utility function for the alternatives available in the choice set. Preference uncertainty causes these utility functions to be ill-defined. In the most basic interpretation of the Random Utility framework (e.g. Li and Mattsson, 1995), uncertain respondents are expected to make more random decisions compared to certain respondents. That is, they are unable to distinguish one alternative from another and base their decision on some unobserved element that is unrelated to the decision at hand. The latter is commonly modelled by imposing high levels of variance on the utility function. Alternatively, respondents may look for simplifying decision rules to reduce the complexity of the choice task (Hensher, 2010a). Different simplifying decision strategies require the specification of



alternative utility functions. Hence, properly taking into account preference uncertainty requires the application of flexible discrete choice models controlling for alternative specifications of the shape and variance of the utility function (e.g. Fiebig et al., 2010; Greene and Hensher, 2010). State of the art discrete choice models are able to control for variations in the variance of the utility across respondents. Moreover, new and more flexible model structures are being developed and applied at a rapid pace across different fields in the economics of non-market valuation, enabling researchers to increase the reliability of their WTP estimates (Hess and Daly, 2010). This thesis contributes to this line of research in various ways (see Section 1.6).

### **1.3.3 Gaps in the literature**

The econometric models introduced at the end of Section 1.3.2 focus on observed and unobserved variations in preference parameters and variance of the utility function across respondents. Indeed, respondents differ from each other in terms of their preferences, which should be taken into account. However, several gaps can still be identified in the current discrete choice modelling literature. First, Greene and Hensher (2003) note that in accounting for respondent heterogeneity researchers are often faced with a trade-off between behavioural relevance and ease of estimation. More specifically, they stress the importance of discrete choice models that ensure behavioural relevance while maintaining a flexible structure. More recently, we have observed a tendency towards more flexible distributional forms capturing heterogeneity in choice patterns across respondents (e.g. Burda et al., 2008; Fosgerau and Hess, 2009; Train, 2008). While flexibility does not necessarily conflict with behavioural relevance, it certainly reduces the ease of estimation. Moreover, the impact of these alternative specifications on the welfare estimates of interest is unclear.

Second, limited attention has been paid to the dynamics in individual choice behaviour as respondents proceed through a sequence of choices (e.g. Brouwer et al., 2010; Hess and Rose, 2009; Holmes and Boyle, 2005). Respondents are likely to refine their preference structure as well as their decision criteria while making a sequence of choices. Learning effects may result in alternative decision strategies, but most importantly better informed decisions (Hensher, 2010a). In long surveys, fatigue effects may occur and result in reduced attention to the choice task at hand and thereby reduce the information content of the observed decisions at the end of the stated choice experiment. Such dynamics in choice behaviour over the choice sequence may have an impact either on the shape or the scale of an individual's utility function. Discrete choice models that simultaneously allow for flexibility in preference

and scale parameters along the dimensions of respondents and choice sequence are still underdeveloped. Moreover, embedding dynamics in choice behaviour over the choice sequence in the recent state of art models has only just set off.

Third and closely related to these dynamics in preference and scale parameters over the choice sequence is the issue of self-reported preference certainty. The former implicitly identifies learning and fatigue effects through a flexible modelling structure. The latter exploits the possibility of stated choice experiments to use explicit follow-up questions to acquire additional information about individual choice behaviour and can serve as an additional measure of observed learning and fatigue effects. Within the contingent valuation literature, a large amount of papers exists relating decision uncertainty to WTP uncertainty (see Akter et al., 2008 for an overview). Many of the modelling approaches developed in the contingent valuation literature on preference uncertainty are not directly transferable to the stated choice experiment literature due to the presence of multiple alternatives in most stated choice experiments (Lundhede et al., 2009). Moreover, it remains unclear whether these implicit and explicit modelling approaches to preference uncertainty reinforce each other and hence whether such follow-up questions can be used to derive more reliable WTP estimates.

#### **1.4 Main objective, central hypothesis and research questions**

Given the large welfare implications of changes in flood risk exposure in the Netherlands and the need to quantify, or value, these changes for a proper evaluation of policy proposals using SCBA, the aim of this thesis is to improve the validity and reliability of WTP estimates by improving existing discrete choice models. Improvements in discrete choice models are presented in line with the three identified gaps in the literature and focus on (i) accounting for respondent heterogeneity, (ii) implicit modelling of preference uncertainty, related learning and fatigue effects by allowing preference and scale parameters to vary over the choice sequence, and (iii) explicit modelling of preference uncertainty using responses to self-reported choice certainty follow-up questions. As such this thesis addresses the following three main research questions:

- 1. How can we account for heterogeneity in response patterns across respondents in discrete choice models in a flexible and behaviourally relevant way, without increasing the complexity of estimation? How sensitive are willingness-to-pay estimates to alternative specifications of preference heterogeneity in discrete choice models?*

2. *To what extent is respondents' choice behaviour subject to preference dynamics, i.e. learning and fatigue effects, over the choice sequence? And does this have an impact on willingness-to-pay estimates?*
3. *Do self-reported choice certainty follow-up questions offer a useful tool to improve willingness-to-pay estimates derived from stated choice experiments?*

The central hypothesis underlying this thesis is that by accounting for preference uncertainty and related preference dynamics, the validity and reliability of WTP estimates for flood reductions can be increased. This hypothesis is tested by (i) implicit modelling of dynamics over the choice sequence in preference and scale parameters in addition to accounting for respondent heterogeneity; and (ii) explicit modelling of preference uncertainty and related learning and fatigue effects by studying measures of self-reported preference certainty. These models are applied to a case study on flood risk valuation in the Netherlands.

### **1.5 Thesis outline**

This thesis consists of eight chapters and is split into three parts. The first three chapters comprise Part 1, which describes the aim of this thesis (Chapter 1), embeds the thesis into the literature on non-market valuation and discrete choice modelling (Chapter 2), and covers the existing theoretical and empirical literature related to the implicit and explicit modelling of preference uncertainty (Chapter 3). Chapter 3 also provides an overview of empirical results regarding the impact of preference uncertainty on willingness-to-pay estimates. Part 2 describes the case study (Chapter 4) and develops and tests new, more advanced, discrete choice models (Chapters 5-7). The case study is specifically designed to test our research questions. Chapter 5 is concerned with heterogeneity in response patterns across respondents. Within the random parameters logit framework alternative specifications of the mixing densities are estimated. While tracing the impact of these specifications on willingness-to-pay estimates, all specifications are evaluated in terms of behavioural relevance, flexibility and ease of estimation. Within this setting a new mixing density is introduced and evaluated. In Chapter 6 we develop a new model to control for preference dynamics in stated choice experiments, which is more suitable than the commonly applied Swait and Louviere (1993) test. In an empirical application we test for the presence and decay of a starting point bias. As such we contrast the Discovered Preference Hypothesis (Plott, 1996) and the theory of coherent arbitrariness (Ariely et al., 2003), which are both contrasting hypotheses on preference dynamics over the choice sequence. Chapter 7 introduces an integrated framework

to learn about preference uncertainty by combining responses to the stated choice experiment and the choice task specific self-reported preference certainty questions. We test whether preference certainty affects willingness-to-pay estimates and how both response formats relate to each other and should be treated in the discrete choice model. Finally, Chapter 8 forms the final part of the thesis and provides a synthesis of the work presented in Chapters 5-7 and presents the main conclusions of this thesis.

## **1.6 Methodological contributions**

Here, we briefly list the main methodological contributions provided in this thesis.

1. In contrast to the dichotomous choice contingent valuation literature, stated choice experiments require respondents to make trade-offs between multiple attributes and multiple alternatives. As such, preference uncertainty can arise both at the level of the alternative and the attribute. We label these as respectively package and trade-off uncertainty. Consequently, not only the scale of the utility function is affected by preference uncertainty (package uncertainty), but also the marginal rates of substitution (trade-off uncertainty). A detailed discussion is provided in Chapter 3.
2. We introduce the asymmetric triangular distribution into the random parameter logit framework and estimate it in a Bayesian fashion. By being restricted to either the positive or negative domain, the distribution produces behaviourally relevant WTP distributions without inducing fat tails. As such, WTP intervals become more valid and reliable. Moreover, by estimating the mode independently from the lower and upper bound, the mixing density has a relatively flexible shape. Chapter 5 discusses whether the new density also results in an improvement in model fit and ease of estimation.
3. Chapter 6 introduces a semi-parametric model, labelled the local multinomial logit model, to control for dynamics in preferences over the choice sequence. By smoothing parameter estimates, the model is much more flexible than the generally applied Swait and Louviere (1993) test procedure. Additionally, the model results in a substantial improvement in efficiency of parameter estimates relative to a set of choice task specific parameter estimates. Furthermore, it can take into account that preferences adjust gradually over the choice sequence.
4. In Chapter 7, we treat preference certainty as a latent variable which affects both the responses to the choice model and the self-reported choice certainty follow-up questions. This resolves some of the possible endogeneity issues being present within the existing literature on the impact of preference certainty in stated choice experiments.



## Chapter 2: Non-market valuation and stated choice experiments

Flood risk reducing policies are designed to prevent the occurrence of one or more floods over a period of time. Reductions in exposure to flood risks are commonly measured by (i) reductions in flood probability and (ii) reductions in expected damages in case of a flood. The benefits of flood risk reducing policies therefore consist of prevented damages. The benefits of prevented damages often extend beyond the range of direct tangible and priced damages, e.g. damage to residences and other property (Brouwer and van Ek, 2004; Jonkman et al., 2008). For example, floods are also associated with loss of life, public health impacts and environmental damages.<sup>3</sup> Non-market valuation techniques are available to assign a monetary value to such indirect and intangible non-priced damages, such that the full monetized benefits of flood risk reducing policies can be contrasted to the monetary costs of implementing them in SCBA. This chapter provides a brief introduction into the theory of non-market valuation, explains the main valuation methodologies available and discusses how they can be applied in the case of flood risk reducing policies.

### 2.1 Choice of welfare measure

Non-market valuation rests on the assumption that despite the fact that a market price maybe lacking for a particular good or policy, it can still have an impact on individual well-being. Accordingly, people are likely to take the provision level of the non-market good into account while making decisions. In terms of flood risks, risk averse people usually prefer to live in areas where they experience limited or no exposure to flood risks. Due to its impact on individual well-being, flood risk exposure can be incorporated as an argument in the individual's utility function. The utility function is assumed to be a complete representation of the individual's preferences. By doing what (s)he likes most, the individual maximizes utility (Flores, 2003; Rieskamp et al., 2006). Utility maximization is subject to income and time constraints, which requires the individual to make trade-offs between alternative goods providing utility (Perman et al., 2003). In the case of flood risks, risk averse people minimize the disutility derived from flood risk exposure or maximize the utility from a reduction in flood risk exposure. Non-market goods have an implicit value due to the fact that individuals are willing to give up money, income, or any other priced good, for improvements in the non-market good. Based on this trade-off, non-market valuation methods derive willingness-to-

---

<sup>3</sup> Additional indirect impacts on individuals and society are discussed, for example, in (Pearce and Smale, 2005) and Jonkman et al. (2008).

pay (WTP) or willingness-to-accept (WTA) compensation values for non-market goods (Freeman, 2003). More formally and without loss of generality, it can be assumed that utility  $U_i$  for individual  $i$  increases from  $U_{i0}$  to  $U_{i1}$  if (s)he experiences a reduction in flood risks. Our interest is in finding a monetary measure for the resulting increase in utility.

### 2.1.1 Equivalent surplus and compensating surplus

Two Hicksian welfare measures, the compensating surplus (CS) and the equivalent surplus (ES), can be used to derive the WTP or WTA measure of interest. The CS measures the maximum WTP of the individual to obtain the improvement in the non-market good, i.e. the amount of income that can be taken from the individual such that, conditional on the improvement, his utility level remains at the same level as before the improvement.<sup>4</sup> The ES measures the minimum amount of compensation required by the individual if the improvement does not occur. The difference between the two measures lies in the entitlement of property rights. If the individual does not have a right to the environmental improvement, the reference utility level is  $U_{i0}$ , i.e. the level before a change in the supply of the non-market good, and CS (or WTP) applies. In the opposite case, the individual is entitled to  $U_{i1}$  and ES (WTA) is more appropriate. Conversely, in case of an environmental deterioration, utility falls from  $U_{i0}$  to  $U_{i2}$ . The same reference utility levels apply based on entitlement, respectively  $U_{i0}$  for CS and  $U_{i2}$  for ES. The CS now measures the minimum amount of compensation (WTA) required to keep the individual at  $U_{i0}$  given that the deterioration occurs. The ES measures the maximum amount the individual is willing to pay to prevent the decrease in environmental quality (Perman et al., 2003). Table 2.1 presents an overview of the discussed monetary measures.

**Table 2.1: Welfare measures associated with changes in the quality and supply of non-market goods**

Case:	Consumer Surplus (CS):	Equivalent Surplus (ES):
Increase in quality or supply (Welfare improvement)	WTP to accept the change	WTA compensation to abstain from the change
Decrease in quality or supply (Welfare deterioration)	WTA compensation for the change	WTP to prevent the change

### 2.1.2 Welfare measure for flood risk valuation in the Netherlands

In the case of flood risk valuation, selection of the most appropriate welfare measure is somewhat complicated. Flood risks in the Netherlands are expected to increase due to climate

<sup>4</sup> It is assumed that utility decreases if disposable income is reduced.

change, thereby decreasing individual utility from  $U_{i0}$  to  $U_{i2}$ . By law specific flood probability levels have been defined for specific areas in the Netherlands (e.g. Bouwer and Vellinga, 2007; see Chapter 4 for more details). As such, it can be argued that respondents are entitled to  $U_{i0}$  and that the government either needs to prevent the increase in flood risks or compensate the residents of these areas for the increase in flood risks. This is in line with a CS welfare measure based on the concept of WTA compensation for the increase in flood risks. These safety standards, however, only serve as a guideline for public policy. Moreover, the Dutch government examines the possibilities to shift the responsibility for flood risk safety from the public domain towards a shared responsibility between public and private stakeholders (Botzen, 2010). These developments suggest that Dutch citizens are not necessarily entitled to the same level of flood risk protection as they were before climate change. On the contrary, in the near future they may need to contribute to additional flood risk reductions either through taxes or insurance payments. This perspective can be best interpreted as an ES WTP approach to prevent an increase in flood risks. The case study presented in this thesis indeed adopts a WTP approach, but takes the scenario under climate change ( $U_{i2}$ ) as the reference point. We examine to which extent respondents are willing to pay for reductions in flood risks relative to this reference point.<sup>5</sup> Therefore, the CS WTP approach is adopted in this thesis.

## **2.2 Valuation methodologies**

As noted in Section 1.2.2, Revealed Preference (RP) and Stated (SP) Preference methods can be used to obtain WTP estimates for reductions in flood risk exposure. Both methods use distinct data sources. RP studies rely on choices made by individuals in existing markets, for example the housing market (Daniel et al., 2009), while SP studies rely on hypothetical choices by respondents in surveys (Champ, Boyle and Brown, 2005). In the former case, the value people attach to flood risk exposure or changes herein is derived from actual purchasing behaviour. In the latter case, respondents state, for example, whether they are willing to move to a specific house, conditional on the level of flood risk exposure, the price and other characteristics of the house. The decision is, however, hypothetical not real. In this section we discuss the existing non-market valuation methodologies in more detail, either using RP or SP methods, and evaluate their applicability for flood risk valuation.

---

<sup>5</sup> Different welfare measures were tested in the pre-testing stages of our survey. Respondents considered the CS WTP formulation as the most comprehensible one.



### **2.2.1 Revealed preference methods**

Revealed preferences build on the assumption that the non-market good is intrinsically related to a traded commodity, which is also known as ‘weak complementarity’ (Perman et al., 2003). Moving decisions, for example, are affected by the number of bedrooms and the quality of the kitchen in the house. Related factors, like the availability of parks and quality of the neighbourhood, also play a role in the decision. Altogether these attributes or characteristics describe the entire house. Hedonic pricing (and wage) studies use the divisibility of a commodity (or job) into identifiable attributes to explain price (wage) differences. Thereby, the hedonic pricing method can indirectly identify to which extent an individual is willing to pay for instance, for an additional bedroom or a lower level of flood risk exposure.

The main advantage of the RP approach is its reliance on actual market data. Consumers experience the consequences of their decisions and therefore have an incentive to optimize their choices. Simultaneously, the RP method also takes into account personal and market constraints on income, accessibility and technology. However, the use of RP methods may come with considerable drawbacks (Hensher et al., 2005). First, the empirical identification of a WTP estimate for changes in a particular attribute requires sufficient variability in the attribute of interest. Often there is limited variability in attribute levels across observations. Even in cases of sufficient variation, multiple attributes frequently show high degrees of correlation, hampering empirical identification. Second, the range of attribute levels is restricted by what is observed in the market. The effect of nonexistent alternatives and attribute levels on WTP cannot be directly estimated, only extrapolated under the assumption that the same relationships and parameters apply to the unobserved part of the utility function. Finally, RP methods mainly rely on information extracted from chosen alternatives, hereby not considering information contained in non-chosen alternatives causing possible sample selection bias. Some of these drawbacks are particularly relevant for the valuation of flood risk reductions.

The Dutch housing market is not ideal for RP studies, because market prices may not represent true shadow prices. The supply of tenant housing is highly regulated and its demand heavily subsidized (CPB, 2008). Simultaneously, the demand for real estate is subsidized by paying a tax refund on the interest payments of the mortgage. Except for the housing market there are no other private markets available in the Netherlands through which individuals can reveal their preferences with respect to their preferred level of flood risk exposure. Flood risk safety is primarily a public responsibility. Specific areas, so-called ‘dike rings’, have been assigned in the coastal areas and along the major rivers with a particular safety level set by the

Flood Protection Act (Stb, 1996). Dikes form the major flood defence systems of these dike rings and they have been designed according to the dike ring specific “design discharge levels” (see Chapter 4 for more details). Since the official flood risk probabilities do not vary within dike rings, there is a limited variation in observed flood risk levels. Moreover, flood risk probabilities are likely to be highly correlated with other regional characteristics, such as population density and economic activity. Flood probabilities are expected to rise due to climate change, which may result in probabilities that fall beyond the range of flood probabilities currently defined in the Flood Protection Act (Maaskant et al., 2009). These arguments confirm the limitations of RP methods as denoted by Hensher et al. (2005).

Two additional arguments for not using RP methods are discussed by De Blaeij (2003). Individuals take into account their perceived risk levels instead of actual risk levels when making safety decisions, while most RP methods assume that individuals act according to actual (observed) risk levels. Also the variability in risk levels, for instance mortality risk in road safety, is very small across alternatives, which makes it hard for individuals to take variation in risk levels into account during the decision making process. The latter argument is easily transferrable to the case of flood risk valuation in which respondents are faced with flood probabilities of once every 10,000 years.

### **2.2.2 Stated preference methods**

Stated preference (SP) methods apply surveys to directly obtain information about the non-market good of interest. The survey approach in SP studies also allows for incorporating questions on individual risk perceptions and providing training opportunities to respondents in order to better comprehend the proposed change in the non-market good. From this perspective, SP methods can overcome the drawbacks of RP studies highlighted above. In addition to the direct contact with the respondent, the researcher is also able to describe and where necessary modify the description of the non-market good, including its attributes and attribute levels, and the institutional settings under which the non-market good is provided (e.g. Brouwer and Akter, 2010). The hypothetical nature of SP studies enables the researcher to introduce sufficient and uncorrelated variability in attribute levels while designing the non-market good. Hence, WTP can be identified more independently from other attributes relative to RP methods. Non-existent alternatives and attribute levels can be defined, as physical and technological constraints observed in the actual market no longer apply. The validity of SP methods has been questioned already since the publication of the contingent valuation manual by Mitchell and Carson(1989). Biases in WTP estimates may arise due to issues related to the

survey instrument and its implementation (Carson et al., 2001); so-called ‘anomalies’ caused by the bounded rationality of the respondent (Braga and Starmer, 2005); and lack of incentives for respondents to reveal their true preferences (Carson and Groves, 2007). This thesis aims to improve the validity and reliability of WTP estimates by assessing and accounting for preference uncertainty as a form of bounded rationality.

SP methods can be split up into two main categories: the Contingent Valuation Method (CVM) and Conjoint Analysis (CA). The CVM elicits preferences for (a change in) a specific good or policy. A properly formulated CVM study describes the current status of the non-market good, i.e. the status quo, and the proposed change in its provision level along with the institutional setting and the payment vehicle. The most simple elicitation format used in the CVM literature is the open-ended (OE) question, which directly asks the respondent to state his maximum WTP for the presented policy scenario. Another elicitation format is the payment card (PC) method in which the respondent can select his maximum WTP by picking a price level from a list with bid levels. A third elicitation format applied in the CVM literature is the dichotomous choice (DC) question, which asks the respondent whether he is willing to pay a specific price  $X$  for the proposed policy (change) or not.<sup>6</sup> If the bid amount is accepted (rejected) the interviewer can ask a second DC question with a higher (lower) price for the same good. In the latter situation the elicitation format changes from a single bounded to a double bounded DC question format, which provides more information on the respondent’s maximum WTP. When even more DC questions are added, the elicitation format is labelled as an iterative bidding game. In that case, the new price level is raised or lowered, conditional on acceptance or rejection of the previous bid, until the respondent’s maximum WTP is reached. However, the incentive compatibility of repeated DC questions is questionable (Carson and Groves, 2007).

While the CVM focuses on a single ‘holistic’ policy scenario, CA confronts respondents with two or more alternative policy scenarios with varying characteristics, generally referred to as attributes. Depending on the elicitation format, preferences can be expressed by selecting the most preferred alternative or by ranking or rating the available alternatives in the choice set. Note that the DC-CV elicitation format can be considered as a specific case of CA, where the proposed policy is only contrasted against the status quo. The price of the proposed policy varies across respondents both in the DC-CV format and in CA.

---

<sup>6</sup> The binary response format in DC-CVM studies closely corresponds to real market situations in which respondents buy goods (or not) at a given price level. Therefore, this approach was endorsed by the NOAA Blue Ribbon Panel (Arrow et al., 1993).

In contrast to the CVM method, which directly estimates WTP for a complete policy package, CA indirectly estimate the WTP for a policy package by calculating the change in consumer surplus from the status quo to the proposed policy. These calculations are usually based on the marginal WTP estimates for and changes in the level of specific policy attributes. By making a sequence of choices, respondents reveal the extent to which they are willing to give up one policy attribute for another, i.e. their marginal rate of substitution (MRS). If price is included as a policy attribute, a monetary marginal WTP value for specific policy attributes can be derived (Hensher, Rose and Greene, 2005).

Four different CA elicitation formats are commonly applied.<sup>7</sup> In a stated choice experiment (SCE) a respondent is presented with a limited set of alternatives. Each alternative is described by a set of common attributes, which vary over the alternatives available in the choice set. First, in the most common elicitation format the respondent is requested to select his preferred alternative. In order to optimally estimate the relationships between the different attributes, experimental designs are used to construct sufficient variation in attribute levels across choice tasks (e.g. Ferrini and Scarpa, 2007; Rose and Bliemer, 2009). Within the environmental valuation literature a status quo option is frequently included in the experimental design. By giving the respondent the opportunity to maintain the current situation, the respondent is allowed to express a zero (or protest) WTP. Not including the status quo option forces the respondent to always accept a change in provision levels, which may not always be realistic or preferred. As a result, WTP estimates are biased upwards when a status quo is not included. Moreover, inclusion of the status quo option allows for derivation of theoretically valid welfare measures for inclusion in SCBA (Boyle et al., 2001).

Second, conjoint ranking studies ask the respondent to rank the available alternatives in the choice set in line with their preferences. This method offers more information to the researcher, since the full preference ordering is obtained. WTP estimates derived from the conjoint ranking format are not necessarily in line with utility theory if the status quo is not included in each choice card. When the status quo is not included, the ranking of alternatives only provides information on demand conditional on the other alternatives in the choice set. Some (or all) of these options may, however, be valued worse than the status quo and therefore fall out of the respondents real demand curve (Louviere et al., 2003).

Third, in the conjoint rating format, the respondent reports his preference for each alternative in the choice set on a specified scale (often 0-10). Thereby, the researcher obtains a

---

<sup>7</sup> Examples for each format can be found in (Hanley et al., 2001).

preference ordering between the alternatives in the choice set and a degree of relative preference strength. Despite being more informative than the conjoint ranking format, it remains unclear how these different ratings can be compared across individuals. Strong assumptions are required to transform the rating scale into a utility measure. Hanley et al. (2001) argue that these assumptions are inconsistent with standard utility theory by definition. The method is therefore hardly used in environmental economics.

Finally, the paired comparison method is similar to the DC-CVM format by presenting the respondent with only two alternatives in each choice set. Like in a SCE, the alternatives in the choice set are allowed to vary on multiple attributes, not only price. Again, the respondent needs to select the most preferred alternative. By systematically contrasting all the policies in the design, a preference order can be established for each respondent. The paired comparison method enables the researcher to test for inconsistencies in the individual's choices (e.g. Brown et al., 2008). In order to obtain valid and reliable WTP estimates, a monetary attribute needs to be included in the design and one of the alternatives needs to be constant across all choices situations. For welfare consistent WTP estimates, this needs to be the status quo (Hanley, Mourato and Wright, 2001). Hence, the paired comparison method reduces to the DC-CVM format and no longer generates the complete preference order over all alternatives. The only difference is that all attributes vary across choice sets, not just price.<sup>8</sup>

### **2.2.3 Elicitation format selection for flood risk valuation**

Both RP and SP methods are associated with several advantages and disadvantages. While RP methods rely on actual behaviour, their applicability to non-existing situations is limited. SP methods overcome these limitations of RP methods by operating in a more flexible (hypothetical) environment, which is fully controlled by the researcher. However, this hypothetical environment may introduce 'anomalies' in behaviour that question the validity of derived WTP estimates (Braga and Starmer, 2005). Overall, the lack of available real market trade-offs on flood risks, the difficulty to introduce non-existent attribute levels and alternatives, and the limited possibility to control for risk perception are important reasons to select the SP method for this thesis. The SP method with its potential drawbacks is also selected with the aim to improve SP methods by controlling for relevant aspects of behaviour related to flood safety decisions, in particular preference uncertainty.

---

<sup>8</sup> As an extension (and different to a stated CE), the strength of preference between the two alternatives in the choice set is frequently measured on a numeric or semantic scale. This method is also known as graded comparison or rated pairs. However, like the conjoint rating exercise, it is difficult to interpret the strength of preference measure across individuals questioning the welfare consistency of the method.

Flood risks are multidimensional events which can be characterized by their probability of occurring and related consequences. Safety programs can target these specific elements, but most likely a policy mix is established. For example, part of the investments can be directed at strengthening and heightening the dikes, while at the same time money is also spent on early warning systems and spatial planning to improve the effectiveness of evacuation schemes. CA provides information on the overall benefits of the safety program and its decomposition into contributions of specific policy attributes. The latter can be useful information for the policy maker in optimizing the safety programs and provides a clear benefit of CA over the CVM.<sup>9</sup> Therefore, a stated choice experiment to value reductions in flood risk exposure is applied in this thesis. Welfare estimates derived by means of SCE's are consistent, i.e. in line with, standard utility theory as long as a status quo option is included in the design of each choice card. Alternative CA preference elicitation formats are not applied, because resulting welfare measures are not necessarily welfare consistent.

### 2.3. The random utility model

The decision framework underlying stated choice experiments can be traced back to Lancaster's (1966) multi-attribute value theory and the random utility model (RUM) (Manski, 1977). The RUM states that the level of utility  $U_{ijt}$  individual  $i$  obtains from alternative  $j$  in choice situation  $t$  can be decomposed into a 'systematic' utility component  $V_{ijt}$  and a stochastic component  $\varepsilon_{ijt}$ . Lancaster's theory of value enters the systematic part of utility by assuming that utility is a function of the row vector of explanatory variables  $X_{ijt}$ , including the policy attributes. The most common functional form for  $V_{ijt}$  is linear in the parameters  $\beta$  and uses  $X_{ijt}$  as inputs. The stochastic component is included in the model, because not all factors affecting utility may be observable to the researcher. Respondents with similar observable characteristics may still make different decisions in the same choice set  $D_{it}$  due to these unobservable characteristics. Since the researcher only observes choices, utility for each alternative becomes a latent variable in (2.1).

$$(2.1) \quad U_{ijt} = V_{ijt} + \varepsilon_{ijt} = X_{ijt}\beta + \varepsilon_{ijt}$$

---

<sup>9</sup> In the socio-psychological literature it appears that lay people usually do not make the explicit distinction between probability and consequence and base their decision primarily on the perceived consequences of a specific risk rather than the perceived probability of occurrence (e.g. Sjöberg, 2000).

Respondents are assumed to maximize utility and therefore select alternative  $j$  if and only if the utility derived from alternative  $j$  is higher than the utility derived from any other available alternative in choice set  $D_{it}$ . Due to the inclusion of a stochastic component, utility becomes a random variable. Accordingly, the probability of preferring alternative  $j$  over alternative  $k$  can be described by (2.2).

$$(2.2) \quad P(U_{ijt} > U_{ikt}) = P[(V_{ijt} - V_{ikt}) > (\varepsilon_{ikt} - \varepsilon_{ijt})] \quad j, k \in D_{it}; j \neq k$$

It is commonly assumed that the random component of utility is independently and identically distributed (i.i.d.) over individuals, alternatives and choice situations following a type I extreme value (Gumbel) distribution with variance  $\sigma_\varepsilon^2 = \pi^2/6$  (Ben-Akiva and Lerman, 1985). Its cumulative probability density function  $F(\cdot)$  is described by (2.3). Given its fixed variance, the scale of the utility function is also automatically normalized. Denote  $\lambda$  as a non-negative scale-parameter that adjusts the scale of the utility function. It is inversely related to the variance of the random component in the following manner  $\lambda = \sqrt{\pi^2/6\sigma_\varepsilon^2}$  such that in the normalized case  $\lambda=1$  (Louviere, Hensher and Swait, 2003).

$$(2.3) \quad F(\varepsilon_{ijt}) = \exp[-\exp(-\varepsilon_{ijt})]$$

This distributional specification implies that the utility difference between two alternatives follows a logistic distribution. As a result the probability that individual  $i$  selects alternative  $j$  in choice situation  $t$  can be described by (2.4), which is also known as the multinomial logit model (McFadden, 2001). It becomes directly apparent that decisions are independent of scale since  $\lambda$  cannot be separately estimated from  $\beta$ , explaining its normalisation to 1 (Train, 2009). Below  $y_{it}$  denotes the observed choice for individual  $i$  in choice task  $t$ .

$$(2.4) \quad P(y_{it} = j | X, \beta) = \frac{\exp(\lambda V_{ijt})}{\sum_{k \in D_{it}} \exp(\lambda V_{ikt})} = \frac{\exp(\lambda (X_{ijt} \beta))}{\sum_{k \in D_{it}} \exp(\lambda (X_{ikt} \beta))}$$

Note furthermore that when taking the ratio of choice probabilities of selecting, for example, alternatives  $j$  and  $k$  the (common) denominator cancels out. Consequently, the relative probability of choosing alternative  $j$  over alternative  $k$  is independent of the other alternatives available in the choice set. The latter is also called the independence of irrelevant alternatives (IIA) axiom (Ben-Akiva and Lerman, 1985). It is a restrictive assumption, because it implies proportional (or symmetric) substitution among all alternatives in the choice set. More advanced econometric models, like the multinomial probit, nested logit and mixed (or random parameters) logit model, are available to relax the IIA property amongst other issues.<sup>10</sup> We discuss the latter type of models in later chapters.

The marginal effect of each policy attribute on utility can be calculated using partial derivatives. In our linear utility specification, this is simply the coefficient  $\beta_a$  associated with the relevant attribute  $a$ . The MRS between two attributes is then defined by the (negative) ratio of their parameter estimates. Accordingly, marginal WTP  $\omega_a$  for a specific attribute  $a$  can be calculated by taking the negative ratio of the attribute's coefficient and the cost coefficient  $\alpha$  as shown in (2.5). The latter is assumed to represent the marginal utility of income. Furthermore, aggregate WTP (or the change in consumer surplus) can be calculated based on the complete estimated utility function by simply including the relevant levels for all attributes. Within this thesis we primarily focus on marginal WTP estimates.

$$(2.5) \quad \omega_a = -\frac{\beta_a}{\beta_{\text{cost}}} = -\frac{\beta_a}{\alpha}$$

## 2.4 Summary

In this chapter we have discussed the theoretical framework underlying the concept of non-market valuation. We reviewed potential methods to derive a monetary value for the benefits of flood risk reducing policies. The most appropriate welfare measure for this purpose is the (Hicksian) consumer surplus based on a WTP for flood risk reduction approach. The method that will be used in this thesis to derive these WTP estimates is a stated choice experiment, which is well-grounded in the stated preference literature. Multiple arguments have been discussed to defend the choice for the SP method, but the potential disadvantages of this method should also be kept in mind. Choice experiments characterize flood risk reducing policies by a set of attributes, which allows the researcher to capture both the multi-

---

<sup>10</sup> Random parameter logit models relax IIA at the aggregate level, i.e. based on expected choice probabilities. Conditional on the individual specific preference parameter IIA still holds.



dimensional aspects of flood risks (i.e. probability and consequences) and the extent to which respondents prefer to reduce their exposure to each of these dimensions through marginal willingness-to-pay estimates. The final section of this chapter introduced the properties of the basic econometric framework underlying choice experiments, the multinomial logit model, and showed how marginal WTP estimates can be obtained from stated choice experiment data.

## Chapter 3: Preference uncertainty in stated choice experiments

The extent to which WTP estimates are valid and reliable hinges on the correspondence between the behavioural assumptions underlying the random utility framework and the actual choices made by respondents.<sup>11</sup> This chapter will focus on the core assumption that respondents have well-defined preferences over the alternatives presented in a stated choice experiment. This assumption is questionable in our case study on flood risk exposure in the Netherlands as discussed in Chapter 1. Moreover, preference uncertainty is likely to be present in various other contexts associated with a lack of experience and familiarity. In fact, preference uncertainty is not restricted to the hypothetical markets established in stated preference research, but may also occur in real world markets (e.g. Brown et al., 2008). The outline of this chapter is as follows. Section 3.1 discusses the basic rationality assumptions underlying the micro-economic framework. Section 3.2 defines the concept of preference uncertainty. Section 3.3 further refines the concept of preference uncertainty and puts it into the context of stated choice experiments, the random utility framework and discrete choice models. An overview of the literature on explicit modelling of preference uncertainty using choice certainty follow-up questions is presented in Section 3.4. Section 3.5 discusses implicit modelling strategies to account for the impact of preference uncertainty, and the closely related concept of choice task complexity, on the scale of the utility function. Section 3.6 extends the discussion to preference dynamics and related learning and fatigue effects over the choice sequence. Section 3.7 offers explanations other than preference uncertainty for the potential dynamics arising in stated choice experiments and Section 3.8 summarizes this chapter.

### 3.1 Rationality assumptions

*Consistency* is one of the basic rationality assumptions imposed when modelling choice behaviour. The concept denotes that when individual  $i$  prefers policy  $j$  over  $k$  in choice set  $D_{it}$ , it is not expected that he will choose policy  $k$  when faced with an alternative choice set that amongst others also includes policy  $j$  (Hanley, Wright and Adamowicz, 1998). Hence, a choice for policy  $j$  means that the individual prefers policy  $j$  over policy  $k$  and all other policies available in the choice set. By answering a sequence of choice tasks the respondent

---

<sup>11</sup> WTP measures are assumed to be valid, i.e. unbiased, if the estimated mean WTP value equals the ‘true’ mean WTP value. Reliability refers to the reproducibility of the WTP estimate and hence mainly concerns its stability over time (Whitehead et al., 1995).

reveals (a part of) his preference relation over the various policy options<sup>12</sup>. The micro-economic framework assumes that the underlying preference relation is *complete* and *transitive*. The former implies that the individual has a well-defined preference ordering over any two policy options in the choice set. Cycles in the preference relation, i.e. simultaneously preferring *j* over *k*, *k* over *l* and *l* over *j*, are ruled out by the transitivity criterion. Violations of consistency will not be observed if the respondent obeys these two rationality assumptions. Moreover, completeness and transitivity imply that the preference relation can be described by a utility function (Mas-Colell et al., 1995). The utility function describes an ordinal relationship between the policies in the choice set and is by definition independent of level and scale.

Rational respondents are assumed to behave in a utility maximizing fashion and select the alternative generating the highest level of utility. According to the completeness assumption the respondent knows in any choice set which alternative he will select; and therefore also knows how changes in specific policy characteristics affect his utility, i.e. personal well-being. More formally, respondents are assumed to know their preference ordering for the alternatives in any choice set and the marginal rate of substitution (MRS) between specific attributes, including price. In stated choice experiments it is commonly assumed that respondents are willing to make trade-offs between all attributes. McIntosh and Ryan (2002, p. 372) describe this underlying *continuity*, or full compensatory, assumption in the following manner: “*it is assumed there always exists some level of improvement in one good (or attribute) which can compensate an individual for a deterioration in another good, attribute or income, whilst leaving the individual on the same indifference curve (level of utility)*”.<sup>13</sup> This set of assumptions provides a sound framework to derive an implicit monetary value for policy changes.<sup>14</sup> However, the extent to which individual choice behaviour fulfils these rationality requirements embedded in neo-classical welfare economics is disputable (Rieskamp, Busemeyer and Mellers, 2006).

---

<sup>12</sup> Theoretically, the degree of precision by which the researcher can describe the preference relation depends on the design of the choice experiment. Precision increases with the number of (i) choice tasks and (ii) alternatives presented to the individual and the degree of systematic variation in attribute levels across alternatives and choice tasks (Ferrini and Scarpa, 2007).

<sup>13</sup> Lexicographic preferences are a form of non-compensatory behaviour and violate the continuity assumption. Accordingly there is no relative price that can be estimated (e.g. Campbell et al., 2008).

<sup>14</sup> The ordinal properties of the utility function do not preclude the calculation of marginal WTP, because taking the ratio of two derivatives rules out the level and scale effect of any positive transformation of utility.

### 3.2 Incomplete preference relations

Many non-market valuation studies in environmental economics cover goods or policies that can be characterized by aspects of unfamiliarity and low experience (e.g. Bateman et al., 2008). Making a well-informed decision over a set of policies is therefore a complex task. Consequently, the respondent may be uncertain about his preference order over the alternatives in the choice set; the impact of specific attributes on utility; or both.

To accommodate these aspects of preference uncertainty, we follow Ariely et al. (2003) by specifying a (non-random) additive utility function for policy  $j$ , which comprises  $A$  policy attributes. In (3.1) each policy attribute  $a$  contributes to overall utility through a specific weight  $\beta_a$ . Preference uncertainty is introduced by assuming that the individual does not know his exact vector of attribute weights  $\beta = (\beta_1, \beta_2, \dots, \beta_A)$ . Based on previous experiences in comparable decision environments, the individual may have a limited set  $E$  of acceptable  $\beta$  vectors. Uncertainty regarding the exact level of  $\beta$  may result in an undefined preference order; and therefore a violation of the completeness axiom (see (3.2)).

$$(3.1) \quad U_{ijt} = \sum_{a=1}^A X_{ijt}^a \beta_a$$

$$(3.2) \quad j \succ_E k \text{ iff } \text{Min}_{\beta \in E} \sum_{a=1}^A (X_{ijt}^a \beta_a - X_{ikt}^a \beta_a) > 0$$

Policy  $j$  is preferred over policy  $k$  if and only if all acceptable  $\beta \in E$  make the respondent better off with policy  $j$ . For some  $\beta \in E$ , policy  $k$  may be preferred over policy  $j$  and vice versa. In that case, the respondent does not have a clear preference order over the policies in the choice set. While comparing any two policies, three preference outcomes can be described; (i) policy  $j$  is strictly preferred over policy  $k$ , (ii) policy  $k$  is strictly preferred over policy  $j$ , and (iii) a shaded area in which the preference ordering is not defined. We define preference uncertainty by the third outcome.

**Figure 3.1 Preference uncertainty in choosing between alternatives  $j$ ,  $k$  and  $l$**

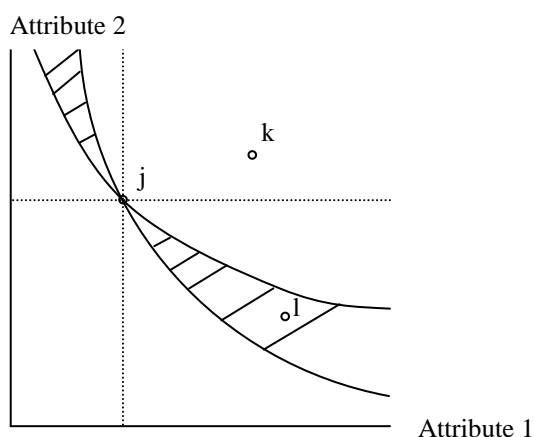


Figure 3.1 illustrates these three outcomes for a policy comprising two attributes. The only rationality assumption we impose is monotonicity with respect to each policy attribute. That is, improvements in an attribute level always have a non-negative (or non-positive) impact on utility (e.g. Geweke, 2011). Without loss of generality, assume utility is non-decreasing in both attributes. Accordingly, four quadrants can be defined around reference policy  $j$ . If policy  $k$  is situated in the north-east (south-west) quadrant, both attributes and utility are improved (reduced) relative to policy  $j$ . Since the respondent is not faced with a trade-off, (s)he will automatically choose policy  $k$  ( $j$ ). In the remaining two quadrants the respondent is faced with an improvement in one attribute, which comes at the cost of deterioration in the other attribute. Under well-defined preferences the respondent knows exactly whether the improvement in attribute 1 offered by policy  $l$  outweighs the disutility of the deterioration in attribute 2, or makes him indifferent between  $j$  and  $l$ . Due to unfamiliarity, as represented by the set  $E$  in (3.2), respondents do not exactly know their MRS between the two attributes. This incompleteness of the preference relation is depicted by the shaded area in Figure 3.1. If a small (large) deterioration in attribute 2 is required for a large (small) improvement in attribute 1, most respondents will be able to indicate a clear preference between the two policies. When the increase in utility offered by the improvement in attribute 1 is nearly offset by a similar level of disutility due to deterioration in attribute 2, a clear decision is difficult for the respondent.

Ariely et al. (2003) state that, in case of an incomplete preference ordering, respondents base their decision on an arbitrary rule unrelated to preferences. In the random utility framework, as described in Section 2.3, incomplete preferences manifest themselves when the stochastic component  $\varepsilon_{ijt}$ , which is unrelated to the characteristics of the policy,

dominates the decision. Note that this happens automatically when policy alternatives get sufficiently close in deterministic utility  $V_{ijt}$ . Respondents that are subject to greater preference uncertainty exhibit more random behaviour and therefore have a larger variance associated with  $\varepsilon_{ijt}$ . Various authors have incorporated this hypothesis in their empirical work by adding an additional (or modifying the) stochastic term to the utility function (Hanemann and Kristrom, 1995; Li and Mattsson, 1995; Vazquez et al., 2006). It should be noted here that, given that RUM assumes respondents are fully rational, the stochastic term of the utility function traditionally only represents noise at the end of the analyst. The work highlighted above and this thesis will exploit the error term to capture beyond measurement issues at the level of the analyst also preference uncertainty at the level of the respondent. Thereby it represents all deviations from the imposed utility function.

### **3.3 Trade-off and package uncertainty**

In Section 3.2 we have defined preference uncertainty as a violation of the completeness axiom. More specifically, in (3.2) the respondent is uncertain about his marginal rate of substitution (MRS) between policy characteristics. We label this form of preference uncertainty arising at the level of policy attributes as ‘trade-off uncertainty’. Trade-off uncertainty is, however, not taken into account by modifying the stochastic term in the RUM framework as done by Vazquez et al. (2006) and others. The latter only capture preference uncertainty arising at the level of the alternative, i.e. the entire policy package. The MRS between specific attributes is assumed to be known by the respondent; uncertainty only arises about the preference order over the alternatives in the choice set. We label the latter form of preference uncertainty ‘package uncertainty’. This distinction in types of preference uncertainty has not been made before in the literature. The distinction is relevant, since respondents in a choice experiment are required to make simultaneous trade-offs between alternatives and attributes.

To illustrate this distinction, imagine someone is buying a car and considers, amongst other options, to install an air conditioning (AC) system. Assume the budget constraint is not satisfied and the AC system is worth to that person about €500. As an additional assumption, the AC is worth definitely more than €450 and certainly not more than €550. In this example, marginal WTP is associated with a degree of imprecision. Since the person lacks experience and knowledge about AC systems, he is uncertain about accepting any offer in between the two price levels. Therefore, he may accept an offer of €525 or decline an offer of €480 and vice versa. Simultaneously, similar uncertainties may exist for other attributes of the car, like

the engine and the upholstery of the seats. These uncertainties at the attribute level contribute to the overall uncertainty regarding the utility derived from a specific alternative. Additional package uncertainty may still arise at the level of the alternative. For example, the final experience of driving a specific car in the choice set may remain unknown, though its characteristics in terms of its attributes are known. This implies that the individual may be uncertain about his WTP for the car in itself (package uncertainty) and his marginal WTP for improving a specific attribute (trade-off uncertainty). Both types of uncertainty are complementary to each other.

### 3.3.1 Modelling trade-off and package uncertainty

A standard RUM additive utility function is described by (3.3), where  $X_{ijt}$  is a row vector of non-price attributes and  $q_{ijt}$  the price of alternative  $j$  presented to individual  $i$  in choice number  $t$ . Moreover,  $\beta$  and  $\alpha$  capture respectively the marginal impact of individual attributes and price on utility, while  $\varepsilon_{ijt}$  represents the usual stochastic component. Marginal WTP  $\omega$  for attribute  $a$  is now defined by  $\beta^a/\alpha$ . Since price is assumed to have a negative effect on utility, marginal WTP is positive if an attribute has a positive impact on welfare. For illustrative purposes, we postpone the introduction of heterogeneity in preferences across respondents to the end of this subsection.

$$(3.3) \quad U_{ijt} = X_{ijt}\beta - q_{ijt}\alpha + \varepsilon_{ijt}$$

Li and Mattson (1995) and Hanemann and Kriström (1995) introduce package uncertainty in the RUM model by means of an additional stochastic component of utility  $v_{ijt}$ , which is independent from  $\varepsilon_{ijt}$ . Together they form the composite error term  $\zeta_{ijt}=v_{ijt}+\varepsilon_{ijt}$ . Package uncertainty thereby flattens the distribution of latent utility  $U_{ijt}$ . The inclusion of a composite error term is, however, uninformative on the possible presence of trade-off uncertainty, since marginal WTP is independent of  $\zeta_{ijt}$ .<sup>15</sup> In line with Ariely et al. (2003), trade-off uncertainty can be characterized by means of a distribution on marginal utility, i.e.  $f(\beta/\Omega)$ , where  $\Omega$  comprises the set of hyper-parameters. Consequently, marginal WTP is no longer specified as a fixed and known entity.<sup>16</sup> The support of  $f(\cdot)$  defines the set of acceptable

---

<sup>15</sup> Nevertheless, overall package uncertainty, as measured by the composite error term can provide useful information on the extent to which respondents make random decisions over alternatives.

<sup>16</sup> Both discrete and continuous distributions offer a natural interpretation of trade-off uncertainty. In particular, for densities with a finite support, such as the triangular distribution, the respondent is able to identify bounds on

parameter values  $E$  as defined in (3.2). The corresponding cumulative density function describes how likely it is that the respondent will accept the offer.

As an aside, Loomes et al. (2009) describe a similar model for trade-off uncertainty based on a discrete distribution. In an analytical fashion they show that uncertain respondents experience exchange resistance and are therefore more likely to stick to the status quo option. This exchange resistance arises as a consequence of loss aversion and non-linear probability weighting. The reference-dependent subjective expected utility model developed in Loomes et al. (2009) is thereby closely linked to prospect theory (e.g. Shaw and Woodward, 2008). Random parameter logit models, which are central to this thesis, are based on maximization of expected choice probabilities, where the probability density  $f(\cdot)$  of the random parameter is treated in a linear fashion. This thesis does not aim to include aspects of prospect theory into the decision framework under preference uncertainty.<sup>17</sup>

Despite the presence of preference uncertainty, an individual often needs to make a decision. Due to the lack of a clear preference ordering over the alternatives in the choice set, the individual in the previous example will sometimes accept a bid of €525 for the AC and sometimes decline the bid. As such, the rates of trade-off displayed by the respondent may vary over the choice sequence, i.e.  $\beta_{it}$ . Here, the assumption is that each decision is based on a particular realization of  $\beta$  from its underlying density. The realized  $\beta_{it}$  is unrelated to the decision environment and thereby comparable to the arbitrary decision described by Ariely et al. (2003). These intra-respondent variations in marginal utility over the choice sequence represent trade-off uncertainty. Preference uncertainty can accordingly be an explanation for decisions that may look inconsistent at first sight (Rouwendal et al., 2010)<sup>18</sup>.

Our interest with respect to research questions one and two is in  $\beta_{it}$ , the parameters of the utility function for individual  $i$  in choice task  $t$ , and their dynamics over the choice sequence and across respondents. It is in general not feasible to estimate  $\beta_{it}$  for every specific individual and every choice task given the limited informational content of choice experiments (see the discussion in Section 3.3.2 and Chapter 5). Various models have been introduced allowing for heterogeneity in preferences across respondents and/or choice tasks. The standard MNL model assumes marginal utility to be constant across respondents and choice tasks. Revelt and Train (1998) discuss two relaxations of the MNL model, the cross-

---

acceptable parameter values. Note that uncertainty about marginal WTP can also be modelled directly by specifying the utility function in WTP-space (Train and Weeks, 2005).

<sup>17</sup> Additional control variables in the utility function may pick up part of these effects.

<sup>18</sup> Using a novel, but deterministic method, Rouwendal et al. (2010) find that approximately 35% of the respondents show an inconsistency in their preference relation over a sequence of ten binary choice tasks.



sectional and panel mixed multinomial logit model (MMNL). These random parameter logit models assume preferences vary respectively across observations and individuals, i.e.  $\beta_{it}$  and  $\beta_i$ . In the cross-sectional mixed logit model all observations are treated as independent and a realization of  $\beta_{it}$  is assumed to be drawn from an underlying mixing density. The mixing density can be described by any type of probability density function (e.g. Balcombe et al., 2009).

Table 3.1, describes the cross-sectional MMNL model for a normal probability density function with mean  $b$  and standard deviation  $\gamma$ . Note that the cross-sectional MMNL model is similar to our specification of trade-off uncertainty above, which does not take into account that preferences are correlated across respondents. In other words, unobserved heterogeneity in preferences across observations was interpreted as trade-off uncertainty. The panel MMNL model takes unobserved heterogeneity in preferences across respondents into account by assuming that respondents base their decisions on a vector of parameters  $\beta_i$ , which is constant over the choice sequence but varies across respondents according to a mixing density. As such, the panel MMNL model assumes each respondent knows his marginal rate of substitution and does not take into account trade-off uncertainty. The panel MMNL model is used as the base model to answer research question one in Chapter 5.

**Table 3.1: Overview of models controlling for inter- and intra-respondent preference heterogeneity**

Independent Random Parameters:		
Model:	Parameter description	Mixing density
MNL	$\beta = b$	Fixed parameter
RPL – Cross Section	$\beta_{it} = b + \vartheta_{it}$	$\beta_{it} \sim n(b, \gamma)$
RPL – Panel	$\beta_i = b + \varphi_i$	$\beta_i \sim n(b, \sigma)$
RPL – Hess and Rose	$\beta_{it} = \mu_i + \vartheta_{it}$	$\beta_{it} \sim n(\mu_i, \gamma)$
	$\mu_i = b + \varphi_i$	$\mu_i \sim n(b, \sigma)$
RPL – Hess and Rose generalized	$\beta_{it} = \mu_i + \vartheta_{it}$	$\beta_{it} \sim n(\mu_i, \gamma_i)$
	$\mu_i = b + \varphi_i$	$\mu_i \sim n(b, \sigma)$
		$\gamma_i \sim IG(v, s)$
Correlated Random Parameters:		
Model:	Parameter description	Mixing density
MNL	$\beta = b$	-
RPL – Cross Section	$\beta_{it} = b + \vartheta_{it}$	$\beta_{it} \sim N(b, \Lambda)$
RPL – Panel	$\beta_i = b + \varphi_i$	$\beta_i \sim N(b, \Sigma)$
RPL – Hess and Rose	$\beta_{it} = \mu_i + \vartheta_{it}$	$\beta_{it} \sim N(\mu_i, \Lambda)$
	$\mu_i = b + \varphi_i$	$\mu_i \sim N(b, \Sigma)$
RPL – Hess and Rose generalized	$\beta_{it} = \mu_i + \vartheta_{it}$	$\beta_{it} \sim N(\mu_i, \Lambda_i)$
	$\mu_i = b + \varphi_i$	$\mu_i \sim N(b, \Sigma)$
		$\Lambda_i \sim IW(v, s)$

Notes:

$n$  represents a normal distribution with mean and standard deviation

$N$  represents a multivariate normal distribution with mean vector and covariance matrix

$IG$  represents an inverted gamma distribution with shape and scale parameter

$IW$  represents an inverted Wishart distribution with shape parameter and scale matrix

$i$  represents inter-respondent preference heterogeneity

$it$  represents inter and intra-respondent heterogeneity

$\vartheta_{it}$  and  $\varphi_i$  are treated as realizations of an underlying zero-mean (normal) distribution

Hess and Rose (2009) and Hess and Train (2011) combine the cross-sectional and panel mixed logit model and thereby allow for both inter- and intra-respondent heterogeneity. Preferences vary across respondents in line with the panel mixed logit model. That is, an individual specific vector of marginal utility parameters  $\mu_i$  is drawn from a (normal) distribution describing the distribution of preferences over the population. Preferences are allowed to vary within respondents by adding a choice task specific effect  $\vartheta_{it}$  to  $\mu_i$ , which is based on a zero mean cross-sectional (normal) distribution with a variance term  $\gamma^2$  common to all respondents. Again the cross-sectional component can be interpreted to represent trade-off uncertainty, since it induces preferences to vary across observations. By imposing  $\gamma^2$  to be the same for each respondent, Hess and Rose (2009) treat trade-off uncertainty as constant across respondents.<sup>19</sup> Due to differences in experience, knowledge and capabilities to deal with the cognitive burden they are exposed to, some people may be more uncertain about their

<sup>19</sup> Hess and Rose (2009) do not attribute intra-respondent heterogeneity to preference uncertainty.

decisions than others. Hence, a generalization of the Hess and Rose (2009) model is proposed below.

In the generalized model, inter-respondent preference heterogeneity is modelled in the same way as in the standard Hess and Rose (2009) model and enters the model through  $\mu_i$ . Intra-respondent preference variation still enters the model through a cross-sectional component  $\vartheta_{it}$ , but  $\gamma^2$  is replaced by an individual specific variance term  $\gamma_i^2$ , or covariance matrix  $\Lambda_i$ , representing individual specific trade-off uncertainty. The off-diagonal elements of this covariance matrix may be non-zero since it is not unlikely that the degree of trade-off uncertainty is correlated across policy attributes. Variation in trade-off uncertainty across respondents is in a similar vein to preference heterogeneity modelled by means of an additional mixing density. A limited set of distributions is available for this purpose, since the variance element  $\gamma_i^2$  needs to be positive, or positive-semidefinite in case of a multivariate distribution. For example, an inverse gamma or inverse Wishart distribution can be applied to represent a valid distribution for respectively a variance term or covariance matrix. Accordingly,  $\beta_{it}$  can be described in the following manner  $\beta_{it} \sim N(\mu_i, \Lambda_i)$  and we maintain the assumption that each decision is based on a realization of  $\beta_{it}$ .

Most important here, is the conceptual aspect that preferences are allowed to vary across respondents and choice tasks, and that the degree of intra-respondent variation may also vary across respondents. These intra-respondent variations in behaviour are attributed to trade-off uncertainty.<sup>20</sup> Not controlling for such unobserved heterogeneity may result in biased welfare estimates, in particular if these heterogeneity effects are correlated over attributes (van den Berg et al., 2010).

### 3.3.2 Econometric limitations

While the respondent makes simultaneous trade-offs between multiple attributes and alternatives in each choice task, the researcher only observes a single choice. The deterministic “half-space method”, as introduced by Rouwendal et al. (2010), makes clear that in each decision, the respondent reveals some information about his MRS, but not the exact levels. A decision for alternative  $j$  reveals that the change in overall utility offered by this alternative outweighs the change in utility offered by the other options in the choice set, including the status quo. This implies that marginal utility and the MRS can only be defined conditional on the MRS between the other policy attributes. Moreover, an upper bound on the

---

<sup>20</sup> Note that the proposed model does not incorporate learning and fatigue effects, which can also be a cause of intra-respondent variation in preferences. See section 3.6 for more details.

MRS cannot be defined through a single choice task. Accordingly,  $\beta_{it}$  lacks empirical identification, which is amplified by the fact that variations in preferences across observations are hard to separate from inconsistent choices arising due to the random component in the RUM.<sup>21</sup> The latter is confirmed in an extensive simulation study by Hess and Train (2011), which also highlights the benefits of working with repeated choice data for each respondent. By presenting the respondent with a sequence of choices with varying levels of trade-off between attributes, the MRS applicable to a single respondent can be narrowed down. Naturally, these benefits are conditional on the assumption that the respondent exhibits a set of stable preference parameters  $\beta_i$ .

It is not surprising that the identification issues regarding  $\beta_{it}$  become worse when one attempts to combine inter- and intra-respondent preference variation in a single model, let alone estimate the generalized Hess and Rose model. Relative to the cross sectional mixed logit model one tries to elicit information from the same choices about individual and observation specific preferences. Hess and Rose (2009) and Hess and Train (2011) acknowledge that not only the computational burden of their model is substantial, but that also the data requirements are considerable. Even in a simulated dataset Hess and Train (2011) obtain high standard errors for the cross-sectional parameter  $\gamma$ . Since the proposed model is a generalization of the Hess and Rose model, even more requirements are imposed on the data. By allowing the degree of intra-respondent variation to vary across respondents a large number of choice tasks per respondent are required. Preferably the choice tasks exhibit a low level of correlation in their attribute levels across alternatives to separate trade-off uncertainty, i.e. variation in  $\beta_{it}$ , from erroneous choices. We deem this not to be a feasible approach, however, because increasing the length of the choice sequence increases the cognitive burden on the respondent and may simultaneously induce related learning and fatigue effects (e.g. Brouwer et al., 2010 and Section 3.6) as embedded in research question two.

Merits may be found in the half space method, since the smallest set of marginal rates of substitution that is consistent with a constant marginal rate of substitution for a specific individual still provides information on the lower and upper bound of his MRS. In a similar vein to Figure 3.1, however, it will still be unknown which alternative the respondent will select if the presented alternatives fall in between these bounds. Moreover, it is in general not feasible to identify bounds for the MRS of each respondent in the sample. The number of

---

<sup>21</sup> Rouwendal et al. (2010) also present a stochastic model, comparable to Balcombe et al. (2007), taking into account the possibility of respondents misrepresenting their preferences.

acceptable intervals in the half space method increases rapidly in the number of alternatives; and relationships between various MRSs become complex when a large number of attributes is included in the choice experiment (Rouwendal et al., 2010). Accordingly, the panel MMNL model seems to be the most reliable model currently available to control for preference heterogeneity. Intra-respondent preference variation will still have an impact on the estimated model parameters, because  $\beta_i$  will be associated with a degree of uncertainty due to the arising preference inconsistencies across choice tasks and limitations of the experimental design. The mixing density will therefore capture aspects of inter- and intra-respondent taste variation which cannot be separated from each other. Since the mixing density is not individual specific, uncertainty regarding  $\beta_i$  is averaged over all individuals.

### 3.3.3 Fuzzy preferences

Van Kooten et al. (2001a) criticize the probabilistic interpretation of preference uncertainty embodied in RUM models (e.g. Loomis and Ekstrand, 1998; Ready et al., 2001). They postulate that preferences are inherently vague, such that the location (or the existence) of the ‘true’ indifference curve is never known. Van Kooten et al. (2001b) argue that due to their limited cognitive ability, respondents never know their preferences for particular amenities. This interpretation of preference uncertainty is in line with our graphical representation in Figure 3.1. According to Van Kooten et al. (2001c), respondents are able to determine their preference order relatively easily when presented with extreme rates of trade-off. At the margin, however, choices become difficult. As an alternative to the probabilistic approach to preference uncertainty, Sun and Van Kooten (2009) therefore work with fuzzy preferences.

In standard preference relations the respondent can only prefer  $j$  over  $k$ ,  $k$  over  $j$ , or be indifferent between the two policies. Only one of these options is possible, i.e. simultaneously preferring  $j$  over  $k$  and  $k$  over  $j$  is excluded. Fuzzy preferences violate this “law of non-contradiction” by allowing preferences to be part of multiple sets. If the preference order is known to the respondent, the decision will belong to a unique set. Preference uncertainty comprises the case in which the sets  $j \succ k$  and  $k \succ j$  are overlapping. The degree of overlap is described by a membership function, which is defined over the interval  $[0,1]$ .<sup>22</sup> Consequently, the respondent can prefer policy  $j$  over  $k$  and policy  $k$  over  $j$  at the same time (Sun and van Kooten, 2009). Reflexivity is embodied in the fuzzy RUM model such that the indifference curve passes through the reference policy. This matches the rationality assumption of

---

<sup>22</sup> If the preference order is known, the membership function takes the value 1 (full member) or 0 (not a member).

monotonicity discussed in Section 3.2. Conceptually, the fuzzy preference model is similar to our interpretation of preference uncertainty. Unfortunately, the fuzzy preference approach also suffers from econometric limitations. The identification of the membership function for fuzzy preferences requires additional information about choice certainty; it cannot be estimated implicitly. Sun and Van Kooten (2009) rely on the use of self-reported choice certainty measures in a DC-CVM setting. Interpreting a choice certainty response in a SCE setting is, however, still under discussion in the literature. Its interpretation differs substantially from the CVM literature (e.g. Lundhede et al., 2009 and Section 3.4). Moreover, the fuzzy preference approach does not alleviate the inherent limitations associated with the informational content of discrete choice data as discussed in the previous subsection. Separating intra-respondent preference variation from erroneous behaviour remains a difficult task. The limited number of applications of fuzzy preferences in the RUM framework, in combination with the limitations of self-reported choice certainty questions in stated choice experiments, does not make the fuzzy preferences approach a suitable alternative to the panel mixed logit model.

### 3.3.4 Outlook

Research questions one and two respectively focus on dynamics in response patterns across respondents and similar dynamics over the choice sequence. This section discussed and proposed a set of models that enable researchers to account for such individual specific and (or) observation specific preference parameters, i.e.  $\beta_i$  and  $\beta_{it}$ , also labelled as inter- and intra-respondent preference heterogeneity. Preference uncertainty is closely related to the latter form of heterogeneity, but it is hard to define whether observed dynamics in response patterns are a result of structural dynamics in the preference relation or a result of random decision making possibly affected by (package) uncertainty. A new form of preference uncertainty was defined affecting observation specific preference estimates  $\beta_{it}$ . Trade-off uncertainty implies that not only the aggregate level of utility is associated with uncertainty. Respondents may also be uncertain about their marginal WTP for (or utility of) a specific attribute.<sup>23</sup> A conceptual model was developed to simultaneously accommodate inter- and intra-respondent heterogeneity, and trade-off and package uncertainty in the RUM framework. The data

---

<sup>23</sup> Loomes et al. (2009) also introduce parameter uncertainty, but assume respondents are strict utility maximizers filtering out all uncertainty by adopting an expected utility approach. The use of such a deterministic approach implies that a set of fixed regularities guides the decision process, restoring completeness of the preference relation. Though neutralizing the effects of preference uncertainty, Loomes et al. (2009) do not fully capture the implications of preference uncertainty on individual choice.

requirements of this model are, however, too high. Preference parameters lack empirical identification at the observation level and introducing an additional layer of parameters to measure trade-off uncertainty associated with these parameters induces more empirical identification problems. Also the fuzzy preference approach does not offer a suitable alternative to identify trade-off uncertainty. Hence, for identification purposes, simpler discrete choice models will be applied in the empirical part of this thesis not accounting for trade-off uncertainty. Chapter 5 will only account for inter-respondent heterogeneity in addressing research question one. Chapter 6 addresses research question two by looking into intra-respondent preference dynamics over the choice sequence while controlling for package uncertainty. Chapter 7 controls for package uncertainty and dynamics in response patterns across respondents and preference dynamics over the choice sequence. Empirical identification issues are alleviated in that chapter using responses to a self-reported choice certainty follow-up question. Within the next section we will review the literature on the empirical modelling of preference uncertainty in the CV and SCE literature through the use of self-reported preference certainty measures.

### **3.4 Explicit modelling of preference uncertainty**

#### **3.4.1 Choice certainty response formats and follow-up questions in the CVM literature**

In an overview article Shaikh et al. (2007) discuss treatments that have been applied in the CVM literature to capture the degree of preference uncertainty respondents experience during a survey and analyze its impact on stated WTP. Two main strands of literature can be identified. The first group of studies applies an adjusted response format, such that respondents can directly express their uncertainty about their decision, e.g. by including a “Don’t Know” (Wang, 1997), “Probably Yes” or “Probably No” option (Alberini et al., 2003). The second group of studies presents the respondents with the standard CVM response format, most often the DC-CVM format, which is followed by a post decision certainty question. The format of the follow-up question can take various forms, using either a numerical scale (e.g. Li and Mattsson, 1995) or text statements (e.g. Blomquist et al., 2009). In both treatments an explicit measure of preference uncertainty is observed.

The econometric approach to analyze the impact of stated preference uncertainty on WTP varies with the response format and the interpretation of uncertain answers (Samnaliev et al., 2006). Wang (1997) and Alberini et al. (2003) use an interpretation closest to the interpretation of incomplete preferences in this thesis. The “Don’t Know” response option is interpreted as a reflection of Figure 3.1’s shaded area. The respondent is uncertain about

accepting or rejecting the nonmarket good at the proposed bid level. Hence, an ordered probit model can be used to estimate the upper and lower bounds of this region. Since the response format applied by Alberini et al. (2003) comprises multiple bounds between “yes” and “no”, an extended model can be estimated identifying the thresholds for switching between various degrees of certainty. Vazquez et al. (2006) adopt the approach originally proposed by Dubourg et al. (1994) and let the respondent directly identify the upper and lower bounds of the shaded area by using a payment card format. Respondents indicate the highest price level at which they would definitely pay and the lowest price level at which they would definitely reject the nonmarket good. A truncated Bayesian regression model is estimated, using the self-reported lower and upper bound on WTP as censoring points.

The Wang-hypothesis states that if bid levels approach the ‘true’ underlying WTP value, respondents become uncertain (Wang, 1997). Expressing a degree of preference uncertainty is a rational response to a set of ill-defined preferences. Li and Mattson (1995) examine this hypothesis by means of a post-decisional preference certainty question using a scale ranging from 0 to 100%. Responses are re-coded from ‘yes’ to ‘no’, if certainty is below 50% and vice versa. The certainty scale is also recalibrated relative to 100% certainty, i.e. ‘no’ with 40% certainty becomes ‘yes’ with 60% certainty. The self-reported certainty levels are used as weights in a weighted binary probit model. This recoding and estimation procedure results in wider confidence intervals on WTP. An alternative recoding approach, labelled as the “asymmetric uncertainty model” (ASUM), is applied in studies using either follow-up questions or a multiple bounded response format (e.g. Champ et al., 1997; Ready, Navrud and Dubourg, 2001; Welsh and Poe, 1998). Not surprisingly, WTP estimates are reduced significantly due to recoding even the slightest degree of preference uncertainty to a ‘no’ response (Brouwer, 2011). Loomis and Ekstrand (1998) propose a symmetric recoding approach (SUM). The original ‘yes’ and ‘no’ responses and a follow-up question are used to construct a new continuous dependent variable on the closed interval [0 1]. First, all certainty levels below 50% are assigned a value of 0.5. All ‘yes’ responses above 50% are coded according to their certainty level and divided by 100. Similarly, all ‘no’ responses are re-coded by subtracting the reported certainty level from 100% and divided by 100. Contrary to Li and Mattson (1995), the ASUM and SUM recoding approaches interpret the recoded responses as choice probabilities, which directly serve as the dependent variable in a linear regression model.<sup>24</sup> The various models applied to control for preference uncertainty in CVM

---

<sup>24</sup> Also the fuzzy preference RUM by Sun and Van Kooten (2009) relies on a recoding approach. It uses the answers to the self-reported certainty follow-up question in establishing the membership functions.



studies do not offer a uniform conclusion with respect to their impact on model performance and WTP estimates (Shaikh et al., 2007). In fact, most of the observed consequences are a direct and predictable effect of the selected re-coding approach (Kobayashi et al., 2012). More recently, Moore et al. (2010) propose an econometric approach which interprets the self-reported preference certainty measure as the probability that the respondent will answer “yes” and does not rely on a recoding exercise. They show that WTP estimates are substantially lower compared to standard modelling approaches not controlling for preference uncertainty. Respondents are, however, only allowed to express their uncertainty for a “yes” answer. Accordingly, the obtained results are again a direct consequence of down scaling the percentage of positive responses. Brouwer (2011) rejects the idea of arbitrary recoding and proposes to include self-reported choice certainty as a separate error component in the estimated choice model. Chapter 7 argues that the latter approach may be associated with endogeneity issues, if not properly controlled for.

### **3.4.2 Choice certainty follow-up questions in stated choice experiments**

The application of choice certainty follow-up questions in choice experiments is not as straightforward as in DC-CVM studies. The dichotomous nature of the response format in the latter method enables the researcher to directly interpret the uncertain responses in terms of WTP uncertainty. In a SCE study the levels of multiple attributes vary across two or more alternatives. Recoding approaches therefore become problematic, because it is not clear which alternative should be considered as second-best. Moreover, the follow-up question only identifies whether the respondent is uncertain about his preference order or not. There is no longer a straightforward relationship with payment certainty, since trade-offs need to be made between multiple attributes, including cost. Lundhede et al. (2009) propose three recoding approaches rooted in the CV literature, but ignore the important issues raised above. The first method eliminates uncertain responses from the sample. The second recodes all uncertain answers into a choice for the status quo, while the third method recodes an uncertain response to a choice for the second best alternative (in terms of utility). This final recoding approach introduces endogeneity in the model, because the second best alternative is a priori unknown in a non-ranking SCE study. Like Kosenius (2009), Lundhede et al. (2009) do not find an improvement in the efficiency of WTP estimates due to these recoding approaches. Balcombe and Fraser (2011) present a SCE format in line with Fenichel et al. (2009), which allows respondents to select a “Don’t Know” option. Based on Shannon’s (1948) entropy, see Chapter 7, they show that the number of “Don’t Know” responses increases when the

alternatives in the choice set become more similar in terms of utility levels. This finding is in line with the Wang-hypothesis and also found by (i) Brouwer et al. (2010) and Olsen et al. (2011) in an (ordered) probit analysis explaining self-reported choice certainty; and (ii) in the recoding approach adopted by Lundhede et al. (2009). Alternatively, Hensher and Rose (2011) follow the approach adopted by Li and Mattson (1995) and weight each observation according to its associated level of self-reported preference certainty. They find an increase in model fit in their study on airline choice.

Overall, the interpretation and modelling of such explicit measures of preference uncertainty in SCE studies remains controversial and the number of applications limited. Simultaneously, most applications prefer to adopt a more implicit modelling approach to account for preference uncertainty in choice experiments. In these models, variation in the scale of the stochastic component of utility is interpreted as a measure of choice consistency.<sup>25</sup> These types of models are discussed in more detail in the next section. Moreover, Chapter 7 proposes a new modelling approach combining both the explicit and implicit measurement of preference uncertainty.

### **3.5 Implicit modelling of preference uncertainty**

Within the CVM framework, self-reported choice certainty follow-up questions are frequently applied to separate package uncertainty  $v_{ijt}$  from the stochastic component of utility  $\varepsilon_{ijt}$ . This separation is not common in the SCE literature. Here, the composite error term  $\varepsilon_{ijt}$  is generally modelled as a (unique) measure of consistency and interpreted in the same way as  $\varepsilon_{ijt}$ . It corrects for all un-modelled (and unobserved) variations in utility across observations both at level of the researcher and the respondent. Preference uncertainty is thereby one of many possible explanations for differences in utility variance across respondents, i.e. variations in the scale of utility  $\lambda_i$ . Uncertain respondents are assumed to have a higher underlying variance (or a lower scale) in their decision process and therefore make more random decision than respondents with a high degree of preference precision. Random behaviour increases the probability that the observed decision is not in line with deterministic part of the utility function. This section will start with a discussion of alternative models for taking into account variations in scale, followed by a review of the various interpretations of variations in scale across respondents.

---

<sup>25</sup> Lundhede et al. (2009), for example, propose to model scale as function of self-reported preference certainty responses.

### 3.5.1 Models taking into account variations in scale

Since utility is independent of level and scale, any positive transformation of the utility function does not affect the decisions made by respondents. It is therefore well known that adjusting the scale of the i.i.d. error term in the family of logit and probit models does not affect the fit of the model, nor marginal WTP estimates (Train, 2009). Typically, the scale is normalized to one in order to obtain an estimate of the preference parameters in the utility function. Accounting for variations in scale across respondents (or over the choice sequence) requires the introduction of scale heteroskedasticity. Rose et al. (2011) discuss nested heteroskedastic error variance models and panoptic heteroskedastic error variance models. The former introduces correlation across alternatives in the choice set through particular nesting structures. The interest here is in panoptic heteroskedastic error variance models accounting for scale heteroskedasticity not related to a subset of alternatives, because preference uncertainty is expected to be primarily related to respondents and (or) specific choice tasks and not to specific alternatives in a choice task. This set of models comprises the heteroskedastic multinomial logit (HMNL) model, the generalized MNL model (GMNL) and the scaled MNL model (SMNL). We describe the characteristics of these models briefly below.

The utility specification for the HMNL model in (3.4) is similar to (3.3), but multiplies the deterministic part of utility by  $\lambda_{it}$ , which represents the scale parameter for observation  $it$ .  $\lambda_{it}$  on itself is a function  $g(\cdot)$  of explanatory variables  $W_{it}$  and associated parameters  $\delta$ . For example, a dummy variable can be included in  $g(\cdot)$  to control for people with prior experience, because they are likely to exhibit less preference uncertainty, i.e. have a higher scale. Critical for the identification of the HMNL model is that the scale parameter of a reference group is normalized to one. This is a direct consequence of the preference parameters being perfectly confounded with scale. Consequently, researchers tend to work with a relative scale parameter highlighting whether specific groups of respondents display a larger scale than others (Swait and Louviere, 1993). Most applications of the HMNL model are conducted in the standard MNL framework, assuming preference parameters are constant across respondents (e.g. DeShazo and Fermo, 2002). Chapter 7 will present an application allowing both preferences and scale to vary across respondents thereby combining elements of the MMNL and HMNL model, which is closely related to the GMNL model discussed below. The respective chapter also argues that including particular control variables in  $g(\cdot)$  may induce endogeneity issues.

$$(3.4) \quad \begin{cases} U_{ijt} = \lambda_{it} (X_{ijt} \beta - q_{ijt} \alpha) + \varepsilon_{ijt} \\ \lambda_{it} = g(W_{it}, \delta) \end{cases}$$

In a series of papers, it is argued that the heterogeneity in preferences as revealed by the MMNL model can mainly be attributed to variations in scale across respondents (e.g. Louviere and Eagle, 2006; Louviere and Swait, 2010; Salisbury and Feinberg, 2010). The GMNL model, as discussed in Fiebig et al. (2010) and Greene and Hensher (2010) accommodates both random scale and preference heterogeneity across respondents. Preference heterogeneity is captured through the standard mixing density in the MMNL model where  $\beta_i$  is assumed to vary across respondents. Random scale heterogeneity is captured by a similar mixing density, which operates within  $g(\cdot)$  as defined in the HMNL model. Typically, a log-normal distribution is defined for random scale heterogeneity to ensure that scale, i.e. variance, is strictly positive.<sup>26</sup> A simpler model specification is the scaled MNL (SMNL) model (Brefle and Morey, 2000). The GMNL model reduces to the SMNL model when there is no unobserved preference heterogeneity across respondents, i.e.  $\beta_i = \beta$ . Although the SMNL model is similar to the HMNL model, the main difference is introduced by making scale a random variable rather than making scale a function of observables.

Hess and Rose (2011) highlight that scale and preference heterogeneity cannot be fully separated from each other within the GMNL model. They argue that inducing scale heterogeneity induces correlation across preference parameters and is thereby comparable to the use of correlated mixing densities in the MMNL model. The GMNL model thereby only offers alternative distributional forms by using a multiplication of particular distributions.

In the models specified above the scale parameter is confounded with all preference parameters in the utility function. In so-called WTP space models, scale is only confounded with the cost parameter (e.g. Train and Weeks, 2005). WTP space models offer the researcher a convenient interpretation of the parameter estimates, since they reflect marginal WTP instead of marginal utility; and a degree of control on the distribution of marginal WTP over the population. However, uncorrelated marginal WTP distributions (including scale) by definition introduce correlated distributions of random marginal utility parameters in

---

<sup>26</sup> Again normalization is required to identify the model and several parameter and sampling adjustments have to be made to estimate the GMNL model (Fiebig et al., 2010 pages 399-400; Greene and Hensher, 2010 pages 416-417). It is not surprising that the GMNL model is a data intensive model and researchers are have to experiment with various starting values to control for the possibilities of getting locked into a local optimum.

preference space and vice versa (Scarpa et al., 2008). Whether models in preference or WTP space should be applied therefore remains an empirical matter (e.g. Balcombe, Chalak and Fraser, 2009). Typically, a reduction in model fit is observed due to the application of WTP space models, while confidence intervals for WTP estimates become more credible. Chapter 5 presents an application of such a WTP-space model.

### **3.5.2 Complexity as an explanation of scale heterogeneity**

A lot of research has focused on identifying and explaining variations in scale across respondents. Payne et al. (1993), de Palma et al. (1994) and others argue that individuals have a limited ability to grasp the full complexity of most choice tasks and also have inherently different abilities to make choices. Hensher (2006) therefore adds choice task complexity to the standard list of explanations for (variations in) error variance, such as omitted variable bias, model misspecification and measurement error.

The complexity of a choice task can be associated with the description of the good itself and its related institutional (or market) setting. The complexity of the institutional (or market) setting may refer to the degree of transparency in the market, but also the amount of information and rules the respondent is required to process and obey. Preference uncertainty is closely related to this form of complexity, because these aspects of complexity make it difficult for respondents to evaluate and contrast various alternatives. Payne et al. (1993) and de Palma et al. (1994) postulate that respondents revert to alternative (simpler) decision strategies in face of complex choice situations and may therefore make suboptimal decisions. Cameron and DeShazo (2010) and Swait and Adamowicz (2001a) argue that the tendency to revert to such heuristics depends on the extent to which respondents are willing to put effort in the choice task. Respondents may perform a simple cost-benefit analysis on whether or not to put additional effort into the choice task.

Misspecifications of the utility function by imposing continuity on all attributes while some respondents choose in a non-compensatory fashion may result in variations in scale across respondents and biases in marginal WTP estimates. The literature on attribute processing strategies (APSs) investigates the development of discrete choice models in which respondents incorporate only a limited number of attributes in their decision framework (e.g. Campbell, Hutchinson and Scarpa, 2008; Hensher, 2010a).<sup>27</sup> It is frequently argued that models that explicitly recognise APS improve model fits and result in better defensible WTP

---

<sup>27</sup> The interested reader is referred to Cameron and DeShazo (2010) for a more detailed discussion on APSs.

estimates (e.g. Campbell et al., 2010b). The number of potential APSs, however, increases rapidly with the number of attributes included in a SCE. For example, in a SCE based on 5 attributes, Campbell et al. (2010b) identify 32 potential constrained latent class structures. Due to the rapidly increasing number of parameters in these models, empirical identification may become problematic when a large number of APSs is considered (Train, 2008). The remainder of this section, will be restricted to a discussion of findings in the literature regarding the impact of choice task complexity on the scale of utility, i.e. package uncertainty. We consider that increased choice task complexity results in random behaviour, which increases the probability of inconsistent behaviour. This hypothesis will be embedded in Chapter 7 as a part of answering research question three.

Hensher (2006) and others focus on how the dimensions of the choice task itself affect the cognitive load respondents face in SP surveys (see Bech et al., 2011 for an overview). The dimensions of the choice task are characterized by (i) the number of alternatives presented within a choice task, (ii) the number of specified attributes, (iii) alternative formulations of variation in the attribute levels, and (iv) the number of choice tasks presented to respondents. Swait and Adamowicz (2001a) combine these various aspects of complexity into a single entropy measure, a specification we follow in Chapter 7.<sup>28</sup> Swait and Adamowicz (2001a) embody the entropy measure within the HMNL model and apply it to ten different datasets covering both RP and SP studies in marketing, transport and environmental applications. In eight out of the ten cases they do not find a significant impact of complexity on scale, suggesting the absence of preference uncertainty due to choice task complexity.

In contrast to Swait and Adamowicz (2001a), DeShazo and Fermo (2002) model scale by means of an exponential function in which the measures of complexity are introduced in an independent and additive fashion. Besides varying the number of alternatives and attributes across choice cards, controls are included for the number of attributes whose levels differ across alternatives (NADA) in the choice set; and for the standard deviation of normalized attribute levels within and across alternatives. With respect to the number of attributes and alternatives in the choice set, a countervailing hypothesis is used. It is assumed that by more completely describing the good and market under consideration, a larger part of unexplained variance can be described. However, in both cases the amount of information to be processed may also result in higher error variance. For this reason, a quadratic relationship is expected between complexity and the number of alternatives in the choice set. The hypotheses are

---

<sup>28</sup> Although this facilitates model estimation, the use of a single entropy measure prevents identification of the impact of specific complexity dimensions on scale.

tested on data for recreational demand in national parks in Guatemala and Costa Rica. As expected, scale is decreasing in the number of attributes. With respect to the number of alternatives in the choice set, an optimal number exists as the variance first decreases and then starts to rise again. In this application the optimal number of alternatives is approximately three. Scale is, however, mainly affected by the variation in attribute levels and in particular the within alternative variation. Welfare estimates are furthermore significantly affected by controlling for scale heterogeneity.

Hensher (2006) argues that the only way to assess the impact of the underlying experimental design on choice consistency is to allow for systematic variations in the design dimensions within a choice experiment. In a series of papers, labelled as the 'design of designs' five design dimensions are identified and various design set-ups constructed using the principles of optimal designs. For example, Caussade et al. (2005) develop 16 alternative design set-ups varying in the number of choice cards, number of alternatives in each choice card, number of attributes per alternative, number of attribute levels and the range of the attribute levels. Scale is modelled as a function of these variables and additional respondent characteristics in the HMNL framework. In a route choice study in Chile, Caussade et al. (2005) find that it is mainly the number of attributes in the choice experiment that affects the respondent's ability to make decisions. Together with the number of attribute levels, it has a negative impact on scale. Like DeShazo and Fermo (2002), they also find that presenting respondents with only a limited number of alternatives in each choice set increases scale. The optimal number of alternatives in this case is four. Finally, increases in scale are observed when a limited range of attribute levels is applied as opposed to a wide range. The latter finding is also reported by Ohler et al. (2000) and is in line with the within alternative variation effect detected by DeShazo and Fermo (2002). In addition to Chile, the design of designs set-up has been applied in Australia and Taiwan (Rose et al., 2009). These spin-off papers mainly focus on the impact of the design dimension on marginal WTP estimates and alternative decision strategies, not on scale (e.g. Hensher, 2006).

Arentze et al. (2003) investigate whether the use of visual information reduces the cognitive burden on respondents and thereby increases the validity of the obtained welfare estimates. In a mode selection choice experiment for working trips in South Africa, task complexity is varied across respondents in the number of attributes and alternatives presented and by the presentation method used. Using the scaling approach, as described in Swait and Louviere (1993), they find that increasing the number of attributes significantly reduces choice consistency, but the same effect is not found for increasing the number of alternatives

in a choice task. The use of visual information in itself did not seem to affect scale. Finally, Dellaert et al. (1999) apply the HMNL model to estimate the impact of differences in attribute levels across alternatives on the scale parameter. In line with DeShazo and Fermo (2002) and Wang (1997), they hypothesize that the greater the similarity between the alternatives in the choice set, the more difficult decisions become. Results are somewhat surprising as scale decreases when price level differences and absolute price levels increase.

Overall, these papers support the notion that choice task complexity affects the scale parameter and possibly welfare estimates in choice experiments. The impact of specific design dimensions varies across papers, all results suggest that researchers should take the limits of an individual's cognitive ability into account when designing a choice experiment. The complexities imposed by the experimental design dimensions are complementary to the concept of preference uncertainty central to this thesis. Preference uncertainty is mainly associated with the properties of the good considered in the choice experiment and may vary across applications depending on the experience and familiarity of respondents. Naturally, more complex goods require additional attributes to describe their properties as completely as possible, thereby also increasing the cognitive burden through the dimensions of the experimental design. The experimental design on itself can, however, add to the amount of information the respondent needs to process and thereby influence the consistency of decisions. Chapter 7 will explicitly control for the impact of choice task complexity on responses patterns in SCE studies.

### **3.6 Preference dynamics**

In the previous sections, preference uncertainty was treated as a somewhat static phenomenon only varying across respondents and possibly choice tasks due to respondent and choice task characteristics. However, preference uncertainty related to lack of (choice) experience and familiarity can be partly overcome by presenting respondents with a sequence of choices. Over a sequence of choice tasks dynamics may arise in the individual's decision process (e.g. Bech, Kjaer and Lauridsen, 2011). Learning and fatigue effects (and other dynamics) arising as a result of the cumulative cognitive burden can increase, or reduce, the amount of behavioural noise in the choice model or even shift decision strategies. In terms of modelling, learning and fatigue effects imply that both preference and scale parameters may vary over the choice sequence, an issue already briefly discussed in Section 3.3.1.



### **3.6.1 Discovered Preference Hypothesis versus Coherent Arbitrariness**

Through repetition respondents are able to make better informed and more consistent decisions, because they learn about the survey format, the (hypothetical) market and their own preferences (List, 2003). Braga and Starmer (2005) label the former two processes as ‘institutional learning’ and the latter as ‘value learning’. Bateman et al. (2008) use these learning effects as a critique against the one shot nature of single bounded DC-CVM studies, because a set of stable and coherent preferences is typically lacking at the start of a survey. Well-formed preferences are a product of experience built up through repetition. The work in the papers referred to above is based on Plott’s (1996) ‘Discovered Preference Hypothesis’ (DPH). The DPH assumes that respondents have a pre-existing set of stable and coherent preferences, but that inexperienced respondents do not know these preferences when first coming to the market. The first rounds will therefore be characterized by a trial and error process potentially resulting in inconsistent decisions. After a couple of rounds respondents learn how to behave more optimally in the market, by experiencing the consequences of their own actions and those of other actors. Eventually behaviour is hypothesized to converge to the stable set of underlying preferences. Preference uncertainty at the start of the survey is therefore a likely explanation for observed behavioural anomalies, such as the WTP/WTA disparity and preference reversals, and their decay (e.g. Loomes et al., 2010).

The convergence of preferences predicted by the DPH is in contrast to the predictions from the literature on preference construction (Lichtenstein and Slovic, 2006). The latter branch of literature also predicts that preferences are dynamic and converging over a choice sequence, but argues that the convergence process is guided by so-called ‘framing effects’. A pre-existing (latent) set of stable preferences is lacking when respondents participate in a survey. The most prominent example is described by Ariely et al. (2003). The basis for his theory of ‘coherent arbitrariness’ (CA) is that respondents with incomplete preference orderings are assumed to make decisions based on an arbitrary component unrelated to the decision itself. Dynamics in and stabilization of preferences is ensured by assuming respondents remember past decisions and have an internal drive to be consistent with those past decisions. New choice tasks are therefore answered in a coherent fashion, precluding violations of transitivity. The impact of arbitrary elements on decisions under uncertainty in the CA framework translates into an unknown point of convergence for the stable set of

preferences, hence the term Coherent Arbitrariness.<sup>29</sup> In a series of six experiments, Ariely et al. (2003) show that initial value clues, also known as anchors, have an impact on the decision process and welfare estimates. For example, when people are reminded before the experiment about the last two numbers of their social security number, they are willing to pay more for the same bottle of wine if they have a high number compared to people with a low social security number. This anchoring or starting point bias does not disappear after a sequence of choices. In SP studies respondents are argued to mainly experience aspects of institutional learning as they do not experience the consequences of their choices in a hypothetical context. Nevertheless, Bateman et al. (2008) and Ladenburg and Olsen (2008) find a rapid decay in the starting point bias in a repeated DC-CVM and a stated choice experiment respectively, supporting the DPH. Groeneveld (2010), however, does not observe a decay in starting point bias in a survey on reducing damages to a specific ecosystem and contextual framing effects are not observed by Carlsson and Martinsson (2008).

### **3.6.2 Cumulative cognitive burden**

The positive predictions of respondents making more informed decisions as they proceed through a sequence of choices due to learning effects can be somewhat offset by the possibility of fatigue, complexity and boredom effects. Swait and Adamowicz (2001b), for example, argue that the number of choice tasks adds to the cumulative cognitive burden inducing respondents to adopt alternative decision strategies. They find that respondents are more likely to select the status quo option as the complexity of the choice task increases. Moreover, it seems as if respondents use successively less information as they progress through the choice sequence. Swait and Adamowicz (2001b) find shifts in attention towards the brand name of the product, rather than considering the full set of attributes, as respondents proceed through the choice sequence. According to Bech et al. (2011), dynamics in decision strategies are detrimental for welfare analysis, because in that case it becomes unclear which choices should be used as an input in SCBA. Here the focus is on two aspects, dynamics in the scale parameter over the choice sequence and similar dynamics in the preference parameters; and not the specification of the utility function. The topic of dynamics in functional form, e.g. switches towards non-compensatory decision-making and other response strategies, are an important area of future research, but is beyond the scope of this thesis.

---

<sup>29</sup> The unidirectional refinement of preferences underlying CA is completely driven by the respondent's internal drive for consistency. Post-decision evaluation might, however, show that the individual made a wrong choice. This is not incorporated in the theory of CA, but can be a part of the DPH.

The standard econometric approach to test for dynamics in preference and scale parameters, also applied in Chapter 6, has been to apply the Swait and Louviere procedure (1993). The test consists of three stages. First, the researcher splits the sample into alternative subsamples, for example based on the position of the choice task in the sequence, and then estimates an unrestricted model generating a set of unique preference parameters for each subsample. Second, a model is estimated with a common set of preference parameters, but a varying (relative) scale parameter across the subsamples.<sup>30</sup> The two models are contrasted by means of a likelihood ratio test. The test checks whether the preference parameters are equivalent across both samples. If the null-hypothesis of equivalent preferences is rejected, the samples cannot be combined and it is unknown whether the observed differences arise due to variation in preferences or also variations in scale. The third and final step is only conducted when the former null-hypothesis is not rejected and tests whether scale is equivalent across both samples. A pooled model with common scale and preference parameters is estimated and its log-likelihood value is contrasted against the second stage model using again a likelihood ratio test. With respect to learning and fatigue effects, increases in the scale parameter (i.e. lower variance) over the choice sequence indicate more consistent decisions and therefore learning effects. Fatigue is associated with a decrease in scale over the choice sequence.

In a study concerning forest management, Holmes and Boyle (2005) find a structural change in preference parameters over the choice sequence based on the Swait and Louviere procedure. While responses to the first three questions were comparable in terms of preference and scale parameters, the fourth (and final) choice task was answered in a significantly different way. They attribute this to a learning effect and consider the last response as the most informative. Holmes and Boyle (2005) also test for context dependence and contrast the current choice task to the preceding and following choice tasks in terms of losses and gains in attribute levels. The likelihood of selecting an alternative increases when the current alternative is cheaper than the alternatives offered in the previous choice task. Since respondents have no clue as to the composition of the next choice task in a SCE, their conclusion that respondents are mainly backward looking is not surprising. Indications of an increase in the scale parameter over a sequence of five choice tasks are also found by Brouwer et al. (2010). However, they cannot reject the null-hypotheses of stable preference and scale parameters in a study concerning water scarcity in Australia. A known problem with

---

<sup>30</sup> Due to preference and scale parameters being confounded, variations in scale can only be retrieved after imposing equivalence of preference parameters (Train, 2009).

the Swait and Louviere test procedure is that by comparing preferences across (smaller) subsamples, the efficiency of the parameter estimates reduces. Moreover, the full experimental design may not be presented to the respondents in a single choice task (e.g. Ladenburg and Olsen, 2008). To this end, Carlsson et al. (2010) present respondents the same experimental design twice in one choice experiment. Preferences are not found to be stable between the two repetitions. Their findings, however, suggest respondents primarily experience institutional learning effects in the first choice task. The largest number of inconsistent decisions (27%) is observed when comparing the first to the ninth choice task, i.e. the start of the repeated sequence. Preferences are then found to be stable when comparing tasks 2-8 with tasks 9-16. Bech et al. (2011) also find variations in preference parameters across subsamples of respondents, after presenting each subsample with an alternative number of choice cards, respectively 5, 9 and 17. However, the set-up of their study does not permit the identification of learning or fatigue effects based on the Swait and Louviere procedure. Savage and Waldman (2008) compare the scale parameter across two subsets of choice tasks in both an internet and a standard mail survey. They only observe fatigue effects for the internet survey through a decrease in scale. Hence, they recommend exploiting the low marginal cost of internet surveys by increasing the sample size, but presenting the respondents with a smaller number of choice tasks. For the initial set of four choice tasks the quality of responses in both survey formats turns out to be equivalent. In summary, results from the Swait and Louviere test procedure in various papers do not reveal a consistent pattern of dynamics in preference and scale parameters over the choice sequence.

A frequently adopted assumption is that respondents initially experience learning effects and then fall subject to fatigue effects. Caussade et al. (2005) therefore model scale in their HMNL using a quadratic term and find that scale is highest around the tenth choice task. In general, respondents are presented with a number of choice tasks ranging between 1 and 16, with an average of approximately eight choice tasks. Brazell and Louviere (1998) conducted a SCE with 10 attributes and 12, 24, 48 and 96 choice tasks and did not find reductions in response consistency over the choice sequence across the different versions of the design. Hence, they argue that respondents are able to handle dozens of choice tasks in a single experiment. Carlsson and Martinsson (2008), Stopher and Hensher (2000) and Hensher (2004) indeed find that the number of choice tasks only has a marginal impact on the behavioural responses, including fatigue effects. Similarly, Arentze et al. (2003) do not find significant fatigue effects over a choice sequence of 16 choice tasks. Ortuzar et al. (2000), on the other hand, found too many inconsistencies and fatigue effects over a choice experiment

of similar length in the pre-testing stage. To reduce the cumulative cognitive burden on respondents, they had to block the design into smaller subsets. In a pair wise comparison task over 10 to 16 choice tasks, Bradley and Daly (1994) find a continuous decrease in scale already after the second choice task. These fatigue effects only manifest themselves through a scale effect as the preference parameters remain constant over the choice sequence. Accordingly, Bradley and Daly (1994) conclude that by randomizing the order of the choices across respondents, researchers can limit the bias effect caused by the cumulative cognitive burden on respondents. In contrast, Brown et al. (2008) observe an increase in consistency due to fine tuning of preferences in a pair wise comparison exercise using public and private goods.<sup>31</sup> Finally, Bech et al. (2011) observe a somewhat higher response variance in the HMNL model when the number of choice tasks falls beyond a certain threshold level, but its impacts on welfare estimates are only minor. Hence, the impact of the number of choice tasks on choice consistency turns out to be ambiguous and its impacts on marginal WTP tends to be limited. Hess et al. (2012) go even further and state that the concerns in the literature about fatigue are possibly overstated, because they do not find a clear decreasing trend in scale across choice tasks while reviewing a wide range of the studies.

Besides the implicit modelling of dynamics in scale parameters, a limited number of studies looked into the dynamics in self-reported choice certainty over the choice sequence. Brouwer et al. (2010) find significant increases in choice certainty when moving from the first to the second and from the fourth to the fifth choice task. The effect of the choice task number, however, disappears when simultaneously controlling for the impact of respondent and design characteristics on self-reported choice certainty. Olsen et al. (2011) also observe ambiguous results. In a survey on the construction of a new motorway in Denmark they observe a learning effect over the choice sequence, but the choice task number is one of the few parameters not significant in an equivalent ordered probit model explaining choice certainty in a national park study. Lundhede et al. (2009) add to this ambiguity by finding a significant negative impact of the choice task number on scale, i.e. fatigue effects, for the same study using a HMNL model in which scale is a function of characteristics found to affect self-reported choice certainty in the Olsen et al. (2011) study.

---

<sup>31</sup> Brown et al. (2008) do not contrast DPH and CA as they lack contextual cues in their design.

### **3.7 More than preference uncertainty**

Thus far, we have assumed respondents behave in a non-strategic way and attempt to select the alternative they prefer most in each choice card. Preference uncertainty and related learning and fatigue effects are therefore primary sources of noise within the described random utility framework. Samnaliev et al. (2006), however, argue self-reported choice certainty levels may also be a way for respondents to express protesting beliefs (Meyerhoff and Liebe, 2009) or strategic behaviour (Carson and Groves, 2007; Collins and Vossler, 2009) by over or understating their WTP. These forms of behaviour are typically related to the hypothetical nature of the stated preference framework and are therefore commonly labelled as ‘hypothetical bias’ (e.g. Hensher, 2010b). It is beyond the scope of this thesis to discuss these aspects in detail, but these are potentially additional sources of bias in marginal WTP, affecting scale and preference parameters. By using self-reported choice certainty measures as a tool to mitigate hypothetical bias, Ready et al. (2010) highlight the close connection between these topics. It also underlines the difficulty faced by researchers. During a sequence of choices, a whole range of closely related dynamics may arise, but these remain latent to the researcher and can only be identified through implicit modelling strategies which identify variances (not their causes) or through follow-up questions. The interested reader is referred to the papers mentioned in this section for a more detailed discussion on protest behaviour, incentive compatibility and hypothetical bias.

### **3.8 Summary**

Section 3.1 provided a description of the standard rationality assumptions underlying the random utility framework, which assumes that respondents have a set of well defined preferences prior to engaging in a stated choice experiment. Preferences are, however, likely to be ill-defined for goods and services characterized by low experience levels and high degrees of unfamiliarity as described in Section 3.2. In those cases two forms of preference uncertainty may arise in choice experiments. First, trade-off uncertainty describes the fact that respondents are uncertain about the extent to which they are willing to give up one (policy) attribute for another. That is, they are uncertain about their marginal rate of substitution, and in particular their marginal WTP. Second, package uncertainty arises at the level of the alternative and denotes that despite the fact that the composition of the policy is known in terms of its attributes, the respondent may still be uncertain about his preference order. Both forms of uncertainty, as described in Section 3.3, are complementary to each other. The distinction between these two forms of preference uncertainty has not been made in the

literature before. The empirical identification of trade-off uncertainty is, however, hampered by the nature of attribute based discrete choice data and the non-deterministic aspects of the random utility framework. Alternative modelling approaches, such as fuzzy preferences, do not offer a solution to this problem either. Given these empirical identification issues, trade-off uncertainty will not be covered in the rest of this thesis. Moreover, research question one regarding heterogeneity in response patterns across respondents will be analyzed in Chapter 5 using the MMNL model abstaining from dynamics in preference and scale parameters over the choice sequence. Package uncertainty will be taken into account in Chapters 6 and 7 using explicit and implicit modelling methods as discussed in Sections 3.4 and 3.5. The former relies on the use of self-reported choice certainty questions, while the latter controls for dynamics in the scale parameter across respondents and over the choice sequence. Uncertain respondents are assumed to exhibit larger error variance (i.e. lower scale) and make more random decisions than certain respondents.

Section 3.5 emphasizes that the cognitive burden of a choice task is not only affected by preference uncertainty at the level of the respondent, but also by the dimensions of the design. As respondents need to process more information when being presented with, for example, more attributes or alternatives, the degree of consistency in their decisions may decrease. Empirical evidence of this effect is, however, ambiguous, but make clear that a stated choice experiment should be carefully designed to minimize the degree of behavioural noise. This is further highlighted in Section 3.6 where learning and fatigue effects are discussed. In the course of proceeding through a choice sequence respondents may exhibit aspects of institutional and value learning, but may also experience cumulative cognitive burden effects and provide less informative answers. Two contradictory hypotheses are discussed, namely the discovered preference hypothesis and coherent arbitrariness. Evidence on the existence of dynamics over the choice sequence is, however, ambiguous. Finally, we have highlighted in Section 3.7 that potential dynamics in choice behaviour and biases in marginal WTP estimates may also be a result of strategic behaviour due to a lack of incentive compatibility in stated choice experiments.

In the remainder of this thesis the concepts of preference uncertainty and related learning and fatigue effects will serve as the basis to answer research question two and three as defined in Chapter 1. Chapter 6, in particular, contrasts the discovered preference hypothesis and theory of coherent arbitrariness and thereby analyzes the dynamics in preference and scale dynamics over the choice sequence and the impact of these dynamics on WTP estimates. For example, the Swait and Louviere (1993) test procedure as described in

Section 3.6 is applied in that chapter. Chapter 7 works with most of the concepts discussed in this chapter. First, its econometric model combines the explicit and implicit modelling strategies by learning about preference uncertainty through self-reported choice certainty follow-up questions and the choices made in the choice sequence. Second, apart from controlling for learning and fatigue effects it uses the concept of entropy (Shannon, 1948) to trace the impact choice task complexity on preference uncertainty. The next chapter discusses the set-up of a stated choice experiment and their relation to the research questions.





## Chapter 4: Case study, survey set-up, hypotheses and experimental design

The purpose of the case study presented in this thesis is to explore the sensitivity of welfare measures derived in stated choice experiments to preference uncertainty and related dynamics in preference and scale parameters across respondents and over the choice sequence. To this end, an online survey was conducted in which respondents were presented with a series of questions eliciting their preferences for changes in their exposure to flood risks in the face of climate change. The main component of the survey consisted of a stated choice experiment (SCE) in which alternative public safety programs are presented to the respondents limiting the impacts of climate change on flood risk exposure. The first part of this chapter describes the study area, the general structure of the survey and the pre-testing stages. The second part comes up with a set of testable hypotheses and discusses how these hypotheses were taken into account in the experimental design of alternative versions of the SCE.

### 4.1 Study area

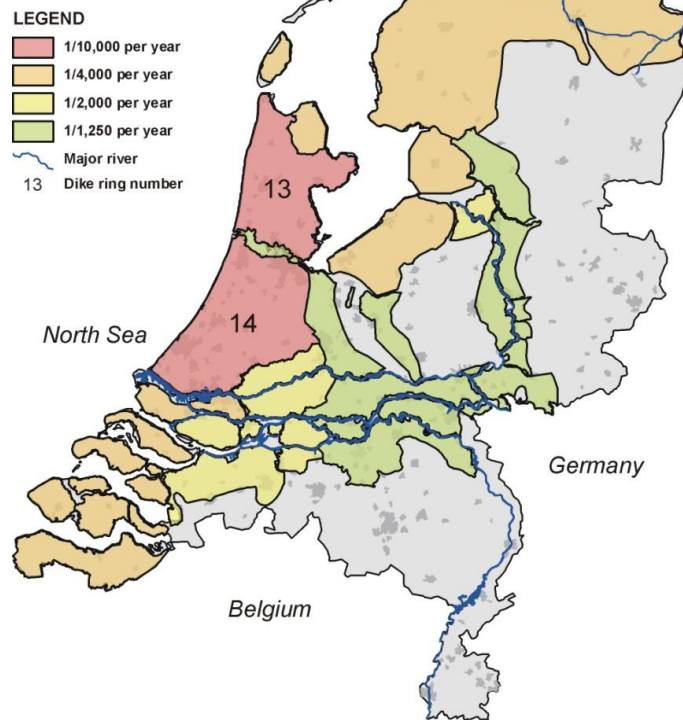
The selected case study area covers two so-called ‘dike rings’ in the Netherlands. A dike ring is defined as a connected series of water defence structures, e.g. dikes, dunes or high ground, protecting a specific geographical area from flooding. The Netherlands is divided into 53 dike rings, each with a designated flood probability as depicted in Figure 4.1.<sup>32</sup> Flood probabilities in the dike rings range between once every 1250 years along the main rivers and once every 10,000 years in densely populated areas (Bouwer and Vellinga, 2007; de Moel, Aerts and Koomen, 2011). We focus on dike rings 13 and 14, which are situated along the Dutch (west) coast and are threatened by coastal flood risks from the North Sea. The designated flood probability for dike rings 13 and 14 is set at the lowest level, i.e. once every 10,000 years. Impacts of a flood will be substantial as most of the country’s economic activity and population is located within these dike rings. As such, the case study area matches with our focus on low-probability-high impact events with which respondents lack experience. More specifically, the last major coastal flood in the Netherlands took place in 1953 and primarily hit the province of Zeeland located south of the case study area.

---

<sup>32</sup> The protection levels indicate that embankments should be high enough to withstand water levels that occur only once every X years, where X refers to the defined flood return period in the dike ring (de Moel et al., 2011). These flood probabilities do not take into account the possibility of failing of particular safety structures by collapsing, piping or erosion (Rijkswaterstaat, 2005b).

**Figure 4.1: Dike rings within the Netherlands**

**Safety standards for dike-ring areas**



Source [www.helpdeskwater.nl](http://www.helpdeskwater.nl); figure used with courtesy of H. de Moel

Most of the area within the two dike rings is situated below sea level and major cities like Amsterdam, Den Haag and Rotterdam are found here, contributing to the large damage potential. For example, Rijkswaterstaat (2005b; 2005c) provides a worst case scenario damage estimate of €290,000 million and €58,000 million for dike rings 14 and 13 respectively. These estimates are based on a global damage calculation method, where a scenario is fed into a hydraulic model generating maximum expected water depths and velocity at which water enters the dike ring during a flood. This is then combined with land use patterns to identify the extent to which land and property will be affected by a flood. Finally, the direct financial damage is calculated based on the associated land and property values as included in the “HIS-schade en slachtoffer module (version 2.1)” (Rijkswaterstaat, 2005a). The latter module describes values for land and property based on replacement values. Loss of life and other non-market goods are not taken into account in the global damage calculation method.<sup>33</sup> Moreover, the global damage calculation method results in an overestimation of expected flood damages, since it is based on a worst case scenario.

<sup>33</sup> Some of these non-market goods, like the value of a statistical life, have been included in the HIS-module. These values are also applied in a more detailed impact analysis for dike ring 14 (Rijkswaterstaat, 2005c).

The worst case scenario considered for dike rings 13 and 14 is when a heavy western storm coincides with high tides on the North Sea. In that case, sea water is pushed towards the Strait of Dover, which increases pressure on the defence structures along the Dutch western shores. In contrast to river floods, such extreme water levels and storm surges cannot be predicted long in advance, usually not more than one to one-and-a-half day (Rijkswaterstaat, 2008). Within such a short time span, effective preparations in terms of, for instance evacuation, are difficult, in particular because the exact location at which the storm will strike strongest is very uncertain at that stage. On the day of the storm surge, the location can be predicted more accurately and water levels can be identified with a 20cm uncertainty range. It is predicted that only 1% of the inhabitants will have sufficient time to leave the area at risk under these conditions. Eventually, 4,340m<sup>2</sup> will be flooded and about 2.3 million inhabitants will be affected by such an extreme flood (Rijkswaterstaat, 2008).

Due to climate change, flood risk exposure within the selected dike rings is expected to increase (Maaskant, Jonkman and Bouwer, 2009). Based on a set of (alternative) future scenarios regarding sea level rise and peak discharge levels in rivers by 2100, Aerts et al. (2008) analyse the impacts of climate change on flood probabilities within the various dike rings. In the absence of additional preventive measures, flood probabilities in dike rings 13 and 14 are predicted to increase to respectively once every 4,273 and 5,012 years conditional on a sea level rise of 24cm. The latter scenario is the most positive scenario considered. Maaskant et al. (2009) use a similar set of scenarios and predict an increase in flood probabilities to once every 4,000 years by 2040 due to higher discharge levels in rivers and a sea level rise of 30cm. Additionally, they also look at the damage potential in dike ring 14 and observe that flood risk exposure also increases due to an increase in the size of economic activity in the region and population growth. For example, a sea level rise of 30cm by 2040 is expected to result to increase the number of fatalities by 20 percent. When also accounting for population growth, the number of expected fatalities nearly doubles. To prevent an increase in the frequency of flooding and its related damages, additional investments in flood defence structures or alternative adaptation mechanisms are required. Kabat et al. (2005) label this as 'climate proofing' the Netherlands. From a cost-benefit analysis perspective, the amount of money that can be invested in such projects hinges on the balance between investment costs and expected benefits from reductions in flood risk, i.e. (expected) prevented damages (e.g. Brouwer and Kind, 2005; Brouwer and van Ek, 2004).

The presented damage estimates by Rijkswaterstaat (2005b; 2005c) based on the global damage assessment approach are incomplete with respect to many non-material

damages to property and individuals. As argued in Chapter 2, SCEs, as a form of stated preference research, can offer a suitable alternative to derive economic values for specific non-market goods and services, which could also be included in the HIS-module. Additionally, SP methods also offer more general insights into public (risk) perceptions and individual willingness-to-pay for specific flood protection projects reducing exposure to flood risks (e.g. Pearce and Smale, 2005). In this thesis we adopt the latter approach and identify which aspects of flood risk policies are important for individuals and how much they are willing to pay for a specific policy scenario. To this end, we embed a SCE within an online survey across a random selection of residents in dike rings 13 and 14 concerning their personal (coastal) flood risks exposure and flood risk perception.

#### **4.2 Survey set-up**

The survey was divided into five sections. The first section comprised a set of introductory questions regarding the current place of residence of the respondents, why they moved there and whether flood risk exposure played a role in this decision. Questions were posed related to various types of floods, e.g. coastal, river or water overflow due to heavy rainfall, which might occur within the area where they live, including which type(s) they expect to experience at least once during their lifetime. Also previous experiences with floods and related consequences were identified. These questions were aimed at measuring flood risk perceptions and slowly familiarizing respondents with their personal level of flood risk exposure. Flood risk perception was then measured more directly by asking respondents how often they expect a coastal flood to occur in their direct surroundings and how high they expect the water will rise during such a flood. In this first part of the survey respondents were also allowed to express that they were uncertain about flood probabilities and elevation levels in case of a coastal flood.

The second part of the survey aimed at updating flood risk perceptions by gradually introducing respondents with information regarding their actual (personal) flood risk exposure. To this end, respondents were asked to report their postal code. This information was directly fed into an online tool on the website [www.ahn.nl](http://www.ahn.nl) (AHN, 2010), which reports the average elevation level within each postal code area relative to sea level. Additionally, a map of elevation levels within the provinces of South- and North-Holland is shown to the respondents. This allowed them to identify their relative flood risk exposure. The question about expected water levels during a coastal flood was then repeated. The next step introduced the concept of flood probabilities. To explain frequencies of once in 10,000 years a

risk ladder was shown to respondents positioning flood risks relative to other probabilities of other risks like car theft and dying in a car crash (e.g. Corso et al., 2001). After the explanation, respondents were again asked to report the frequency at which they expect coastal floods to occur in their direct surroundings. The section was finished by a set of questions about the expected (financial) consequences of a coastal flood and the impacts of climate change on flood probabilities. In this part of the questionnaire respondents were no longer allowed to state they were uncertain neither about flood probabilities, nor about potential water levels during a flood.

The impacts of climate change on flood risks and related policy proposals were introduced in the third part of the survey as a preamble to the SCE. Respondents were informed that flood probabilities are expected to increase to once every 4,000 years by 2040 in the absence of additional preventive measures. The set of preventive measures, i.e. the attributes in the SCE, was then introduced and respondents were told that they would face a set of ten choice tasks in which they are requested to state their preferences for specific policy packages (more details on the SCE in Section 4.4). The fourth part of the survey considered the SCE and a set of methodological follow-up questions. Respondents were questioned about the way they responded during the SCE. Questions considered which policy characteristics were most important, how respondents evaluated choice tasks and whether they changed their evaluation criteria during the choice sequence. Related issues regarding the credibility of the choice tasks were also covered, including an open ended contingent valuation question related to the most preferred policy option. The fifth and final part of the survey covered respondent's demographic and socio-economic household characteristics.

### **4.3 Pre-testing**

The survey builds upon preceding stated preference research on flood risk valuation in the Netherlands (Bockarjova et al., 2010; Botzen, 2010; Brouwer and Schaafsma, 2011). Though these surveys had a different focus, being either private insurance programs to mitigate flood risks or measuring the value of statistical life, they faced the same challenges of measuring risk perceptions and communicating small risk levels. Consultations with the respective researchers resulted in useful insights into the extent to which respondents comprehended the concept of risk and the related choice tasks. The initial version of the survey comprised a mixture of the questionnaires used in the previous studies, which have been extensively pre-tested on their own, using face to face, Computer Assisted Personal Interviews (CAPI) and online response formats, and were conducted in different parts of the Netherlands, including

parts of our study area. Below we briefly discuss the four pre-testing stages that were conducted before launching the final version of the survey as presented in the rest of this thesis.

In the first stage, the survey was internally distributed to a group of researchers working on economic valuation in the Institute for Environmental Studies at VU University Amsterdam including the researchers mentioned above. The main purpose was to identify whether the concepts of interest (see Section 4.4 for more details) could be identified based on the presented study. In general, the survey was perceived as too long. Also the language style was considered too formal. Additional remarks were made about the graphical tools to clarify flood probabilities. With respect to the SCE, a group discussion was held to discuss the clarity of the attributes and their levels. The most important issue during this discussion was that the compensation attribute was considered to be mis-specified. The presented compensation levels in the range of €100,000 - €300,000 were not considered applicable to everybody, since expected damage in the case of a flood may go well beyond or below these levels. Suggestions were made to explain the current compensation structure through the “Calamities and compensation act” (Botzen and Van Den Bergh, 2008) and include questions about the expected degree of compensation under current conditions. During the next pre-testing stage the attribute levels were maintained, but discussions were held with respondents on this topic. Fundamental identification issues were not detected during the first pre-testing stage.

By the end of November 2009, a total of seven CAPI's were held to test the length of the survey, and discuss the presented concepts and communication methods. This second stage did not have the purpose of obtaining analyzable data. It took the respondents about 30 minutes to work their way through the survey. Hence, it was considered important to reduce the length of the survey. The general comment was that the concept of risk was still considered complex. Consequently, we put more emphasis on coastal flood risks specifically and adjusted our communication tools to better present coastal flood risk probabilities and their consequences, both verbally and graphically. For example, we added the external link where people could enter their postal code and get the average elevation level in their direct surroundings. The number of questions before the SCE was also reduced to prevent fatigue effects before the SCE. In the SCE, the pictograms explaining the policy attributes were considered to be clear enough. As expected from the first pre-test stage, the compensation levels were not considered realistic by all respondents. Discussions revealed that people preferred to get a percentage of their damage compensated, instead of a maximum amount of money. The levels were adjusted accordingly to percentages. The probability levels were still

hard to comprehend for everyone. We therefore added relative indications, for instance “2x smaller” than the status quo option, to the presented probability levels. The probability levels were not changed. Also the choice certainty follow-up question was slightly adjusted, since the 0-10 scale was considered too broad and uninformative. We therefore adopted a five point scale running from ‘very certain’ to ‘very uncertain’.

In December 2009 the survey was uploaded for a new test among 45 respondents. The sample in stage three was not representative, and consisted only out of residents living within the study area. The main goal was again to check the length and content of the questionnaire. Respondents were requested to send their remarks by e-mail to the researcher. The length of the survey had improved. It took respondents this time on average about 20-25 minutes to complete the survey. Nevertheless, some questions were removed, as we aimed for approximately 15-20 minutes. The main comments focused on the phrasing and interpretation of particular questions. These remarks usually did not require major changes. Some minor changes were made to the risk communication tools, resulting in the final version of the risk ladder. In the SCE each question was given a number and people were offered the opportunity to re-read the SCE-explanation in each question. The former was done to allow respondents to monitor their own progress, the latter to improve respondent understanding of the choice task.

The final pre-testing stage was conducted in January 2010 and completed by a representative sample of 100 respondents from the study area. The sample was obtained through Multiscope, a commercial company selling panel data, which was also hired to provide the panel of respondents for the final version of the survey. The average respondent took about 13.5 minutes to fill out the questionnaire and only four respondents took over twenty minutes to finalize the survey. Based on the response patterns, we decided to remove the “Don’t Know” option from some questions in order to force respondents to provide an answer regarding their risk perception. Out of the 100 completed SCEs, nineteen respondents consistently selected the status quo option. In order to identify the reasons why respondents did not change their responses, a follow-up question was added to the final version asking the respondent why they consistently chose the status quo. Basic choice models revealed that all attributes had a significant impact on utility and the signs on the coefficients were as expected. This provided sufficient confidence to proceed with the survey. Overall, the pre-tests revealed that the cognitive load on respondents was considerable, but that the adjustments made the survey easier to complete and more comprehensible for respondents.



#### **4.4 The stated choice experiment**

The stated choice experiment (SCE) is considered the most important part of the survey. By assigning respondents to alternative versions of the SCE, we are able to test a set of hypotheses driven by the research questions described in Section 1.4. The SCE comprises three alternative versions, each being a variation to a generic format described in Section 4.4.1. Section 4.4.2 presents the three versions and their underlying hypotheses. A set of unique choice cards has been generated for each of the three samples in the SCE.<sup>34</sup> Section 4.4.3 describes these experimental designs and how they allow for hypotheses testing. Finally, Section 4.4.4 pays specific attention on the extent to which our design allows to derive choice task specific welfare estimates.

##### **4.4.1 Choice tasks, alternatives and attributes**

Within the SCE each respondent is presented with a sequence of ten choice tasks. The ten choice tasks are split into a set of ‘fixed’ and ‘random’ choice tasks. The two fixed choice tasks are located at the start and at the very end of the SCE. Moreover, this final choice task is identical to the first choice task.<sup>35</sup> The remaining eight (random) choice tasks presented to each respondent were based on a so-called experimental design. More details on the specifics of the fixed choice tasks are provided at the end of this section. The random choice tasks are discussed in Sections 4.4.3.

In each choice task a respondent is presented with a choice card. The choice card consists of two policy alternatives and an opt-out (or status quo) option. Each policy alternative is characterized by four attributes, which vary in their levels over the set of presented policy alternatives. In other words, within the choice card the same policy is not presented twice and the proposed policies vary over the choice sequence. Repetitions of policies, not choice tasks, are not excluded over the choice sequence. Table 4.1 summarizes the attributes and attribute levels applied in the SCE, and describes the status quo (SQ) option. The graphics represent the way each policy attribute was communicated to the respondent. As described in Section 4.3, attribute levels and risk communication measures were extensively pre-tested.





---

<sup>34</sup> The word choice *task* refers to the actual question presented to the respondent at a specific moment in the choice sequence. The task is characterized by a choice *card*, which comes from an experimental design. In other words, in each choice task a respondent is presented with an alternative choice card.

<sup>35</sup> Identical choice tasks were included to test for consistency in responses over the choice sequence. This test is not included in this thesis. We conduct all our analyses in later chapters on choice tasks 2-9 and treat the initial choice task as an instructional choice task explaining the properties of the choice experiment. Indeed, the first choice task included more explanatory text compared to the rest of the choice tasks.

The first attribute describes the probability of flooding within the study area, henceforth labelled as PROB. In the status quo (SQ) option, flood probabilities increase to once every 4,000 years due to climate change. The proposed policies offer improvements, i.e. reductions in flood probabilities, to respectively once every 6,000, 8,000 or 10,000 years. A reduction in flood probability is not necessarily always provided in each policy proposal. Therefore, the PROB attribute can also take the level of 4,000 in specific policy alternatives.

**Table 4.1: Attributes, possible attribute levels and the Status Quo option in the stated choice experiment**

Attribute	Possible attribute levels			
Probability 	1 in 4,000 years	1 in 6,000 years (1.5x smaller)	1 in 8,000 years (2x smaller)	1 in 10,000 years (2.5x smaller)
Compensation 	0%	50%	75%	100%
Available evacuation time 	6 hours	9 hours	12 hours	18 hours
Increase in annual tax 	€40	€80	€120	€160
	Probability	Compensation	Evacuation	Tax
Status Quo	1 in 4,000 years	0%	6 hours	€0

The second attribute covers compensation in case of a flood, henceforth labelled as COMP. Currently, compensation is only provided if the government designates the flooded area as a disaster area. In that case the government acts upon the “Calamities and compensation act” (Botzen and Van Den Bergh, 2008) and provides damage relief. Compensation through this act is, however, highly uncertain and only partial. Therefore, the rate of compensation in the SQ option is determined at 0%. Policies offer improvements of the COMP attribute by covering 50%, 75% or 100% of the suffered material damage. Respondents are informed that compensation payments are done from a compensation fund, which is specifically established for this purpose. The COMP attribute can also remain at the 0% level in the proposed alternatives.

The third attribute takes into account the amount of evacuation time available in case of a coastal flood. Based on the worst case scenario, as discussed in Section 4.1, we have set the SQ level for this attribute at 6 hours. Though the location of a flood may be known longer in advance, organizing and communicating an evacuation plan will take some time. As such, we expect respondents to have approximately 6 hours to leave their residence after being informed about the threat. Investments in improved prediction mechanisms and more efficient organization and communication strategies will increase the amount of available evacuation time for households. The level of the evacuation attribute, henceforth EVAC, can improve to 9, 12 and 18 hours. Again, improvements are not necessarily offered in the proposed policy alternatives.<sup>36</sup>





Finally, the cost attribute covers the annual payment per household to finance the reductions in flood risk exposure provided by the policy alternatives. Payments are in the form of an increase in annual tax per household to the water board, which is the local government body responsible for managing flood defence structures. Naturally, the cost for the SQ alternative is set to €0 as no improvement is provided. For the alternative policy options price levels are €40, €80, €120 and €160 per household per year. Zero payments are not possible, since an improvement is offered in each of the policy options relative to the SQ option. It was emphasized to respondents that these flood risk reducing policies could not be financed by a reallocation of existing funds.

---

<sup>36</sup> To prevent repetition of the SQ option as a policy alternative, we forced policy alternatives to provide an improvement in at least one of the non-price policy attributes (PROB, COMP or EVAC) relative to the SQ.

**Figure 4.2: (Fixed) instructional choice task presented to respondents in the CCF sample**

Please select your best option from the plans presented below:

	Plan A	Plan B	Status Quo
Probability 	1 in 8,000 years (2x smaller)	1 in 10,000 years (2.5x smaller)	1 in 4,000 years
Compensation 	75%	50%	0%
Available evacuation time 	9 hours	18 hours	6 hours
Increase in annual tax 	€120 per year	€160 per year	€0 per year
Choice:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The purpose of the fixed choice task at the start of the stated choice experiment was to familiarize respondents with the settings in the choice tasks. Ladenburg and Olsen (2008) label this as the ‘instructional choice set’. We vary the attribute levels of all policy attributes across alternatives to present the respondent with as much potential attribute levels as possible. The policies presented in the first choice task are equivalent in all three versions of the SCE in terms of the level of the non-price policy attributes, but prices may vary across versions (see Sections 4.4.2 and 4.4.3). An example of an instructional choice task is provided in Figure 4.2. The instructional text explained that policy A is cheaper and offers more compensation, but comes with a higher probability of flooding and less evacuation time relative to policy alternative B.

#### 4.4.2 Hypotheses and versions of the experiment

Respondents were presented with one of three alternative versions of the stated choice experiment, and respectively formed the Choice Certainty Follow-up, High-Starting-Bid and Low-Starting-Bid sample. In this subsection we describe the hypotheses underlying the versions of the SCEs, how they relate to the research questions described in Section 1.4 and where these will be tested in this thesis.

The first research question addresses heterogeneity in preferences across respondents and is not related to a specific version of the stated choice experiment. To this end, Chapter 5 will combine all three samples and analyze variations in response patterns, i.e. marginal willingness-to-pay, across respondents. Since we are interested in preferences at the level of the individual respondent, the experimental design applied is not of too much importance.<sup>37</sup> Hypotheses are not specifically formed for this chapter, neither are they related to the specific versions of the stated choice experiment. Our prime aim is to specify a mixing density that is behaviourally relevant. In terms of the experimental design, this implies that improvements in the non-cost policy attributes should result in a non-negative impact on utility. Tax increases should not have a positive impact on utility. Other evaluation criteria applied in selecting the mixing density are flexibility and ease of estimation. A trade-off is required between these criteria (more details in Chapter 5).

Chapter 6 tests the stability of willingness-to-pay estimates over the choice sequence using the HSB and LSB sample. More specifically, we anchor respondents in both samples on the price attribute in the first choice task. Respondents in the High-Starting-Bid (HSB) sample are all assigned the same initial choice task as depicted in Figure 4.2. The prices used for these policies cover the high-end of the price vector described in Table 4.1. The Low-Starting-Bid (LSB) sample is presented with exactly the same initial choice task, but the policies in this sample differ in their price levels. The annual tax payments were replaced by respectively the values €40 and €80 representing the lower end of the price vector. The subsequent choice tasks are identical in the HSB and LSB samples (more details in Section 4.4.3). By anchoring respondents on the price attribute, we attempt to induce a starting point bias in a similar way as Ladenburg and Olsen (2008). Starting point bias is one method to contrast the predictions of the Discovered Preference Hypothesis and theory of Coherent Arbitrariness (see Section 3.6.1). We approach this topic from a positive side and formulate Hypothesis 1 as close to the micro-economic framework as possible. The micro-economic framework predicts a set of stable preferences over the choice sequence, irrespective of the framing of the choice tasks. Therefore, neither do we expect to find differences in marginal willingness-to-pay estimates between the HSB and LSB sample, nor do we expect to find dynamics in marginal willingness-to-pay over the choice sequence. In contrast to the micro-

---

<sup>37</sup> In fact, variations in the experimental design may induce a particular (correlated) distribution of preferences over the sample of interest. Commonly adopted mixing densities, like the normal distribution, may not be suited to accommodate a distribution of such form. If the shape of the mixing density is flexible enough, it will accommodate impacts of the experimental design, if not, the researcher can control for the impact of the version using additional control variables. The flexibility of the mixing density is of interest in Chapter 5, not the latter.

economic framework, both the DPH and theory of CA predict a discrepancy in marginal WTP estimates between the HSB and LSB sample at the start of the survey, due to this starting point bias. Preference dynamics over the choice sequence will either result in disappearance (DPH) or persistence (CA) of the starting point bias. As such, the HSB and LSB samples are used in Chapter 6 to test for framing effects and preference dynamics over the choice sequence and thereby focus on the second research question.

*Hypothesis 1: Preference parameters are constant over the choice sequence and are not affected by arbitrary framing effects, such as the use of initial value clues.*

Chapter 6 primarily tests for dynamics in preference parameters over the choice sequence induced by a starting point bias. One of the econometric test procedures applied in that chapter also allows controlling for dynamics in the scale parameter over the choice sequence. The scale parameter is the best measure at hand to quantify the degree of package uncertainty experienced by respondents, since it measures overall uncertainty about the utility for a specific policy (see Section 3.3). Preference uncertainty arising at the level of the alternative, can be expected to decrease over the choice sequence due to preference learning, in particular institutional learning (Braga and Starmer, 2005).<sup>38</sup>

Hypothesis 2 sticks to the micro-economic framework predicting no dynamics in the scale parameter over the choice sequence. Chapter 6 and Chapter 7 both pay attention to dynamics in the scale parameter, conditional on a set of stable preferences over the choice sequence (see footnote 7). Hypothesis 1 and 2 are therefore both related to the second research question.

*Hypothesis 2: Scale parameters are constant over the choice sequence and not affected by any possible learning effects.*

Chapter 7 is based on the Choice Certainty Follow-up (CCF) sample, which is characterized by a similar set of choice tasks, but after each choice task respondents are requested to answer a follow-up question regarding their choice certainty. As such, the CCF sample can measure preference uncertainty both in an implicit and explicit fashion through

---

<sup>38</sup> Note that both the DPH and theory of CA do not discuss dynamics in the scale parameter over the choice sequence. Dynamics in the estimated marginal rates of substitution lie at the heart of these models. Moreover, scale dynamics cannot be identified separately from dynamics in marginal rates of substitution (e.g. Swait and Louviere, 1993).

the responses to the choice tasks and follow-up questions respectively. Thereby, the CCF sample is used in Chapter 7 to answer research question three. If we assume that responses to the choice certainty follow-up questions are in line with the true degree of choice certainty experienced by the respondent, then the responses to these follow-up questions can be used together with the responses to the choice tasks in a unified model. Since both the choice task and the follow-up question embody information about the degree of preference uncertainty, more (statistical) information is at the disposal of the researcher and more valid and reliable willingness-to-pay estimates may therefore be obtained when the measure of interest is affected by preference uncertainty at the level of the respondent. Hypothesis 3 summarizes these arguments and states that self-reported choice certainty questions can be a useful tool in deriving more valid and reliable willingness-to-pay estimates. For example, the follow-up question can pick up the complexity of the choice task at hand, because when the policy alternatives in the choice task are closer in utility levels compared to other choice tasks, the decision becomes more difficult and respondents may report a lower level of choice certainty. The response to the follow-up question can thereby help in identifying the size of the scale parameter in a specific choice task. However, as discussed in Chapter 7, self-reported choice certainty responses may be subject to a measurement error of true preference uncertainty. Moreover, their use may potentially induce endogeneity issues, which can have detrimental impacts on the validity and reliability of derived welfare measures.

*Hypothesis 3: Self-reported choice certainty responses are in line with the actual degree of preference certainty experienced by respondents in stated choice experiments. Therefore, such follow-up questions are a useful tool in improving willingness-to-pay estimates when response patterns are affected by preference uncertainty.*

Table 4.2 summarizes how the alternative versions of the stated choice experiment are used to test the hypotheses and appear in the following chapters. The set of hypotheses in combination with the setup of the alternative versions of the stated choice experiment allows for a thorough analysis of preference uncertainty and related preference dynamics, including possible learning effects. The next subsection continues by discussing the design of the choice cards presented to the respondents.

**Table 4.2: Overview of the samples, the chapters in which they are used, and the related hypotheses**

	CCF	HSB	LSB
Description	Choice certainty follow-up questions	High starting bids in the first choice task	Low starting bids in the first choice task
Purpose	Implicit and explicit measurement of preference uncertainty	Contrast DPH and CA	Contrast DPH and CA
Used in testing hypotheses:			
Hypothesis 1		x	x
Hypothesis 2		x	x
Hypothesis 3	x		
Applied in chapters:			
Chapter 5	x	x	x
Chapter 6		x	x
Chapter 7	x		

#### 4.4.3 Experimental design of the random choice tasks

The eight random choice tasks presented to each respondent were based on a fractional factorial experimental design. A total of 24 unique choice cards were generated per sample, divided into three blocks of eight choice cards. Every respondent was presented with one of those blocks. The attribute levels described in Table 4.1 were used as inputs in constructing these choice cards. The following common design criteria were applied to the generation of the design across all samples. First, it was ensured that the fixed choice task presented at the start of the survey was not repeated in the experimental design. Second, the SQ alternative could not be repeated within the other policy alternatives. An improvement was always offered for at least one of the non-price attributes. Third, since all attributes characterizing the policy alternatives were expected to have a clear positive (or negative) impact on utility we excluded the possibility of dominant alternatives. That is, improvements in PROB, COMP and EVAC are expected to have a positive impact on utility, while additional taxes are perceived as negative. Accordingly, policies which offer higher improvements on all non-price attributes but come at a lower price than the other policies are not permitted, because respondents are not really faced with a trade-off in that case. The sample specific designs were generated using the software package NGENE (2010).

The designs applied within the CCF, HSB and LSB samples are so-called d-efficient designs (e.g. Scarpa and Rose, 2008). The efficiency of a design is related to the informational content available from a set of observed choices. More specifically, researchers want to minimize standard errors in the estimation of the parameters of the utility function. In this



way, changes in marginal willingness-to-pay (WTP) and consumer surplus can be identified as accurately as possible. Standard errors are obtained from the asymptotic covariance matrix of parameter estimates in the model, i.e. the negative inverse hessian matrix, which is based on the information matrix. Besides the standard errors, the covariance matrix also informs the researcher about the extent to which parameters of interest can be identified independent from other parameters (explanatory variables) in the model. High correlation between parameter estimates in the model indicates confounding between the effects of explanatory variables. To prevent confounding between parameter estimates, much attention has been paid to the generation of orthogonal designs minimizing correlation between the explanatory variables. Huber and Zwerina (1996) highlight that the principle of orthogonality is not as relevant for highly non-linear models, like the family of logit models, as it is for linear regression models. Therefore, the use of efficient instead of orthogonal designs, which enable the researcher to obtain parameter estimates with smaller standard errors, has been proposed (e.g. Rose and Bliemer, 2009).

In principle, efficient designs are generated by maximizing the information matrix of the proposed model specification, i.e. minimizing the asymptotic variance covariance (AVC) matrix. Since a matrix is multi-dimensional, optimization is not possible unless the researcher specifies a one-dimensional measure summarizing the information matrix and thereby allowing the comparison of efficiency across alternative designs. Kessels et al. (2006) provide an overview of alternative efficiency measures of which the d-efficiency measure is applied most frequently. A more recent discussion on d-efficient designs is provided in Vermeulen et al. (2011). A d-efficient design is based on minimization of the determinant of the AVC matrix and can be used to compare designs with a similar number of parameters. Rose et al. (2008) note that the determinant of the AVC matrix depends on the elements of the design, i.e. the levels and use of the attribute levels as explanatory variables, and the value of the parameters themselves. Regarding the latter, since designs are specified prior to the actual SCE, efficient designs are optimized based on a set of prior expectations about the parameters in the model (e.g. Ferrini and Scarpa, 2007).<sup>39</sup>

We have selected the d-efficiency measure not only for its common application, but also because it takes into account the off-diagonal elements of the AVC matrix. For example, the a-efficiency measure (Kessels et al., 2006) minimizes standard errors based on the trace of

---

<sup>39</sup> Uncertainty about proper priors can be taken into account by specifying a distribution of prior values and comparing designs based on expected efficiency measures. This is done by averaging the efficiency criterion over a sequence of draws from the prior distribution. This is also referred to as a Bayesian efficient design.

the AVC matrix, neglecting possible confounding between parameter estimates. Moreover, due to our interest in accurate welfare estimates we do not apply G- and V- optimality criteria, which look at minimizing prediction errors rather than standard errors of parameters (Kessels, Goos and Vandebroek, 2006). More recently, efficiency criteria are being developed taking into account the standard errors of marginal WTP estimates. For example, Vermeulen et al. (2011) propose the use of a WTP-optimality criterion minimizing the trace of the variance covariance matrix of marginal WTP estimates. Though suitable for our purposes, this type of efficiency criterion was not yet available within the NGENE software version used in this thesis.

Finally, it should be noted that an efficient design represents a pseudo optimal design for a specific model specification and is conditional on the range of attribute levels and associated set of prior parameter values. In general, the number of possible designs is too large to evaluate. Hence, a range of design algorithms has been developed to systematically optimize the design (see Kessels, Goos and Vandebroek, 2006 for a discussion). The optimal design remains, however, unknown explaining the choice of wording for *d-efficient* rather than *d-optimal* designs. Efficiency is increasing in the number of choice tasks presented to a respondent, and in the number of respondents answering a particular sequence of choice tasks. For a given amount of respondents, efficiency can increase substantially using efficient designs relative to standard orthogonal designs (Huber and Zwerina, 1996; Rose and Bliemer, 2009). The latter improvement is decreasing in sample size and only holds if the model is correctly specified. A priori, it is unclear how alternative designs perform under misspecification of the model. More robust efficient designs can be specified by optimizing the average efficiency criterion over a set of alternative model specifications. However, Ferrini and Scarpa (2007) show in a simulation study that under misspecification researchers might be just as good off with designs based on standard (orthogonal) fractional factorial designs for linear models. Huber and Zwerina (1996) still find an improvement in efficiency, despite severe misspecification in their prior parameter estimates. A detailed analysis on this issue is beyond the scope of this thesis.

Now we discuss the experimental designs underlying the three samples in more detail. The 24 choice cards in the CCF design were based on a standard multinomial logit (MNL) attributes only model with prior parameter values based on the results from the online pre-test

among 100 respondents.<sup>40</sup> Within the three blocks of eight choice cards attribute level balance is maintained ensuring respondents are not only presented with high or low levels of a particular attribute. Second, based on the prior parameter values the set of dominant alternatives could be defined and therefore ruled out by the syntax. The final d-efficiency measure for the CCF sample was  $d\text{-error} = 1.5 E-05$ . After obtaining the data from the CCF sample a quick analysis was conducted to update the priors in the model for the HSB and LSB samples. For these two samples the d-efficient design was based on a panel random parameter logit model. Independent normal distributions were identified for the non-price attributes and a zero mean error-component term was assigned to the non-SQ policy alternatives (e.g. Scarpa et al., 2005). The underlying simulations required to evaluate the AVC matrix were based on a sequence of 100 Halton draws (Halton, 1960) over 200 respondents, which would be just below the expected effective sample size within each sample. A common design (d-error 0.001) was generated for the HSB and LSB sample in order to contrast within and between sample preference dynamics in a clean experimental setting. By imposing the same experimental design, potential differences in preferences can only be induced by anchoring respondents on the price vector in the first choice task or due to heterogeneity in preferences across respondents.

#### **4.4.4 Choice task specific parameter estimates**

Given our focus on preference dynamics over the choice sequence, it is crucial that each choice card in the design is presented at each moment in the choice sequence to various respondents. If this is not the case, it becomes hard to derive accurate willingness-to-pay estimates at the choice task specific level. Moreover, by presenting alternative trade-offs to respondents at different moments of the sequence it is impossible to disentangle the impact of the design from possible learning effects. Ladenburg and Olsen (2008) study preference dynamics while presenting a single choice card to all respondents at a particular moment during the choice sequence. Their results may have been affected by the issue pointed out above. We work around it by systematically rotating the order of appearance of the choice cards contained in each block of eight choice cards. That is, version 1 presented respondents with choice cards 1-8 in ascending order. Version 2 started with choice cards 2-8 and ended with choice card 1. This rotation procedure yielded 8 versions per block adding to a total of

---

<sup>40</sup> The number of choice cards in the design is higher than the number of choice tasks per respondent in order to increase the number of identifiable parameters. The design is efficient for the combination of three blocks of eight choice cards.

24 versions. Note that this is done for each of the three samples. Finally, we varied the order of appearance of the first and second policy alternatives to prevent effects from reading from left to right. Accordingly, the number of versions was doubled to 48 per sample. Respondents within a sample were randomly assigned to one version. We exploit this particular property of the design in Chapter 6 when testing for preference dynamics over the choice sequence.

#### **4.5 Summary**

The case study presented in this thesis relies on an online survey considering flood risk exposure in the Netherlands in the face of climate change. The main part of the survey consists of a stated choice experiment focused at deriving a marginal WTP estimate for reductions in coastal flood risk exposure. Respondents are randomly selected residents of the provinces of North- and South Holland who live within dike rings 13 and 14, which are currently characterized by a flood probability of once every 10,000 years. The population density and presence of major economic activities within the dike rings contribute to the low probability – high impact setting of the stated choice experiment. In combination with limited experience with flood risks and related (monetary) trade-offs, the low probability-high impact setting is likely to result in preference uncertainty regarding the proposed flood risk reducing policies. By building on existing stated preference studies regarding flood risk exposure in the Netherlands and by going through a sequence of pre-testing stages, the survey gradually informed respondents in a comprehensible fashion about the low probability – high impact characteristics of their coastal flood risk exposure level (see Sections 4.2 and 4.3). The careful development and testing of risk communication measures helped in bringing about the required information to respondents, but is unlikely to have entirely eliminated preference uncertainty at the level of the respondent. To test for the impacts of preference uncertainty on welfare estimates and related preference dynamics over the choice sequence, a set of three alternative versions of the stated choice experiment is developed. Section 4.4 described the basic hypotheses serving as the basis for the alternative versions and related them to the main research questions defined in Chapter 1. In the next chapters these hypotheses will be systematically tested. The unique aspect of this thesis can be found in the careful design of the version specific choice cards to match these hypotheses, as described in Sections 4.4.3 and 4.4.4. Not only does the design allow for a comparison of preferences across versions of the SCE, but also across different moments in the choice sequence.



## Chapter 5: Simple asymmetric mixing densities for MMNL models<sup>41</sup>

The mixed multinomial logit model (MMNL), also referred to as the random parameters logit (RPL) model or random coefficients logit (RCL) model, currently represents the most popular econometric model used to analyze discrete choice type data. Rather than being a single model, the MMNL model represents a family of models with alternative specifications offering various potential benefits over standard multinomial logit models. Advantages of the MMNL model include i) the ability to model heterogeneity in the patterns of choices across respondents<sup>42</sup> (see e.g. Ben-Akiva et al., 1993; Bhat, 1998; Brownstone and Train, 1998; Hensher and Greene, 2003; McFadden and Train, 2000), ii) non-constant error variances across alternatives via a relaxation of the Independence of Irrelevant Alternatives assumption (see e.g. Hensher and Greene, 2003; Train, 2009) and iii) the potential accommodation of correlation in choices made by the same respondent (see e.g. Revelt and Train, 1998). In this chapter the version of the MMNL used takes into account that patterns of choice vary across respondents and that a single respondent makes a sequence of choices rather than a single choice.<sup>43</sup> We focus on the first research question by investigating the appropriateness of alternative specifications of the probability density functions underlying the random parameters in the MMNL model and their impact on marginal WTP estimates. The main evaluation criteria applied are behavioural relevance and ease of estimation. To this end we introduce a new probability density function in the MMNL framework, the asymmetric triangular distribution, which produces behaviourally relevant marginal WTP estimates while maintaining a simple functional form at the same time.

### 5.1 Flexibility and empirical identification

Controlling for heterogeneity in response patterns across respondents is a complex task. It is generally not feasible to identify for each respondent the true set of preference parameters. The informational content of commonly applied multi-attribute and multi-alternative choice

---

<sup>41</sup> This chapter is based on joint work conducted with Prof. John Rose during a research visit to the Institute for Transport and Logistic Studies at the University of Sydney. A paper version is submitted to the Journal of Choice Modelling.

<sup>42</sup> Whilst it is common to interpret the random parameter coefficients as representing purely preference heterogeneity, due to the perfect confoundement with scale and preference in most discrete choice models, including the MMNL model, any modelled heterogeneity should more correctly be interpreted as representing a mixture of both preference and scale or error heterogeneity over the sample (Fiebig et al., 2010; Hess and Rose, 2011). For this reason, the heterogeneity is not simply referred to as preference heterogeneity, but rather as heterogeneity in the patterns of choices.

<sup>43</sup> This chapter does not account for intra-respondent heterogeneity as explained in Section 3.3.1.

experiments is typically not enough to obtain sufficiently small confidence intervals around individual specific parameter estimates. Moreover, preference parameters are likely to be confounded with each other and scale at the individual level (Rouwendaal et al., 2010). The primary cause for this identification problem is the limited number of choice tasks typically presented to respondents within a choice experiment. Recent developments in the experimental design literature resulted in improvements in informational content based on a given number of choice cards (e.g. Rose and Bliemer, 2009; Scarpa and Rose, 2008), but additional steps still need to be made in order to derive individual specific parameter estimates (Louviere et al., 2008). Accordingly, the true distribution of response patterns over the population of interest is generally approximated by a (continuous or discrete distribution), also known as the mixing density. The mixing densities discussed in this chapter can easily be applied to the cross-sectional random parameters logit model, which allows preferences to vary across respondents and choices, but does not take into account the fact that a single respondent makes multiple choices.<sup>44</sup> Nevertheless, this chapter adopts the panel MMNL as discussed in Revelt and Train (1998) and Section 3.3.1.

In selecting the mixing density that best approximates the true distribution of response patterns, the researcher is faced with a range of trade-offs. First, a decision needs to be made between a (i) multivariate distribution capturing correlation in preferences across particular attributes or (ii) a set of independent distributions. In the framework of continuous distributions, the choice for a multivariate mixing density restricts the set of available distributional forms to the normal distribution and its related transformations, simply because it is one of the few distributional forms which can take into account correlation patterns across random parameters (Train and Sonnier, 2005).<sup>45,46</sup> Latent class (LC) models represent the multivariate case for discrete distributional forms (Greene and Hensher, 2003). Correlation in preferences over the attributes is taken into account by estimating a block of parameters for specific types of choice patterns, i.e. classes. Since each class represents a specific mass-point, these correlated discrete distributions allow latent class models to take a very flexible

---

<sup>44</sup> Applications of cross-sectional mixed logit models in a Bayesian framework can be more problematic as the prior tends to dominate the posterior and shrinks the observation specific augmented parameters to the mean (Rossi et al., 2005).

<sup>45</sup> Correlation can also be induced by multiplying all (or some) random parameters with a common scalar, as discussed by Fiebig et al. (2010) and Hess and Rose (2011). This is not part of this thesis.

<sup>46</sup> Model estimations within the classical maximum likelihood framework generally do not take into account such correlation patterns. Within the Hierarchical Bayesian (HB) literature, applications of multivariate (normal) mixing densities are more common (e.g. Rigby and Burton, 2006). An important cause of this property is that the stability of the off-diagonal Cholesky terms is questionable when using a relatively small number of draws in the maximum likelihood estimation procedure. Since HB-models do not rely on optimization procedures, such stability issues are no longer relevant.

form. Allenby and Rossi (1998) and Elrod and Keane (1995), however, note that by using discrete mass-points the model may understate the degree of heterogeneity in the data. Burda et al. (2008) alleviate this limitation by estimating a class specific multivariate normal distribution within the LC model. Such mixture of normal models are also known as “Bayesian semi-parametric” models, which can represent an (in)finite number of classes and thereby approximate the true distribution of preferences over the population of interest. The number of parameters required estimating such a model, or a LC model in general, creates, however, a major drawback. That is, the number of parameters in the LC model increases in a linear fashion with the number of classes. Consequently, estimation in the classical maximum likelihood framework becomes problematic if the number of classes becomes too large (Train, 2008). Moreover, too many classes may also result in an over-identification of the model and produce large standard errors and confounding between parameter estimates. On the contrary, the multivariate normal distribution is not as flexible in shape as the mixing densities obtained by means of the LC or the Burda et al. (2008) model. Transformations of the multivariate normal distribution, like the Johnson SB distribution, can take a more flexible form (e.g. Train and Sonnier, 2005). Train and Sonnier (2005), however, note that estimation of the Johnson SB is frequently hampered by empirical identification issues due to high degrees of correlation between the hyper-parameters characterizing the distribution.

By selecting a set of independent distributions, more alternative (continuous) distributional forms are at the disposal of the researcher. These range from simple and (or) bounded distributions, like the standard normal, uniform and triangular, to more complex shapes like the beta, gamma and Weibull distribution (e.g. Meijer and Rouwendal, 2006). The flexible shape offered by the latter distributions comes at the cost of additional parameters characterizing the location, shape, scale and support of the distribution.<sup>47</sup> For example, relative to a standard normal distribution, the beta distribution is characterized by two additional hyper-parameters. As in the multivariate case, a discrete mixing density can be specified. However, for each mass-point added, a probability and preference parameter needs to be estimated. Relative to the LC model, the independent discrete distributions model is more complex, since the closed form of the model requires the evaluation of all potential classes in the model. That is, in a model with two discrete random parameters, both having three mass-points, already nine classes (all possible combinations of mass-points) are present. The number of classes increases rapidly as additional random parameters and mass-points are

---

<sup>47</sup> The support of a distribution refers to the levels the variable of interest can take. For example, the support of a standard uniform distribution is defined over [0,1].



added to the model.<sup>48</sup> Similar to the Burda et al. (2008) model, Campbell et al. (2010a), Fosgerau and Hess (2009) and Train (2008) introduce the use of mixture modelling for a set of independent normal distributions in the classical framework. Combining discrete and continuous distributions again increases flexibility, but increases the complexity of the model and therefore requires highly informative datasets. Fosgerau and Bierlaire (2007) increase the flexibility of the mixing density by transforming a base distribution using a Legendre series polynomial. Compared to the use of discrete distributions additional flexibility in this semi-nonparametric (SNP) approach only requires the specification of a single additional parameter. Nevertheless, the complexity of the model still rises and properties of the selected base distribution, like the support, are still present in the final approximation of the true density. Finally, non-parametric approaches are being developed within the classical estimation framework allowing for flexibility in preferences over specific dimensions (e.g. Fosgerau, 2007; Koster and Koster, 2011). Such non-parametric approaches are discussed in more detail in Chapter 6 of this thesis.

## **5.2 Behavioural relevance and ease of estimation**

The discussion above made clear that more flexible distributions can approximate the true mixing density more accurately, but the computational burden and requirements related to the data and estimation procedure increase substantially. As such, empirical identification and convergence issues may arise. This applies to both multivariate and independent specifications of the mixing density. Balcombe et al. (2009) introduce the Bayesian concept of marginal likelihood to identify the most appropriate distributional form, using statistical model fit as the prime selection criterion. Alternatively, information criteria can be used to compare the fit of various (non-)nested models (e.g. Train, 2008). Both approaches penalize for the inclusion of additional parameters in the model. Model fit and information criteria, however, do not take into account the behavioural relevance of the model. For example, respondents are not expected to experience positive utility from increasing prices for exactly the same product. Strictly positive price coefficients should therefore not be included in the support of the mixing density.

Two issues contrasting aspects should be clarified here. First, in principle the support of the mixing density should not be restricted to a particular domain during estimation. Non-restricted mixing densities can identify whether there are issues with the data. For example, if

---

<sup>48</sup> Moreover, model estimation becomes complex as each preference and probability parameter affects multiple classes. Within the latent class model each parameter only affects a single class.

the range for the levels for the cost attribute in the stated choice experiment is not large enough, then respondents may not trade with respect to price and display zero (or positive) cost sensitivity. Second, including zero in the support of the mixing density for the cost coefficient may seem feasible, but this will cause issues in deriving WTP measures (Daly et al., 2011b). A trade-off therefore needs to be made between representing choice behaviour or deriving valid moments for the WTP distribution.

A common approach to work around the second point is to model the cost coefficient in the form of a lognormal distribution. Although this restricts marginal utility to the positive domain, or negative after multiplying the distribution by -1, a fat upper tail is imposed implying some respondents exhibit very high cost sensitivity.<sup>49</sup> Hensher and Greene (2003) therefore put emphasis on the development of truncated or constrained distributions, of which the (constrained) triangular distribution has been applied most frequently (e.g. Brouwer et al., 2010; Greene et al., 2006). A last consideration is the ease of implementation. Complex models are generally not included in standard software packages and significantly increase computation time. Particularly continuous random parameter models increase computation time, due to their reliance on Monte Carlo simulations to approximate expected choice probabilities (Hess et al., 2006).

Overall, selecting the mixing density becomes a delicate balance between flexibility, ease of estimation, computation time and behavioural relevance. Hence, this calls for the implementation of simple, but flexible distributional forms complying with behavioural expectations while taking into account that misspecification of the mixing density can result in biased welfare estimates (Fosgerau, 2006). Picking up the suggestions by Hensher and Greene (2003), the use of a new distribution is proposed, i.e. one that has not been applied before in MMNL models. The proposed (a)symmetric triangular distribution has a set of convenient properties. It has a bounded support, which can be restricted to a particular domain, and does not result in fat (upper) tails and may therefore produce more reliable marginal WTP estimates. In addition, the mode of the distribution can be defined independent of the lower and upper bound. Hence, the distribution can accommodate any type of skewness present in the distribution of preferences over the population of interest. This is in contrast to the normal and lognormal distribution, which cannot take into account skewed distributions or

---

<sup>49</sup> Another inconvenient property of the lognormal model, as discussed by Rigby and Burton (2006), is the lack of mass at zero, precluding the event of respondents having a zero marginal utility (or WTP) for a particular attribute.

only right skewed distributions.<sup>50</sup> By having only three parameters, a relatively simple model structure arises, not expected to introduce empirical identification issues. However, the simplicity of the proposed density comes at a cost. It cannot take into account multi-modality and, like the normal distribution, is sensitive to observations in the tails. Finally, the triangular density cannot account for correlation across parameters. Whether its advantages outweigh its disadvantages remains an empirical question.

In the following sections, the proposed triangular distribution will be applied to our case study. The asymmetric triangular mixing density will be contrasted with a set of common specifications of the mixing density, including the standard normal and lognormal distribution. Model estimations are conducted within a Bayesian framework to facilitate the comparison of the non-nested model structures. Moreover, models will be estimated directly in WTP-space as this is expected to increase reliability of WTP estimates (Campbell et al., 2010a; Scarpa, Thiene and Train, 2008) and prevents issues with the identification of the moments of the marginal WTP distribution when estimated in preference space (Daly, Hess and Train, 2011b). Compared to models in preference space, WTP-space models allow in some cases for alternative distributional patterns (Train and Weeks, 2005) The model set-up is discussed in more detail below.

### 5.3 Contrasting alternative continuous distributions

Adopting the notation introduced in Chapter 2, the utility respondent  $i$  derives in choice task  $t$  from alternative  $j$  is described by the utility function  $U_{ijt}$  in (5.1). The specification in WTP-space is a transformation of the standard linear utility specification, where  $X_{ijt}$  represents a vector of explanatory variables, including the policy attributes but not price, and  $q_{ijt}$  the price of the specific alternative.  $\varepsilon_{ijt}$  follows a standard i.i.d. extreme value distribution. Individual specific marginal WTP parameters  $\omega_i$  are obtained by dividing the marginal utility parameters  $\beta_i$  by the cost parameter  $\alpha_i$ .<sup>51</sup> The restriction  $\alpha_i > 0$  is imposed to ensure price has a negative impact on utility.<sup>52</sup> Unless indicated otherwise, we comply with this restriction by specifying a lognormal distribution for  $\alpha_i$ . Alternative independent mixing densities are applied for  $\omega_i$ . In generic notation, the joint mixing density is denoted by  $f(\alpha_i, \omega_i | \Omega)$ , where  $\Omega$  comprises the set of hyper-parameters. Accordingly, the likelihood function for the mixed logit model using

---

<sup>50</sup> By mirroring the lognormal distribution in the origin makes it possible to take into account left-skewed distributions. This is, however, a decision of the researcher prior to estimation, not something revealed by the data.

<sup>51</sup> By taking the ratio of two parameters, marginal WTP parameters are no longer confounded with the scale parameter, except for the cost parameter  $\alpha_i$  and any fixed coefficients  $\beta_j$  in the model.

<sup>52</sup> This restriction may result in a reduction in model fit (Train and Weeks, 2005).

continuous distributions is described by (5.2), where  $y$  denotes the vector of observed choices. Denote  $j$  as the chosen alternative in choice task  $it$ , i.e.  $y_{it} = j$ . In the actual estimation a fixed alternative specific constant  $Z_{ijt}$  is included in the model, which is represented by the fixed coefficient  $\beta_f$  and influences the likelihood function through the MNL choice probability.

$$(5.1) \quad U_{ijt} = Z_{ijt}\beta_f + \alpha_i(X_{ijt}\omega_i - q_{ijt}) + \varepsilon_{ijt}$$

(5.2)

$$L(y | X, Z, q, \beta_f, \Omega) = \prod_{i=1}^n \int_{\alpha_i, \omega_i} \left( \prod_{t=1}^T \frac{\exp(Z_{ijt}\beta_f + \alpha_i(X_{ijt}\omega_i - q_{ijt}))}{\sum_{k \in D_{it}} \exp(Z_{ikt}\beta_f + \alpha_i(X_{ikt}\omega_i - q_{ikt}))} f(\alpha_i, \omega_i | \Omega) \right) d(\alpha_i, \omega_i)$$

### 5.3.1 Data augmentation and prior distributions

The Bayesian estimation framework relies on exactly the same likelihood function as the classical estimation framework, but offers through data augmentation a convenient way to work around the integral over all possible values for  $\alpha_i$  and  $\omega_i$  (e.g. van Dyk and Meng, 2001). Conditional on the individual specific value of the random variable(s), the likelihood function reduces to the standard MNL specification

$$L(y | X, Z, q, \beta_f, \alpha, \omega) = \prod_{i=1}^n p(y_i | X, Z, q, \beta_f, \alpha_i, \omega_i), \text{ since the mixing density only}$$

influences the likelihood function through  $\alpha_i$  and  $\omega_i$ . The latter are no longer treated as latent variables, but as “known” model parameters with actual realizations under data augmentation. Bayesian estimation requires the specification of a set of prior distributions for all model parameters, including the augmented variables. For the fixed parameters a multivariate normal prior is applied with mean vector  $m_f$  and covariance matrix  $V_f$ . The mixing density  $f(\alpha_i, \omega_i | \Omega)$  describes the uncertainty about the random parameters  $\alpha_i$  and  $\omega_i$  and therefore serves as a prior. Still, the hyper-parameters of the mixing density  $\Omega$  are treated as model parameters requiring an additional layer of prior distributions.

In this chapter four different types of (continuous) mixing distributions are contrasted: (i) the independent normal; (ii) the independent lognormal; (iii) the independent censored normal; and (iv) the asymmetric triangular distribution.<sup>53</sup> The normal mixing distribution is characterized by two parameters, the mean  $\mu$  and the variance  $\sigma^2$ . The vector of mean

---

<sup>53</sup> Since the triangular distribution cannot control for correlation across random parameters, we do not contrast it against correlated (transformations) of the normal distribution.

parameters  $\mu$  is assigned a normal prior with mean vector  $m_r$  and variance  $V_r$ . The prior for the variance term  $\sigma^2$  is specified by an inverse gamma distribution with shape and scale parameters  $\eta$  and  $\kappa$ . These are conjugate priors for the hyper-parameters of the normal distribution facilitating drawing from the conditional posterior densities in the Gibbs Sampler (GS) (e.g. Koop, 2003; Train, 2009).<sup>54</sup> The probability density function of the asymmetric triangular distribution described in (5.3) is characterized by three hyper-parameters  $a, b$  and  $c$ . The lower and upper bounds  $a$  and  $b$  define its support, while the mode  $c$  determines the skewness of the density. The set of priors for these hyper-parameters need to satisfy the constraints on their ordering.<sup>55</sup> We define a set of uniform priors and restrict the prior on  $a$  in (5.4) to fall between  $l_a$  and  $u_a$ , respectively  $a$ 's lower and upper bound.<sup>56</sup> Conditional on  $a$ , the prior for the mode  $c$  can be restricted to the interval  $[a, u_a]$ , which is again assigned a uniform density with density  $(1/(u_a - a))$ . A similar conditional uniform prior distribution is constructed for  $b$  over the interval  $[c, u_a]$  with density  $(1/(u_a - c))$ . Appendix 5.A describes input values for the prior distributions applied in the alternative model specifications presented in Section 5.4.

$$(5.3) \quad p(\omega_i | a, b, c) = \begin{cases} \frac{2(\omega_i - a)}{(b - a)(c - a)} & \text{for } a \leq \omega_i \leq c \\ \frac{2(b - \omega_i)}{(b - a)(b - c)} & \text{for } c \leq \omega_i \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$(5.4) \quad p(a | l_a, u_a) = \begin{cases} \frac{1}{u_a - l_a} & \text{for } l_a \leq a \leq u_a \\ 0 & \text{otherwise} \end{cases}$$

### 5.3.2 Conditional posterior distributions

Applying Bayes' rule by combining the joint prior distribution, including the mixing densities, and the augmented likelihood function, results in the full posterior distribution,

---

<sup>54</sup> Train and Sonnier (2005) explain that the lognormal ( $\exp(n(\mu, \sigma^2))$ ) and censored normal ( $\min(0, n(\mu, \sigma^2))$ ) distributions are simple transformations of the normal distribution. Minor adjustments are required in the calculation of the utility function.

<sup>55</sup> We work with the lower and upper bound, because it allows easily restricting the distribution to a particular domain. Alternatively, one can work with a location parameter for  $c$  and two spread parameters  $s_1$  and  $s_2$  alleviating the imposed ordering restrictions.

<sup>56</sup> Bounds on the support of the prior are introduced to generate a set of proper priors, such that the marginal likelihood can be derived for the purpose of model comparison. Proper priors imply that the mass under the prior density integrates to one. If the Gibbs Sampler tends to go to the bounds, they can easily be extended such that the prior has no influence on the parameter estimates.

which is hard to evaluate analytically. Hence, a set of conditional posteriors is defined. Train (2009) shows that parameters from the conditional posteriors for the fixed parameters  $\beta_f$  in (5.5) and the augmented parameters  $\alpha_i$  and  $\omega_i$  in (5.6) and (5.7) need to be drawn using a Metropolis-Hastings (M-H) algorithm, irrespective of the specified mixing distribution.<sup>57</sup> The latter is a result of the presence of the MNL choice probability of the chosen alternative in the conditional posterior density, abbreviated here to  $P_{ijt}$ . The term cannot be rewritten in a convenient distributional form from which one can easily take draws. Note the specification of respectively a (standard, log or censored) normal and triangular mixing density in (5.6) and (5.7). Transformations from the normal distribution to the lognormal or censored normal are dealt with in the calculation of  $P_{ijt}$ .

$$(5.5) \quad p(\beta_f | m_f, V_f, \alpha, \omega, y, X, Z, q) \propto \exp\left(-\frac{1}{2}(\beta_f - m_f)' V_f^{-1} (\beta_f - m_f)\right) \prod_{i=1}^n \prod_{t=1}^T P_{ijt}$$

$$(5.6) \quad p(\alpha_i | \mu, \sigma^2, \beta_f, \omega_i, y, X, Z, q) \propto \exp\left(-\frac{1}{2}(\alpha_i - \mu)' (\sigma^2)^{-1} (\alpha_i - \mu)\right) \prod_{t=1}^T P_{ijt}$$

$$(5.7) \quad p(\omega_i | a, b, c, \alpha_i, y, X, Z, q) \propto \prod_{t=1}^T P_{ijt} \cdot \begin{cases} \frac{(\omega_i - a)}{(c - a)} & \text{for } a \leq \omega_i \leq c \\ \frac{(b - \omega_i)}{(b - c)} & \text{for } c < \omega_i \leq b \\ 0 & \text{otherwise} \end{cases}$$

Unlike the case of the (transformed) normal distribution (as described in Train, 2009), the hyper-parameters of the triangular distribution do not have a conditional posterior distribution of convenient form that can be used to generate draws. Additional M-H steps are therefore required in the Gibbs Sampler. For completeness, the full set of conditional posteriors is reported in Appendix 5.B.

---

<sup>57</sup> Metropolis-Hastings algorithms assist in generating draws from a density of which one cannot easily draw a random number. It uses a candidate generating function from which one can easily draw values. The true density is evaluated at the candidate draw and contrasted against the value of the density at the previous draw. Broadly speaking, the candidate draw is only accepted if it outperforms the previous draw, otherwise the old draw is maintained. Hence, the challenge is in finding a candidate density function close to the true density function to increase efficiency of the GS. See for more details Train (2009, chapter 9.2.9)

### 5.3.3 The Gibbs Sampler

Model estimation proceeds by repeatedly drawing in a stepwise fashion from the set of conditional posterior distributions. First, a set of starting values for all model parameters are defined, which are then fed into a particular conditional posterior. Conditional on the starting values, a draw from the conditional posterior is taken to update the value of the chosen parameter. The updated parameter is then fed into another conditional posterior to update that one's value. Both updated values are then fed into the next conditional posterior. When all parameters are updated, the process is repeated for a number of rounds. This sampling procedure is called the Gibbs Sampler (GS). The draws obtained from the GS are assumed to converge to draws from the joint posterior distribution and thereby provide information on the parameters of interest (e.g. Koop, 2003). To ensure convergence, a set of burn-in iterations is specified. Here, 10,000 burn-in draws are applied for the models based on a (transformed) normal mixing distribution and 40,000 for the triangular. Then a set of draws is maintained to characterize the posterior distribution (30,000 for the normal models and 100,000 for the triangular). The triangular model requires more draws, because the sampler is less efficient, due to the use of additional M-H steps in the GS. Arbitrary starting values were used for the model parameters ( $a, b, c, \beta_f$  and  $\sigma^2$ ) to initiate the GS. The augmented variables  $\alpha_i$  and  $\omega_i$  were also initialized within those bounds. The models showed limited sensitivity to the specified starting values.

Model estimations were conducted in MATLAB and random walk chains are applied for the M-H steps in the Gibbs Sampler (e.g. Chib and Greenberg, 1995). Convergence of the GS was monitored by examining plots of the posterior draws and using Geweke's convergence diagnostic (Geweke, 1992). To reduce the autocorrelation across draws from the GS only every 3<sup>rd</sup> draw is kept in case of a (transformed) normal density and every 10<sup>th</sup> draw in the triangular case. This process is commonly labelled as thinning and the set of 10,000 maintained draws serves as the basis for the posterior analysis and model comparison.

### 5.3.4 Model comparison

For model comparison the method of Gelfand and Dey (1994) is used as introduced in the mixed logit framework by Balcombe et al. (2009). Bayesians contrast the model fit of models  $M_1$  and  $M_2$  by means of the posterior odds ratio  $p(M_1/y)/p(M_2/y)$ , where  $p(M_1/y)$  represents the probability that  $M_1$  is the correct model after observing the data. Simple rules of probability imply  $p(M_1/y) = p(M_1)p(y/M_1)$ . Commonly, the prior probability  $p(M_1)$  that  $M_1$  is the correct model is set to 0.5, such that  $p(M_1)$  and  $p(M_2)$  cancel out in the posterior odds ratio, which

then only depends on the ratio of marginal likelihoods  $p(y/M_1)/p(y/M_2)$ . The ratio of marginal likelihoods is also known as the Bayes Factor. The marginal likelihood of a model is equivalent to the normalizing constant, which ensures that the joint posterior density of model parameters integrates to unity. Calculating the marginal likelihood therefore requires evaluating the integral over the joint prior and the likelihood function as denoted by (5.8), where  $\theta$  comprises all parameters in the model.<sup>58</sup> Two options are available, one can use the high dimensional set-up based on the augmented likelihood function or the non-augmented likelihood can be evaluated treating  $\alpha_i$  and  $\omega_i$  as latent variables. Since the method of Gelfand and Dey (1994) is known to perform poorly in high-dimensional problems, the non-augmented likelihood is evaluated (Balcombe, Chalak and Fraser, 2009). The method of Gelfand and Dey (1994) approximates the natural log of (5.8) by (5.9), where  $G$  is the set of maintained draws from the GS and  $\psi$  a truncated normal density function (for more details see Koop, 2003, pages 104-106). The truncation ensures that the fraction in (5.9) is bounded from above. The likelihood function is equivalent to (5.2) and is approximated by 1000 Halton draws (Halton, 1960).<sup>59</sup> Kass and Raftery (1995) provide a useful discussion on how Bayes Factors can be used for model selection.

$$(5.8) \quad p(y_M | X, Z, q) = \int_{\theta} p(\theta) L(y | \theta, X, Z, q) d\theta$$

$$(5.9) \quad \ln(p(y_M | X, Z, q)) = -\ln \left[ \frac{1}{G} \sum_{g=1}^G \frac{\psi(\theta_g)}{p(\theta_g) L(y | \theta_g, X, Z, q)} \right]$$

### 5.3.5 The dataset

In this chapter all three samples of the stated choice experiment, as described in Chapter 4, are applied to investigate the degree of heterogeneity in response patterns across respondents. The dataset comprises 701 respondents making eight choices each. Within each choice task a respondent is requested to pick his preferred alternative from two policy options and a status quo (SQ). By pooling the three alternative samples model efficiency is improved substantially. The model results presented in Section 5.4 were also estimated for each individual subsample, but this did not affect the general conclusions.

---

<sup>58</sup> For clarification, in the asymmetric triangular specification we evaluate the parameters  $a$ ,  $b$ ,  $c$ ,  $\beta_f$ ,  $\mu$  and  $\sigma^2$ .

<sup>59</sup> The stability of the likelihood function for a model was evaluated by taking (at each maintained draw of the GS) differences between the likelihood values at a given number of draws against the same model based on 2,500 Halton draws. Errors became sufficiently small around 1,000 draws.



#### 5.4 Results – Continuous distributions

Table 5.1 reports the results from three models, respectively using the normal, asymmetric triangular and lognormal mixing density. Given that all models are estimated in WTP-space, which requires that the cost (i.e. scale) coefficient is positive, the random parameter for cost is assumed to follow a lognormal distribution in each specification. Model 1 assumes that the marginal WTP parameters are normally distributed over the population of interest. Based on this specification, respondents are willing to pay, on average, €6.32 to increase the denominator of the flood probability by 1,000 years. That is, reducing the flood probability from for example, once every 4,000 years to once every 5,000 years. For an additional percentage point of compensation respondents are willing to pay €0.75. An extra hour of evacuation time is valued at €1.20. Payments are in euros (€) per household per year. For these non-cost attributes the standard deviations of the normal mixing density are larger than their mean values. Consequently, there is a high probability of respectively 29 percent, 22 percent and 40 percent that the respondent values an improvement in that attribute negatively. Although a share of the respondents may not feel a need for reductions in flood probability, improvements in compensation levels or available evacuation time, it is unlikely that they are willing to pay for deteriorations in these attributes.<sup>60</sup> Since the normal density has a symmetric shape, the substantial probability mass in the counterintuitive domain may be a result of the normal distribution attempting to take into account mass close to zero. It is around this point where we expect to find respondents that neglect a particular attribute in their decisions. Alternative inconsistencies with the shape of the normal density, like multimodality, may also be a cause of the estimated high standard deviations. As discussed in Campbell et al. (2010a), the tails of the normal distribution become more pronounced due to such inconsistencies between the true and modelled heterogeneity in response patterns.

---

<sup>60</sup> Since we are only offering improvements in flood risk exposure in the experimental design, respondents are not able to indicate whether they want to receive financial compensation for deteriorations in particular attributes.

**Table 5.1: Results attributes only models in WTP-space**

Name	Hyper - parameter	(1) Normal		(2) Triangular		(3) Lognormal	
		Post. Mean	Post. Std. Dev.	Post. Mean	Post. Std. Dev.	Post. Mean	Post. Std. Dev.
ASC	-	-2.99	0.13	-2.91	0.13	-2.04	0.12
PROB	Mean (mode)	0.63	0.07	-0.61	0.23	-1.64	0.22
	St. dev (lower bound)	1.13	0.08	-1.25	0.22	2.01	0.17
	(upper bound)	-	-	4.41	0.28	-	-
COMP	Mean (mode)	0.75	0.05	-0.32	0.16	-0.88	0.10
	St. dev (lower bound)	0.98	0.05	-0.77	0.16	1.59	0.10
	(upper bound)	-	-	3.65	0.18	-	-
EVAC	Mean (mode)	0.12	0.03	-0.35	0.11	-3.77	0.46
	St. dev (lower bound)	0.47	0.04	-0.63	0.10	2.71	0.32
	(upper bound)	-	-	1.52	0.12	-	-
COST (lognormal)	Mean	-1.16	0.06	-1.15	0.05	-1.16	0.06
	St. dev	1.01	0.06	0.92	0.12	1.05	0.06
<i>Model fit statistics</i>							
Log Marginal likelihood		-4332.53		-4328.23		-4230.02	
Log BF vs N		-		4.30		102.51	
Log BF vs T				-		98.21	
<i>Sample</i>							
Number of respondents		701		701		701	
Number of choice observations		5608		5608		5608	
<i>Marginal WTP densities evaluated at the posterior means</i>							
Attribute	(1) Normal	(2) Triangular	(3) Lognormal				
	Mean WTP <sup>+</sup>	Mean WTP <sup>+</sup>	Mean WTP <sup>+</sup>				
PROB	€6.32 (29%)	€8.50 (32%)	€14.80 (0%)				
COMP	€0.75 (22%)	€0.85 (24%)	€1.48 (0%)				
EVAC	€1.20 (40%)	€1.80 (43%)	€9.56 (0%)				

- a- WTP for increase by 1000 years in denominator of probability
- b- WTP for an additional percent point of compensation
- c- WTP for an additional hour of evacuation time
- + - Percentage of cdf below 0, i.e. with a negative marginal WTP between brackets

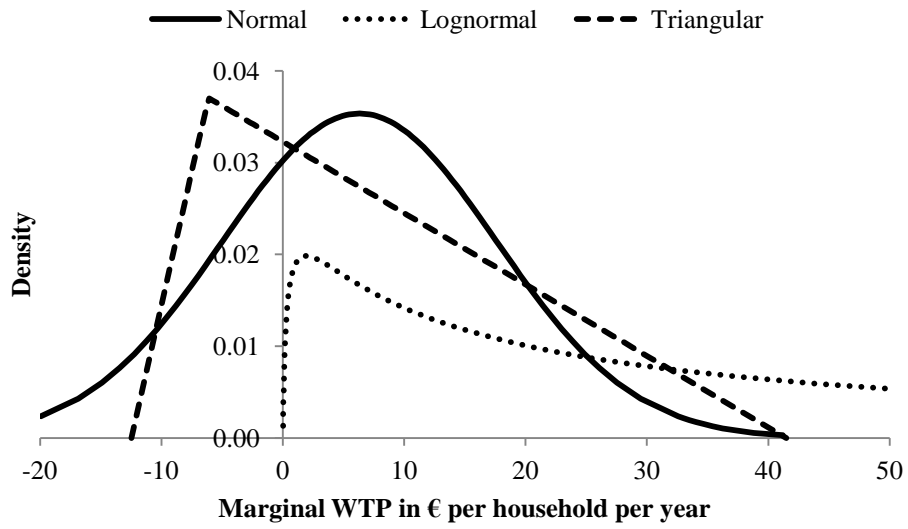
Model 2 comprises a set of asymmetric triangular distributions for the non-cost attributes. Table 5.1 reveals that marginal WTP for each non-cost attribute follows a right skewed distribution over the population of interest, i.e. the mean of the mixing density is larger than the median, creating a long right tail. However, all the estimates for the lower bounds and modes exhibit a negative sign and thereby fall within the counterintuitive domain from a behavioural perspective. Evaluated at the posterior mean of the GS, respectively 32

percent, 24 percent and 43 percent of the respondents reveals a negative marginal WTP for the probability, compensation and evacuation attribute.<sup>61</sup> Although the shares in the counterintuitive domain are higher than those for Model 1, Model 2 outperforms Model 1 in terms of model fit as indicated by the (log) Bayes Factor of 4.30. On the scale of Kass and Raftery (1995), there is ‘positive’ evidence in favour of Model 2. Figures 5.1-5.3 provide some explanation for the improvement in model fit. The lower tail of the normal distribution is much fatter than that for the triangular distribution and allows some respondents to have a very high negative marginal WTP for the non-cost attributes. The lower bound of the triangular distribution clearly corrects for these unexpected behavioural predictions of the normal distribution. The fat lower tail of the normal distribution is partially explained by the mode of the triangular distribution. Since the normal distribution is symmetric by definition, its standard deviation needs to correct for a skewed distribution of preferences, increasing the mass in the tails. In all three figures it becomes apparent that the upper bound on marginal WTP is estimated more accurately than its lower bound. That is, the mass in the upper tail of the normal distribution becomes negligible near the upper bound of the triangular distribution. This observation is also supported by the correlation matrix of parameter estimates. Correlation levels between the upper bounds and modes are approximately -0.25, indicating that the location of the upper bound can be defined in an independent fashion. On the contrary, there is a high degree of (negative) correlation between the draws for each lower bound and its respective mode (about -0.65). The latter is not surprising for choice experiments, since we are interested in identifying an individual’s maximum willingness-to-pay. Identifying lower bounds is more problematic, because in this part of the distribution we are more likely to find people that neglect a particular attribute, generating mass around zero, and those making inconsistent decisions.

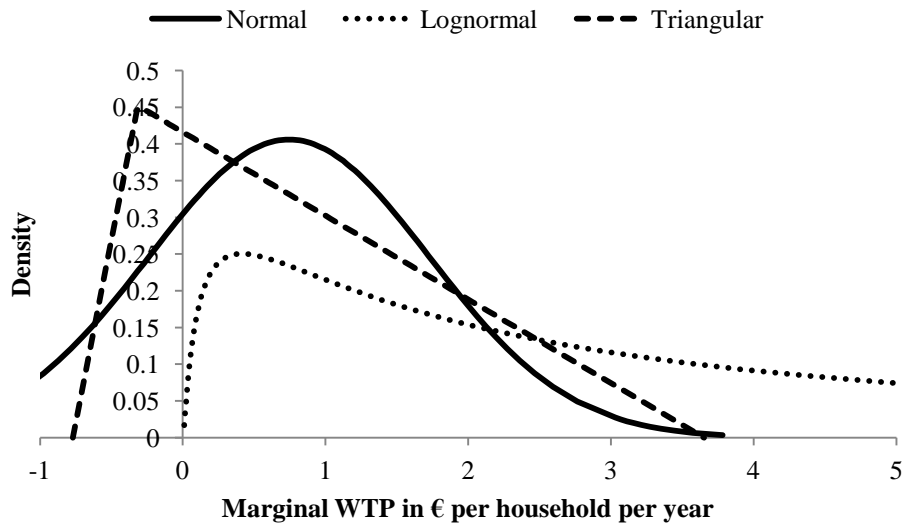
---

<sup>61</sup> Note that a negative lower bound is not necessarily inconsistent with the actual marginal WTP levels revealed by a respondent. As discussed in Rouwendal et al. (2010), while making trade-offs between multiple attribute in a single choice task, researchers only observe a single decision. Accordingly, the estimated set of MRSs becomes a function of the MRSs (or marginal WTPs) for other attributes in the choice experiment. Nevertheless, such a finding remains inconsistent from a behavioural perspective.

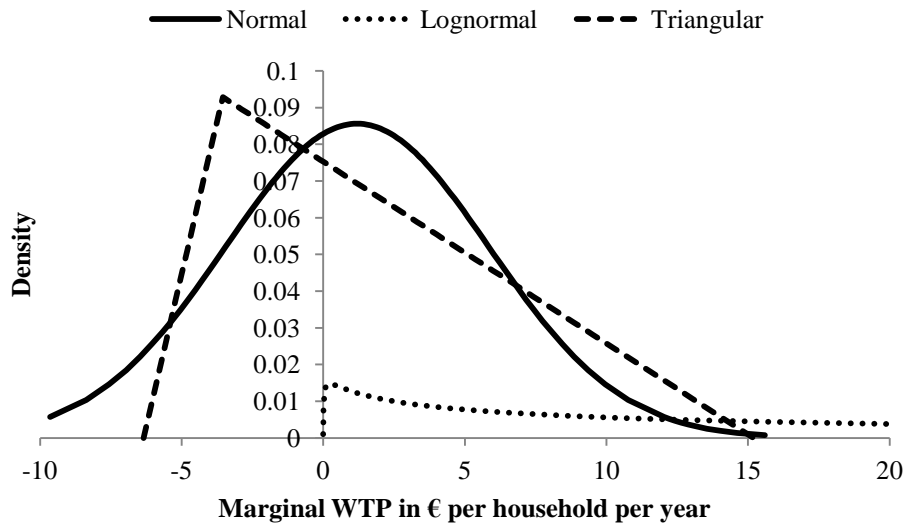
**Figure 5.1: Marginal WTP density for the probability attribute**



**Figure 5.2: Marginal WTP density for the compensation attribute**



**Figure 5.3: Marginal WTP density for the evacuation attribute**



The hyper-parameters of the (triangular) mixing density distribution attempt to correct for two elements. First, they try to match the shape of the mixing density with the shape of the true distribution of response patterns over the sample of interest. However, since the distributional form is not flexible enough, they attempt to match expected choice probabilities with observed choices as much as possible. This second correction may induce different moments on the mixing density than those existent in the true distribution of preferences. This inconsistency is driven by the limited flexible shape of most mixing densities, including the normal and triangular distribution. The most apparent consequence of the latter correction is that a substantial share of the population of interest tends to fall within a domain that is inconsistent from a behavioural perspective, i.e. exhibit a negative WTP for improvements in a policy attribute. Estimated mean marginal WTP levels for the triangular distribution support this notion. Mean marginal WTP levels increase respectively to €8.50, €0.85 and €1.80.

In order to restrict marginal WTP for the non-cost attributes to the positive domain, we estimate an alternative model using the lognormal distribution (Model 3) as the mixing density. Results for Model 3 are reported in Table 5.1. Remarkable are the high standard deviation estimates for the underlying normal distributions of the probability and evacuation attributes. Figures 5.1-5.3 show that by increasing the variance, the lognormal distribution becomes more *L*-shaped. The mode and most mass moves towards the origin, but some observations with a high marginal WTP in the right tail are still present. As such, Model 3 indicates that the probability and evacuation attributes have a lot of respondents with a low marginal WTP level or even close to zero. Nevertheless, the values in the right tail increase

mean WTP estimates to respectively €14.80, €1.48 and €9.56. In terms of model fit, Model 3 has a decisively better model fit compared to Models 1 and 2. It is therefore highly unlikely that a part of the sample has marginal WTP levels below zero. However, also the lognormal distribution appears to be inconsistent with the true distribution of preferences over the sample. Correlations over the draws in the GS between the hyper-parameters of the mixing density increase respectively to -0.79 and -0.90 for the probability and evacuation attribute. The mean of the underlying normal distribution attempts to correct for the impact of the high variance estimate on the mean and variance of the (transformed) marginal WTP distribution.<sup>62</sup> In a final attempt to approximate the WTP distributions for probability and evacuation using standard distributional forms, we estimate a model using a censored normal distribution. Correlation across the draws for the mean and variance of the evacuation attribute, however, increase to -0.97. Serious empirical identification issues are therefore evident within this model.

Overall, Figures 5.1-5.3 show that all three mixing densities have difficulties approximating the true mixing distribution of marginal WTP over the sample of interest. In particular on the right hand side of the mode all three mixing densities have a problem in offering flexibility. The triangular distribution has more mass at high marginal WTP estimates relative to the normal distribution, which is a direct consequence of the triangular and bell-shaped form of the two mixing densities. This effect is even more pronounced for the lognormal distribution, since the variance strongly affects the shape and moments of the mixing density. The set of tested models highlights that the chosen ‘simple’ mixing densities (i.e. the normal, lognormal, asymmetric triangular and censored normal) are likely to be misspecified and inconsistent with the true distribution of preferences over the sample. This is of particular concern when moving towards a more individual level of modelling. Remember, however, that in that case respondents are answering only to eight choice tasks, whilst trying to identify at least 4 parameters per respondent. This implies that the MMNL received a lot of identification from the mixing density itself and assuming that uncertainty regarding  $\alpha_i$  and  $\omega_i$ , as measured by the mixing density, is the same across respondents. Indeed, an attempt to estimate a model with an (improper) uniform prior on the individual level parameter estimates revealed this lack identification and showed high levels of confounding. From that perspective, the mixing density is highly influential in determining the shape of the individual

---

<sup>62</sup> The mean of the lognormal distribution is defined by  $\exp(\mu + \sigma^2/2)$ . We also accommodated the domain restrictions on marginal WTP in our triangular model by fixing the lower bounds at zero. Similar to Model 3, the mode approaches zero for the probability and evacuation attributes. Results can, however, not be reported due to convergence problems during the estimation.

specific posterior distribution on  $\alpha_i$  and  $\omega_i$ . Policy makers are, however, more interested in the moments of the mixing density. The sensitivity of mean marginal WTP estimates to the mixing density reveals the potential of severe biases in welfare estimates as a result of alternative mixing densities. Sensitivity tests are therefore recommended. Nevertheless, the improved model fit found by the Bayes Factors of the triangular and lognormal distribution relative to the normal mixing density, indicates that response patterns are asymmetrically distributed over the population of interest. The superior model fit of the lognormal indicates that respondents value the non-price attributes positively, but some respondents may have a tendency to neglect a particular attribute and reveal a (near) zero marginal WTP.

### 5.5 A Latent Class approach

In order to obtain additional insights into the distribution of marginal WTP over the sample, the use of latent class (LC) models is proposed. To keep the model specification as simple as possible, we abstain from mixture models such as applied by Burda et al. (2008) and Fosgerau and Hess (2009) and related independent discrete distributions. The LC specification presented here is a standard linear utility function in preference space, since it is possible to take the ratio of coefficients at specific mass-points rather than dividing entire distributions. Identifying the moments of WTP is therefore less problematic (Daly, Hess and Train, 2011b).

Estimations are again conducted in a Bayesian framework to facilitate non-nested model comparison. Train (2008) provides an insightful discussion on estimating a LC model with a large number of classes in the classical framework. The likelihood function presented in (5.10) describes a panel set-up over  $S$  classes, where  $\pi_s$  describes the probability of individual  $i$  belonging to class  $s$ . A set of class specific parameter  $\beta_s$  is estimated for each class and is summarized in the vector  $\beta$ .<sup>63</sup> Within the LC model, class membership is treated as an unobserved variable, like the individual specific parameters  $\alpha_i$  and  $\omega_i$  in the continuous MMNL model. In analogy to the continuous MMNL model, the vector of class-probabilities  $\pi$  serves as the mixing density and is replaced by the augmented class membership vector  $e_i$  in the augmented likelihood function to facilitate model estimation (Koop, 2003, Ch. 10). Denote augmented class membership by the binary variable  $e_{is}$ , which takes the value one if an individual belongs to a specific class and zero otherwise. Since a respondent can only

belong to a single class the restriction  $\sum_{s=1}^S e_{is} = 1$  is imposed. By definition,  $e_i$  follows a

---

<sup>63</sup> Here, also the ASC is class specific, but a set of parameters constant across classes can be easily added through  $\beta_f$  as an additional block in the GS.

multinomial distribution  $e_i \sim M(1, \pi)$ . Again, prior distributions are assigned to the model parameters, i.e.  $\pi$  and  $\beta_s$ . The class specific parameters are assigned a relatively uninformative, normal prior with mean  $m_s$  and variance  $V_s$ .<sup>64</sup> The vector  $\pi$  is assigned a conjugate Dirichlet prior with the  $S$  by 1 vector  $\varphi$  as its hyper-parameter (Koop, 2003). Since we have no information about class probabilities a priori, each element in  $\varphi$  is assigned the uniform value  $1/S$ . Similar to the continuous MMNL model, a set of conditional posteriors is obtained by combining the priors and the augmented likelihood function. The posteriors are described in more detail in Appendix 5.C.

$$(5.10) \quad L(y | X, \beta) = \prod_{i=1}^n \sum_{s=1}^S \pi_s \prod_{t=1}^T P_{ijt|\beta_s}$$

To ensure model identification, an ordering restriction is imposed in the GS on the class probabilities, such that  $\pi_s < \pi_{s+1}$ . The latter is required to prevent label-switching, since the likelihood value is the same under alternative orderings of classes in the LC model (Koop, 2003). Draws from the posterior on  $\pi$  that do not satisfy this constraint are not considered. The GS consists of 30,000 burn-in draws and a set of 50,000 maintained draws. Convergence and acceptance rates of the draws in the GS are evaluated in similar vein to the continuous MMNL model and thinning is applied to store 10,000 draws in total for each model. Finally, the method of Gelfand and Dey (1994) is again used for model comparison.

## 5.6 Results – Latent Class

Table 5.2 provides an overview of the marginal likelihoods obtained for a range of attributes-only models varying in the number of latent classes taken into account. Based on model fit, the optimal number of classes is defined at six, which provides a worse model fit than the lognormal specification in Table 5.1.<sup>65</sup> This has two reasons. First, the number of parameters has increased substantially from nine to 35, for which the model is penalized by the set of uninformative priors. Second, by having only six mass-points the true degree of heterogeneity is likely to be understated in the model. Nevertheless, when including more than three classes, the model performs at least as good as or better than the normal and triangular specifications in Table 5.1. The flexible shape of the mixing density arising from the LC model thus has a

<sup>64</sup> Its prior values are equivalent to those assigned to  $m_f$  and  $V_f$  in Appendix 5.A.

<sup>65</sup> The marginal likelihood only provides ‘positive’ support for the model with seven latent classes relative to the model with six classes in terms of the Kass and Raftery (1995) scale. Selecting the latter model is therefore somewhat arbitrary, but justifiable. We prefer the model with six classes, since it comprises less parameters.



clear benefit over the latter mixing densities. Despite its flexible shape, correlation patterns across the draws of the GS do not indicate serious identification issues in the LC-model. Only when the number of classes increases, the correlation across preference parameters within a specific class increases. The latter is not surprising, because data augmentation splits the sample into various smaller classes. Consequently, the effect of specific attributes on utility cannot be identified as independently from other attributes as when the full sample is considered. For example, for the LC model with six classes this implies that the highest correlation of 0.81 is observed between the ASC and the cost coefficient in the third class (S3). Respondents in this class are either highly cost sensitive or have a strong tendency to select the status quo.<sup>66</sup>

**Table 5.2: Marginal likelihoods of attributes only Latent Class models**

# of classes	Marginal likelihood	Bayes Factor (relative to previous class)	# of parameters
2	-4518.74	-	11
3	-4400.02	118.73	17
4	-4332.43	67.59	23
5	-4314.45	17.98	29
6	-4297.97	16.48	35
7	-4292.77	5.20	41
8	-4292.53	0.24	47

Marginal WTP estimates for the non-cost attributes in the model with six latent classes can only be obtained for four out of six classes as revealed by Table 5.3. In the smallest class (S1) and S4 the cost coefficient does not vary from zero. Respectively, 25% of the respondents fall in these two classes. Neglecting the cost attribute implies that the moments of the attribute specific marginal WTP distributions are not defined for the entire model. The shares of GS draws in the counterintuitive domain, as reported in Table 5.3, also highlight that the evacuation attribute is particularly neglected in the second class (S2). Similar observations can be made for the probability and compensation attribute in the smallest class (S1). Moreover, when interpreted in classical terms at the 5% significance level, in most classes at least one of the non-cost policy attributes has an insignificant impact on utility. Only in class six trade-offs are made between all policy alternatives.

<sup>66</sup> For this group of people it could have been the case that their maximum annual WTP for most policy packages falls below €40 per year. Indeed, this class is represented by a high cost parameter, i.e. low marginal WTP parameters.

**Table 5.3: Class probabilities and % of draws in the GS falling in the unexpected domain for LC 6**

Class	Class Probability	Preference parameters (% of draws in negative domain)				Marginal WTP parameters <sup>+</sup> (% of draws in negative domain)		
		PROB	COMP	EVAC	COST*	PROB	COMP	EVAC
S1	6%	13%	13%	0%	42%	39%	40%	41%
S2	13%	8%	3%	29%	0%	8%	3%	29%
S3	16%	3%	0%	6%	0%	3%	0%	6%
S4	19%	0%	7%	6%	23%	23%	26%	26%
S5	21%	0%	0%	7%	0%	0%	0%	7%
S6	25%	2%	0%	0%	0%	2%	0%	0%

\*% of draws in positive domain

<sup>+</sup> based on ratio at each draw of the GS

Table 5.4 refutes the results from the continuous MMNL models presented before based on a set of normal and triangular mixing densities. Class specific preference parameter estimates of the LC-model all fall in the expected domain. Moreover, the low (or insignificant in classical estimation terms) cost coefficient in classes one and four explain in part the good fit of the lognormal distribution. Mean marginal WTP estimates will be high for the respondents in these classes and approximate infinity. Class three, and to a lesser extent class six, however, also underline the limitations of the lognormal model. In S2 and S5 respondents are highly sensitive to the cost coefficient inducing marginal WTP estimates to be close to zero. Moreover, since the lognormal distribution cannot take into account mass at zero (e.g. Rigby and Burton, 2006) it has a hard time explaining the insignificant marginal WTP estimates for the probability, compensation and evacuation attributes across the various classes. The normal and triangular distributions have difficulties assigning mass to in the upper tail as this would also assign significant mass to intermediate values and the lower tail. The lognormal distribution is more suited in this case, in particular also by the infinite marginal WTP for classes one and four. For the evacuation attribute most mass points will be centred close to zero, due to the insignificant marginal WTP estimates in most classes. The lognormal simultaneously tries to take into account classes one and four where marginal WTP approaches infinity. Clearly, this stimulates the L-shaped form of the lognormal distribution, but this is not likely to be entirely consistent with the true shape of the mixing density.

Overall, the LC model with six classes reveals that not each respondent takes all attributes into account, including the cost attribute, and that when respondents take an attribute into account, they comply with a priori behavioural expectations. This results in some mass-points of marginal WTP close to zero and some approaching infinity. The lognormal distribution is best capable to approximate these forms of behaviour, but its

distributional shape remains inconsistent with the true distribution of preferences over the sample of interest. An additional drawback of the lognormal distribution is that it is not able to take into account attribute non-attendance (i.e. zero marginal WTP).

**Table 5.4: Estimated preference parameters for the policy attributes in LC6 (ASC not reported)**

Class	Class Probability		PROB		COMP		EVAC		COST	
	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.
S1	0.06	0.01	0.09	0.09	0.06	0.06	0.51	0.15	-0.01	0.05
S2	0.13	0.02	0.08	0.06	0.06	0.03	0.01	0.03	-0.24	0.03
S3	0.16	0.01	0.09	0.05	0.16	0.03	0.03	0.02	-0.62	0.06
S4	0.19	0.01	0.33	0.03	0.02	0.02	0.02	0.01	-0.01	0.02
S5	0.21	0.01	0.20	0.05	0.53	0.05	0.03	0.02	-0.18	0.03
S6	0.25	0.02	0.06	0.03	0.19	0.03	0.04	0.01	-0.34	0.03

## 5.5 Summary

In this chapter the focus was on the impact of the mixing density in random parameter logit models on marginal WTP estimates. Random parameters are commonly used to approximate the variation in response patterns across respondents. Specification of the mixing density introduces a delicate decision between behavioural relevance, model fit and ease of estimation. Regarding the former, this chapter has shown that simple unconstrained distributions, like the normal distribution and the (new) asymmetric triangular distribution, have problems in approximating the revealed distribution of response patterns over the sample of interest. According to these distributions, a substantial share of respondents has a marginal WTP which falls in the domain that is not a priori predicted by economic theory. More detailed analysis showed that the shape of these distributions is inconsistent with the actual heterogeneity in marginal WTP across respondents. For example, the lognormal distribution restricts the domain of the mixing density to the relevant domain from a behavioural perspective and thereby induces a large improvement in model fit. However, the shape of the latter distribution also turned out to be inconsistent with the revealed distribution of preferences over the sample, resulting in empirical identification issues. The prime cause of this inconsistency is that respondents neglected particular attributes when going through the choice experiment. The lognormal distribution is not able to capture the resulting mass-points at zero. It gets even more problematic when respondents neglected the cost attribute. In that case marginal WTP estimates approach infinity. The fat upper tail of the lognormal

distribution is more suitable to capture these extreme choice patterns compared to the normal and asymmetric triangular distribution.

The empirical identification issues related to the lognormal distributions call for a different approach. More flexible continuous distributions, however, put heavy requirements on the informational content of the data and are likely to introduce additional empirical identification problems. This chapter has shown that the mixing density receives most identification by assuming that uncertainty about individual level parameters (i.e. the shape of the mixing density) is the same across respondents. Marginal WTP at the individual level can, however, be barely identified illustrating the limited informational content of choice experiments for welfare analysis at the individual level. Individual specific posterior estimates are highly influenced by the shape of the mixing density. Therefore, additional research is needed regarding the applicability of more flexible continuous distributions to stated choice experiments. Latent class models offer a relatively simple, but flexible modelling approach. By using discrete mass-points, variation in preferences can be identified without relying on a particular distributional shape. Despite running the risk of understating heterogeneity in preferences, the set of latent class models indeed revealed that respondents tended to neglect particular attributes in their decision strategies. Even here, the requirements on the data become too high when a large number of classes is specified. Accordingly, the researcher needs to decide on which criteria to select the mixing density. Despite their worse model fit, relative to the lognormal model, latent class models seem to be better suited to reveal decision patterns which are more consistent with actual behaviour underlying this study.

Finally, marginal WTP estimates vary substantially across specifications of the mixing density. Hence, the model selection criterion will also have an impact on the welfare implications of specific policies. Researchers should be aware of this sensitivity and explore alternative specifications to find out whether simple specifications of the mixing density, like the normal or asymmetric triangular, confirm with observed behaviour. Of particular concern is that marginal WTP estimates approach infinity if respondents neglect the cost attribute in their decisions. Accounting for such behaviour in welfare analysis should be studied in more detail (e.g. Hensher and Greene, 2010).

## Appendix 5.A – Applied prior values

The prior values applied during the estimation are reported in Table 5.A1. In order for the priors to be uninformative on the location of the parameters of interest during the actual estimation, we have set a high variance for the normal priors. Here, the uniform prior on the hyper-parameters of the triangular distribution only restricts the domain of the posterior, but provides no information on the exact location of the parameter.

**Table 5.A1: Applied prior values during the estimation**

Prior Parameter	Description	Affects parameter	Size	Value
$m_f$	Prior mean for the fixed parameters in the choice model	$\beta_f$	$k_f$ by 1	0
$V_f$	Prior variance for the fixed parameters in the choice model	$\beta_f$	$k_f$ by $k_f$	$100 \cdot I_{k_f}$
$m_r$	Prior mean for the mean of a normally distributed mixing density	$\mu$	scalar	0
$V_r$	Prior variance for the mean of a normally distributed mixing density	$\mu$	scalar	100
$H$	Prior shape parameter for the inverse gamma prior density on $\sigma^2$	$\sigma^2$	scalar	0.5
$K$	Prior scale parameter for the inverse gamma prior density on $\sigma^2$	$\sigma^2$	scalar	0.5
$l_a$	Lower bound on the uniform prior for the lower bound of the triangular distribution	$a$	scalar	-100
$u_a$	Upper bound on the uniform prior for the hyper-parameters of the triangular distribution	$a, b$ and $c$	scalar	100

## Appendix 5.B – The conditional posteriors in the continuous MMNL model

This appendix describes the full set of conditional posteriors in which the mixing density for the marginal WTP parameters  $\omega_i$  is a triangular distribution and  $\alpha_i$  is described by a lognormal density. Hence, the probability of picking the observed choice is:

$$P_{ijt} = \frac{\exp\left(Z_{ijt}\beta_f + e^{\alpha_i}(X_{ijt}\omega_i - q_{ijt})\right)}{\sum_{k \in D_{it}} \exp\left(Z_{ikt}\beta_f + e^{\alpha_i}(X_{ikt}\omega_i - q_{ikt})\right)}$$

The fixed parameters\*:

$$p(\beta_f | m_f, V_f, \alpha, \omega, y, X, Z, q) \propto \exp\left(-\frac{1}{2}(\beta_f - m_f)'V_f^{-1}(\beta_f - m_f)\right) \prod_{i=1}^n \prod_{t=1}^T P_{ijt}$$

The augmented parameters\*:

$$p(\alpha_i | \mu, \sigma^2, \beta_f, \omega_i, y, X, Z, q) \propto \exp\left(-\frac{1}{2}(\alpha_i - \mu)'(\sigma^2)^{-1}(\alpha_i - \mu)\right) \prod_{t=1}^T P_{ijt}$$

$$p(\omega_i | a, b, c, \alpha_i, y, X, Z, q) \propto \prod_{t=1}^T P_{ijt} \cdot \begin{cases} \frac{(\omega_i - a)}{(c - a)} & \text{for } a \leq \omega_i \leq c \\ \frac{(b - \omega_i)}{(b - c)} & \text{for } c < \omega_i \leq b \\ 0 & \text{otherwise} \end{cases}$$

The hyper-parameters of the lognormal mixing density:

- Where  $\bar{\mu}$  represents the (normal) posterior density for hyper-parameter  $\mu$

$$p(\mu | m_r, V_r, \sigma^2, \alpha) \propto \exp\left(-\frac{1}{2}(\mu - m_r)' V_r^{-1} (\mu - m_r)\right) \prod_{i=1}^n \exp\left(-\frac{1}{2}(\alpha_i - \mu)' (\sigma^2)^{-1} (\alpha_i - \mu)\right)$$

$$\bar{\mu} \sim n(\mu_1, V_1)$$

$$V_1 = \left(\frac{1}{V_r} + \frac{n}{\sigma^2}\right)^{-1}$$

$$\mu_1 = V_1 \left(\frac{m_r}{V_r} + \frac{1}{\sigma^2} \sum_{i=1}^n \alpha_i\right)$$

- $ig$  stands for the inverse gamma distribution.

$$p(\sigma^2 | \eta, \kappa, \mu, \alpha) \propto (\sigma^2)^{-(\eta+1)} \exp(-\kappa/\sigma^2) \prod_{i=1}^n \exp\left(-\frac{1}{2}(\alpha_i - \mu)' (\sigma^2)^{-1} (\alpha_i - \mu)\right)$$

$$\bar{\sigma}^2 \sim ig(\eta_1, \nu_1)$$

$$\eta_1 = \eta + \frac{n}{2}$$

$$\kappa_1 = \kappa + \frac{\sum_{i=1}^n (\alpha_i - \mu)^2}{2}$$

The hyper-parameters of the triangular mixing density\*:

$$p(a | b, c, \omega) \propto \left(\frac{1}{u_a - a}\right) \prod_{i=1}^n \begin{cases} \frac{(\omega_i - a)}{(b-a)(c-a)} & \text{for } a \leq \omega_i \leq c \\ \frac{(b - \omega_i)}{(b-a)(b-c)} & \text{for } c < \omega_i \leq b \end{cases} \quad \text{for } l_a \leq a \leq c$$

$$p(c | a, b, \omega) \propto \left(\frac{1}{u_a - c}\right) \prod_{i=1}^n \begin{cases} \frac{(\omega_i - a)}{(c-a)} & \text{for } a \leq \omega_i \leq c \\ \frac{(b - \omega_i)}{(b-c)} & \text{for } c < \omega_i \leq b \end{cases} \quad \text{for } a \leq c \leq b$$

$$p(b | a, c, \omega) \propto \prod_{i=1}^n \begin{cases} \frac{(\omega_i - a)}{(b-a)(c-a)} & \text{for } a \leq \omega_i \leq c \\ \frac{(b - \omega_i)}{(b-a)(b-c)} & \text{for } c < \omega_i \leq b \end{cases} \quad \text{for } c \leq b \leq u_a$$

Model estimations were conducted in MATLAB and random walk chains are applied for the M-H steps in the Gibbs Sampler (e.g. Chib and Greenberg, 1995). Conditional posteriors assigned with a \* require the use of a M-H algorithm. As the candidate generating density function (cgdf) we use normal distributions with the mean set at the old draw and rejecting draws that fall outside of the support defined by the prior, for example when  $a > c$ . The latter method is also supported by Gelfand et al. (1992) and turned out to work better than using truncated normal distributions as the cgdf. While running the GS, acceptance rates and the share of improper draws are monitored. Variance terms of the cgdf are adjusted accordingly to improve the efficiency of the model. Acceptance rates were aimed at levels between 20 and 35 percent.



### Appendix 5.C – The conditional posteriors in the Latent Class MMNL model

This appendix describes the full set of conditional posteriors for a MMNL model in which the mixing density is in the form of a latent class model in preference space. Hence, the class specific probability of picking the observed choice is described by:

$$P_{ijt} = \frac{\exp\left(Z_{ijt}\beta_f + (X_{ijt}\beta_s + q_{ijt}\alpha_s)\right)}{\sum_{j \in D_{it}} \exp\left(Z_{ikt}\beta_f + (X_{ikt}\beta_s + q_{ikt}\alpha_s)\right)}$$

In the remaining of this Appendix  $\alpha_s$  is dropped to save space. Due to working in preference space, its conditional posterior is comparable to the posterior on  $\beta_s$ .

The fixed parameters\*:

$$p\left(\beta_f \mid \mu_f, V_f, \beta, e, y, X, Z, q\right) \propto \exp\left(-\frac{1}{2}(\beta_f - m_f)'V_f^{-1}(\beta_f - m_f)\right) \prod_{i=1}^n \sum_{s=1}^S e_{is} \prod_{t=1}^T P_{ijt|\beta_s, \beta_f}$$

The fixed parameters are not included in the presented attributes only model. Moreover, an M-H algorithm is required to take draws from the conditional posterior.

The augmented class membership parameter  $e_i$ :

- $M$  stands for the multinomial distribution.

$$\left(e_i \mid \beta, \beta_f, q, y, X, Z\right) \sim M \left( 1, \left[ \frac{q_1 \prod_{t=1}^T P_{ijt|\beta_1, \beta_f}}{\sum_{s=1}^S q_s \prod_{t=1}^T P_{ijt|\beta_s, \beta_f}}, \frac{q_2 \prod_{t=1}^T P_{ijt|\beta_2, \beta_f}}{\sum_{s=1}^S q_s \prod_{t=1}^T P_{ijt|\beta_s, \beta_f}}, \dots, \frac{q_S \prod_{t=1}^T P_{ijt|\beta_S, \beta_f}}{\sum_{s=1}^S q_s \prod_{t=1}^T P_{ijt|\beta_s, \beta_f}} \right] \right)$$

The class probabilities  $q$ :

- $D$  stands for the Dirichlet distribution and  $e_i$  is a row vector of augmented class membership for individual  $i$ .

$$(q | \beta, \beta_f, e, y, X, Z) \sim D(\bar{\varphi})$$

$$\bar{\varphi} = \varphi + \sum_{i=1}^n e_i$$

The class specific marginal utility parameters  $\beta_s^*$ :

$$p(\beta_s | \mu_s, V_s, \beta, e, y, X, Z, q) \propto \exp\left(-\frac{1}{2}(\beta_s - m_s)' V_s^{-1} (\beta_s - m_s)\right) \prod_{i=1}^n \sum_{s=1}^S e_{is} \prod_{t=1}^T P_{ijt|\beta_s, \beta_f}$$

The latter part boils down to evaluating the likelihood function for all respondents assigned to class  $s$ . An M-H algorithm is required to draw from the conditional posterior.



## Chapter 6: Semi-parametric estimation of preference dynamics in stated choice experiments<sup>67</sup>

This chapter addresses research question two by looking into the dynamics of marginal WTP estimates over the choice sequence. More specifically, an alternative method to the Swait and Louviere (1993) test procedure to test for preference dynamics over the choice sequence in stated choice experiments is proposed. The semi-parametric local multinomial (L-MNL) model allows estimating choice task specific preference parameters and related welfare measures more efficiently. The proposed method boils down to estimating a sequence of weighted MNL models, where the weights define the degree of similarity between observations. By controlling the degree of smoothing of choice task specific parameter estimates, the L-MNL model offers a flexible model to account for gradual preference dynamics in stated choice experiments. In the empirical application, both the L-MNL model and the Swait and Louviere (1993) test procedure are used to test the contrasting hypotheses regarding preference uncertainty and related preference dynamics formed by the discovered preference hypothesis and theory of coherent arbitrariness, as discussed in Section 3.6.1. The set-up of the chapter is as follows. Section 6.1 provides a short recap of the underlying theoretical framework and findings in the literature. Section 6.2 discusses our empirical approach to contrasting the two alternative theories. The econometric methods are explained in Section 6.3. The data are briefly described in Section 6.4. Section 6.5 provides the main results and Section 6.6 concludes.

### **6.1 Preference dynamics in stated choice experiments**

#### **6.1.1. Recap of the theoretical framework**

The random utility model (RUM) assumes respondents are rational and have well defined preferences, i.e. complete and transitive utility functions. Hence, when asked to choose between different alternatives in a stated choice experiment, respondents can identify in each choice task their preferred alternative and they know at which rate they are willing to give up one (policy) attribute for another. Non-market valuation studies are, however, frequently applied to (new) goods and services with which the respondent has only limited experience. The existence of well-defined preferences has therefore been questioned (e.g. Bateman et al., 2008; Brouwer et al., 2010; Shaikh, Sun and Cornelis van Kooten, 2007).

---

<sup>67</sup> This chapter is based on joint work with Paul Koster ([p.r.koster@vu.nl](mailto:p.r.koster@vu.nl)) and under review at the Journal of Public Economics.

Two contrasting hypotheses exist regarding the extent to which preferences are subject to change as respondents gain experience with the non-market good, the (hypothetical) market and its institutional settings. Plott (1996) argues in his discovered preference hypothesis (DPH) that by repetitively operating on a market, individuals discover the best way to act and learn about their preferences through experience. Braga and Starmer (2005) refer to these processes as institutional and value learning. Learning opportunities are frequently acknowledged to result in reductions in anomalous behaviour, such as the WTA-WTP disparity (List, 2003; Loomes, Starmer and Sugden, 2010). Plott's hypothesis that preferences evolve over a sequence of choices is congruent with the preference construction literature (e.g. Ariely, Loewenstein and Prelec, 2003; Ariely et al., 2006; Fischhoff et al., 1999; Lichtenstein and Slovic, 2006; McFadden, 1999; Slovic, 1995). For example, Ariely et al. (2003) stipulate in their theory of coherent arbitrariness (CA) that an individual gradually develops a stable set of preferences due to an internal drive for consistency with past decisions. Since past decisions drive future decisions, the convergence process is path dependent and possibly influenced by initial (arbitrary) value cues. As such, the preference construction literature argues that a stable set of preferences is non-existent prior to the choice experiment.<sup>68</sup> In contrast, the DPH predicts convergence to a stable set of preferences, but assumes that the convergence level is path independent (Bateman et al., 2008; Braga and Starmer, 2005). Closely related is the discussion about the existence and decay of a starting point bias, which has been extensively discussed in the contingent valuation literature and more recently also in the stated choice experiment literature (Carlsson and Martinsson, 2008; Groeneveld, 2010; Ladenburg and Olsen, 2008).

Inducing a starting point bias is one approach to contrast the competing hypotheses and is adopted here.<sup>69</sup> In similar vein to Ladenburg and Olsen (2008), we induce a starting point bias and test for dynamics in welfare estimates in a choice experiment on flood risk reduction in the Netherlands. The chapter offers two main contributions to the literature. First, an improved experimental design is applied enabling us to better identify dynamics in welfare measures over the choice sequence (see Chapter 4). Second, we present a novel econometric approach, called the local multinomial logit (L-MNL) model, as an alternative to the

---

<sup>68</sup> Even though the DPH assumes stable preferences exist before respondents come to the (hypothetical) market, both the DPH and theory of CA assume preferences are not (yet) fully known to *the respondent* at that stage, which induces choice uncertainty.

<sup>69</sup> The presence of a starting point bias is interpreted here as if a stable set of preferences is not known to the respondent at the start of the survey. A decay of the starting point bias supports the DPH, while a persistent starting point bias supports the theory of CA. Support for either hypothesis remains conditional on a stable set of preferences by the end of the choice sequence. Alternative tests contrasting both hypotheses may, for example, alter the order of the choice sequence, thereby inducing path dependence.

commonly applied Swait and Louviere (1993) test procedure (e.g. Bech, Kjaer and Lauridsen, 2011; Brouwer et al., 2010; Carlsson et al., 2010; Holmes and Boyle, 2005; Ladenburg and Olsen, 2008). The L-MNL model provides an intermediate solution to either fully combining or entirely splitting samples of interest. The latter is required by the Swait and Louviere test. Moreover, the L-MNL model is expected to significantly reduce the size of standard errors compared to estimating choice task specific welfare measures (Koster and Koster, 2011).

### **6.1.2 Findings in the literature**

The HSB and LSB samples included in our experimental design were specifically designed to contrast the DPH and theory of CA by inducing a starting point bias (SPB) and testing for dynamics in a set of (scale-free) marginal WTP estimates of interest over the choice sequence. Bateman et al. (2008) and Ladenburg and Olsen (2008) find support for the DPH by observing a rapid decay in the SPB in respectively a repeated dichotomous choice contingent valuation study and a stated choice experiment. Persistent contextual framing effects are also not observed by Carlsson and Martinsson (2008). Groeneveld (2010), however, does not observe decay in SPB in a stated choice experiment focusing on reducing damages to specific ecosystems. To our knowledge, these studies cover the limited set of available studies directly testing the hypotheses embodied in the DPH and CA models. Related studies show that marginal WTP estimates are sensitive to the maximum level of the cost attribute (Morkbak et al., 2010), the number of choice cards included in the experiment influences preferences (Bech, Kjaer and Lauridsen, 2011), and dynamics in preferences are observed over the choice sequence (Carlsson, Morkbak and Olsen, 2010; Holmes and Boyle, 2005; Swait and Adamowicz, 2001b). However, these findings are not supported by Brouwer et al. (2010) and Savage and Waldman (2008). Brown et al. (2008) and Kingsley and Brown (2010) find evidence of preference learning by reductions in decisional variance over a sequence of choices. The reverse, i.e. indications of fatigue effects, is also observed in the literature (Arentze et al., 2003; Bradley and Daly, 1994; Caussade et al., 2005; Lundhede et al., 2009). Finally, Day and Pinto (2010) find that the order in which choice tasks are presented may also affect welfare estimates.

## **6.2 Empirical approach**

### **6.2.1 Within and between sample preference dynamics**

Similar to Ladenburg and Olsen (2008), we induce a starting point bias between two independent samples through the experimental design. The only difference between the two

samples, referred to as the Low Starting Bid (LSB) and High Starting Bid (HSB) sample respectively, is found in the first choice task.<sup>70</sup> The policies presented in the first choice task were nearly identical across samples, but only varied in the cost of the proposed policies. LSB was assigned the lowest levels of the price vector and HSB the highest levels for exactly the same policy alternatives.<sup>71</sup> In the following eight choice tasks all respondents were presented with exactly the same experimental design and the same attribute levels for the subsequent alternatives, including the levels of the cost coefficient. Hence, this chapter only takes into account starting point effects, but not the effect of showing different attribute levels to different respondents (e.g. Carlsson and Martinsson, 2008; Hanley et al., 2005; Morkbak, Christensen and Gyrd-Hansen, 2010; Ohler et al., 2000).<sup>72</sup>

By anchoring respondents on the price of the presented policy alternatives in the first choice task, respondents in the LSB sample may be more sensitive to changes in the cost attribute during the rest of the choice sequence. They have a lower reference value induced by the experimental design. Therefore, marginal WTP values are potentially higher in the HSB sample. Such an effect reflects the presence of a potential SPB. As respondents proceed through the remaining eight choice tasks, they encounter several attribute levels and learn about their preferences. Consequently, the impact of the initial choice task on their subsequent choices is likely to decay. Based on the DPH and due to the similarity in design between HSB and LSB, marginal WTP estimates can be expected to stabilize and converge between both samples. However, in Section 4.4.2 two hypotheses were already formulated taking the micro-economic framework as a point of reference. *First, preference parameters are expected to be constant over the choice sequence and are not affected by arbitrary framing effects, such as the use of initial value clues. Second, scale parameters are constant over the choice sequence and not affected by any possible learning effects.* In testing these hypotheses in the remaining of this chapter, we aim to identify two forms of preference dynamics over the choice sequence. First, we test for *within-sample* preference dynamics, where both the DPH and theory of CA predict the emergence of a stable set of preferences due to learning. Second, we test for *between-sample* preference dynamics. The DPH predicts convergence in preferences between samples, whereas the theory of CA predicts a set of stabilizing preferences subject to a persistent SPB throughout the choice sequence.

---

<sup>70</sup> The words choice task and choice card will be used in the remainder of this chapter. The former refers to the position of the choice in the choice sequence. The latter refers to a specific choice situation as included in the experimental design. Respondents were presented with two alternative policies and an opt-out option.

<sup>71</sup> Within each sample everybody answered exactly the same first choice task.

<sup>72</sup> A common finding in the contingent valuation literature is that using systematically higher bid levels results in higher WTP values (Chien et al., 2005).

## 6.2.2 Improving the experimental design

Testing for *within* and *between* sample preference dynamics requires the estimation of choice task specific preference parameters within the LSB and HSB samples.<sup>73</sup> Estimation of a choice model for a specific choice task in each sample requires that all choice cards in the experimental design are answered a sufficient number of times at each moment during the choice sequence. If this is not the case, then the parameter estimates are highly inefficient, because only a limited number of trade-offs are considered. Identifying whether preference dynamics are result of ‘true’ preference dynamics, limitations of the design, or heterogeneity in preferences across respondents becomes hard under these circumstances. This may have played a role in Ladenburg and Olsen (2008), where each respondent was presented with the same choice task at the same moment in the choice sequence. We work around this issue by applying a rotating procedure, which structurally varies the order in which the choice cards are presented to the respondent. More details about the experimental design and rotating procedure can be found in Chapter 4.

Due to our careful experimental set-up, we avoid design elements that could have an impact on differences in preferences between the two samples, except differences in the choice task at the start of the stated choice experiment and (unobserved) differences in respondent characteristics. To minimize the impact of variations in respondent characteristics on choice task specific parameter estimates, we sampled respondents for both versions independently. The hired survey company launched the LSB and HSB separately and demographic and socio-economic characteristics were monitored during the survey to guarantee representativeness at the sample level.

## 6.3 Econometric Methods

### 6.3.1 The Swait and Louviere (1993) test procedure

The most common test procedure applied to control for preference dynamics over the choice sequence is the Swait and Louviere (1993) test, henceforth the SL-test. The properties of the SL-test are described in Section 3.6.2, but effectively it boils down to a likelihood ratio test of equivalence in preference structure across two datasets.<sup>74</sup> If the datasets (do not) have a

---

<sup>73</sup> We do not include responses of the first choice task in our analysis and only use the remaining eight choice tasks in the empirical section of this chapter. Accordingly, we will estimate sixteen choice task specific models, one for each choice task in each sample separately.

<sup>74</sup> The data underlying each of the sixteen models is treated as a separate ‘dataset’. We test for within sample preference dynamics by applying the SL-test to two ‘datasets’ from the same sample at different moments along



similar preference structure, they are fully combined (analyzed separately). Alternative applications have contrasted data from revealed and stated preference studies (Adamowicz et al., 1994; Brownstone et al., 2000; Cameron et al., 2002) or compared welfare estimates across different populations in benefits transfer studies (Colombo et al., 2007; Johnston, 2007; Lusk et al., 2003). Dynamics in the scale parameter over the choice sequence have also been analyzed, which is commonly interpreted as a measure of choice accuracy (e.g. Brown et al., 2008; Swait and Adamowicz, 2001b).

The hard line between either fully combining the datasets of interest or treating them as completely independent, is a major drawback of the SL-test. Treating the datasets as independent significantly decreases the efficiency of parameter estimates. Combining the datasets may, however, result in biased welfare estimates due to neglecting subtle differences in preferences. Therefore there is a trade-off between bias and efficiency. By treating each dataset of interest as independent, the SL-test does not take into account that some datasets are more likely to be comparable than others. In fact, the theory of CA predicts that preferences gradually evolve over the choice sequence before stabilizing at a specific level. More specifically, preferences at a particular stage of the choice sequence are more likely to be similar to preferences revealed in a recent choice task than to those revealed at the other end of the choice sequence. The SL-test therefore has its limitations when aiming to test for within and between sample preference dynamics. In the next section, we propose an alternative econometric model, the local multinomial logit model (L-MNL), which provides an intermediate solution between fully or not combining different datasets.

### 6.3.2 The local multinomial logit model

In stated choice experiments respondent  $i=1,2,\dots,I$  is presented with a sequence of  $T$  choice tasks. In each choice task  $t=1,2,\dots,T$  a limited number of alternatives is included in choice set  $D_{it}$ . Each alternative in the choice set is characterized by a set of attributes, which vary in their levels across alternatives and choice sets. The respondent is requested to select the most preferred alternative from the set of available alternatives. Based on micro-economic theory, the respondent is assumed to select the alternative generating the highest level of utility  $U_{ijt}$ . The random utility model (RUM) describes this utility function as a combination of deterministic and stochastic components, respectively  $V_{ijt}$  and  $\varepsilon_{ijt}$ . For simplicity, we use a linear-additive utility specification  $U_{ijt} = V_{ijt} + \varepsilon_{ijt} = X_{ijt}\beta + \varepsilon_{ijt}$ , where  $\beta$  denotes the vector of

---

the choice sequence. Between sample dynamics are analyzed in the SL-test by contrasting two ‘datasets’ at exactly the same moment in the choice sequence, but taken from a different sample.

marginal utility parameters.  $X_{ijt}$  represents a row vector of explanatory variables characterizing the chosen alternative  $j$  presented to individual  $i$  in choice task  $t$ . The error term  $\varepsilon_{ijt}$ , describes aspects of the choice process that are either unobserved or not explicitly modelled by the researcher. By imposing an i.i.d. extreme value distribution on  $\varepsilon_{ijt}$ , the model belongs to the family of logit models. Accordingly, the log likelihood of observing the vector of choices  $y$  in (6.2) can be described by the sum of logged choice probabilities of the chosen alternatives in each choice task characterized by (6.1).

$$(6.1) \quad P(y_{it} = j | \beta, X_{it}) = \frac{\exp(X_{ijt}\beta)}{\sum_{k \in D_{it}} \exp(X_{ikt}\beta)}$$

$$(6.2) \quad LL(y | X, \beta) = \sum_{i=1}^I \sum_{t=1}^T \ln(P_{ijt})$$

In the standard multinomial logit (MNL) model marginal utility  $\beta$  is assumed to be constant across respondents and over the choice sequence. We are, however, interested in sample  $s=1,2,\dots,S$  and choice task  $t$  specific preference parameters  $\beta_{st}$ . This can be achieved by estimating  $S \cdot T$  independent models, which potentially suffer from the same efficiency problems underlying the SL-test procedure. The L-MNL model increases efficiency by estimating  $\beta_{st}$  whilst using information from all available data. The L-MNL model discussed in Fan et al. (1995), Fan and Gijbels (1996), and Fosgerau (2007), is estimated by estimating  $S \cdot T$  alternative weighted MNL models. In our case, sixteen weighted MNL models will be estimated, so eight unique models within both the LSB and HSB sample. We label these as the locally estimated models.

Each locally estimated model results in a vector of parameter estimates  $\hat{\beta}_{st}$  for the respective local point (choice task  $t$  in sample  $s$ ). Conditional on the local point, (6.3) assigns a weight  $K_{ql}$  to each observation in the dataset, where  $q$  represents the sample and  $l$  the choice task number of the observation.<sup>75</sup> Let  $I_q$  measure the number of respondents in respectively the LSB and HSB sample. The weight is defined by the distance, i.e. degree of similarity, between each observation and the local point. Observations that are considered more similar to the local point, by being in the same sample ( $q=s$ ) or by being positioned at the same

---

<sup>75</sup> Note that the local point varies across models and thereby the weight of each observation in the likelihood function.

moment in the choice sequence ( $l=t$ ), receive a higher weight and therefore have more influence on the likelihood function.

$$(6.3) \quad LL_{st}(y | X, \beta) = \sum_{q=1}^S \sum_{i=1}^{I_q} \sum_{l=1}^T K_{ql} \cdot \ln(P_{qijl})$$

The weights are determined by a kernel density function  $g(\cdot)$ , which requires as inputs: (i) a vector (or matrix)  $Z_{st}$  characterizing the local point; (ii) the value of  $Z$  at a specific observation  $Z_{ql}$ ; and (iii) a set of bandwidth parameters  $h$ , such that  $K_{ql}=g(Z_{st}, Z_{ql}, h)$ . In our estimations, we control for within- and between-sample preference dynamics by means of a two-dimensional kernel density function, modelled as the product of two independent kernel density functions  $K_{ql} = K^1_l \cdot K^2_q$ . Within-sample preference dynamics are characterized by an ordered categorical variable describing the position of the choice task in the sequence. Between-sample dynamics are captured by an unordered categorical (dummy) variable, defining to which sample the local point belongs (LSB or HSB). Racine et al. (2006) show that kernel functions for ordered categorical variables and unordered categorical variables need to have the possibility to be an indicator function; and that it must be possible to smooth out a categorical variable.<sup>76</sup> The shape of the two kernel density functions  $K^1_l$  and  $K^2_q$ , described in respectively (6.4) and (6.5), fulfil these requirements when bandwidth parameters  $h_1$  and  $h_2$  are restricted to the interval  $[0,1]$ .

$$(6.4) \quad K^1_l = \begin{cases} 1 & \text{if } l=t \\ h_1^{|l-t|} & \text{if } l \neq t \end{cases}$$

$$(6.5) \quad K^2_q = \begin{cases} 1 & \text{if } q=s \\ h_2 & \text{if } q \neq s \end{cases}$$

The bandwidth parameters smooth the locally estimated preference parameters. Setting both  $h_1$  and  $h_2$  to one will result in the standard MNL model, since every observation gets the same weight. By setting  $h_1$  and  $h_2$  to zero, the L-MNL model is equivalent to estimating

---

<sup>76</sup> The former implies that the kernel density can take the value  $h=0$  for observations different than the local point. Other observations than the local point are not treated in the estimation of the L-MNL model. The latter ( $h=1$ ) accounts for the fact that within or between sample preference dynamics may not be present.

sixteen independent models.<sup>77</sup> Any intermediate specification is expected to result in an increase in efficiency in parameter estimates relative to this set of independent models, because it draws information from all observations in the dataset. (6.4) reveals that decisions at the other end of the choice sequence are treated as more dissimilar and receive a lower weight compared to choice tasks closer to the local point. If the bandwidth parameter is too large, then there is risk of over-smoothing. Too much detail disappears and parameter estimates may become biased. If the bandwidth is too small, then there is risk of under-smoothing, that is, over-fitting to random fluctuations in the data. Model evaluation criteria, like the *Akaike Information Criterion* (AIC) and *Bayesian Information Criterion* (BIC), can be used to select the optimal bandwidth. Hurvich et al. (1998) note that the AIC can lead to under-smoothing, while the BIC tends to support a high degree of smoothing. In this chapter, we apply the corrected AIC (AICc) as a model selection criterion, which introduces a penalty for additional parameters in the model compared to the AIC. As a rule of thumb, models are considered significantly different if the difference between model criteria is larger than  $3/(T \cdot T)$  (Charlton and Forthringham, 2009). The number of parameters in the model can be approximated by evaluating the trace of the hat-matrix (more details are provided in Appendix 6.A). More specifically, we perform a manual grid search to identify the optimal set of bandwidth parameters.

Note that with  $h$  being greater than zero, the sum of the weights will by definition be larger than the number of observations at the local point. This represents the element of smoothing. Rescaling the overall sum will not affect the parameter estimates at the local point, because all observations are rescaled by the same scalar. Rescaling will, however, affect the hat-matrix and thereby reduce the approximated number of parameters estimated for the overall model, which will affect our model evaluation criterion to determine the optimal bandwidth parameters. Hence, estimations are conducted without rescaling the weights. Indeed, local parameter estimates use data from observation considered to be similar, thereby the same data is used for different local estimates, but using different weights each time. Each of these local estimates is estimated independently and we compare the local parameter estimates using normal testing procedures. Correcting for overlapping samples in these tests is left for future research.

---

<sup>77</sup> For increasing sample sizes, group specific parameter estimates are better identifiable. Therefore, the local estimate becomes less dependent on the revealed behaviour by similar socio-economic groups. In fact, with sample size increasing to infinity, the bandwidth parameters will increase to one if all socio-economic groups have exactly the same preferences or to zero otherwise. In both cases a standard MNL model is estimated, providing consistent parameter estimates.

The set of locally estimated preference parameters is used to derive (scale free) welfare measures of interest. These may include marginal WTP for particular attributes, or mean WTP for a specific scenario (alternative), or changes in consumer surplus in general. Statistical tests are performed to test whether these welfare measures reveal any within or between sample preference dynamics. The same set of tests is performed on the results from the sixteen independent models. We hypothesize that the L-MNL increases efficiency of the parameter estimates and thereby increases the power of the test relative to these independently estimated models. The bandwidth parameters of the L-MNL model are informative on the extent to which decisions at various stages of the choice sequence can be treated as similar. This is comparable to the purpose of the SL-test in testing for within and between sample preference dynamics. If the SL-test finds preference dynamics in the database, the researcher still needs to conduct the same tests to find out whether the dynamics in the preference structure also translate into dynamics in the welfare measures of interest. The L-MNL model and the SL-test are comparable in the sense that both methods perform a preference structure test. The SL-test performs a likelihood ratio test to find out whether a variation in preference parameters results in an improvement in model fit, while the L-MNL has a similar purpose by optimizing the selected information criterion through changes in (local) preference and bandwidth parameters. The L-MNL, however, offers a more flexible approach.

#### **6.4 The dataset**

In each choice task, two alternative (unlabelled) public policy programs and a status quo (SQ) (opt-out) alternative were presented to the respondent. Each policy alternative is described by four attributes: (i) a reduction in flood probability; (ii) compensation of the material damage to each household after a coastal flood has occurred; (iii) available evacuation time; and (iv) an increase in the annual tax to the water authority for all households, including the respondent's household. Table 4.1 already described the design levels of each attribute and the definition of the SQ option. As described in Section 4.4.2, a potential starting point bias is introduced in the first choice task. More specifically, respondents in the LSB (HSB) sample were presented with the cost levels €40 and €80 (€120 and €160) for respectively the first and second alternative in the instructional choice card. The policy alternatives depicted in the instructional choice task were identical for all other attributes in both samples. The remaining eight choice cards presented to the respondents in both versions come from the same experimental design.

The sample size consists of 477 respondents, respectively 247 in the HSB and 230 in the LSB sample. Together these respondents account for 4,293 choices (477 times 9 choice tasks). The independent sampling strategy resulted in two sets of respondents comparable in terms of their main socio-economic characteristics. Statistical tests fail to reject the null-hypothesis of equivalence in the distribution and central tendency of the key socio-economic indicators between both samples, such as income, gender and age (Table 6.1). Accordingly, we present a set of attributes-only models to facilitate the illustration of the L-MNL model.<sup>78</sup>

**Table 6.1: Testing for equivalence in socio-economic sample characteristics**

Variable	Type of variable	Description	Test	d.f.	Test-statistic	p-value
Income	Categorical	10 (ordered) income categories	Chi-square test	9	8.52	0.48
Gender	Dummy	1= male ; 0 = female	Chi-square test	1	1.14	0.29
Age	Continuous	Respondent age (18-65)	Kolmogorov-Smirnov test	-	0.08	0.50

## 6.5 Results

### 6.5.1 Preference dynamics according to choice shares

The development in the choice shares across the alternatives over the nine choice tasks are reported in Table 6.2. Shares for the first choice task highlight that respondents tend to select the cheaper option (alternative 1) and the share of SQ responses in the HSB sample (21%) is higher relative to the LSB sample (13%). The  $\chi^2$ -test rejects the null hypothesis of an identical distribution of choice shares in the first choice task across the two subsamples at the 10% level. To test whether the price difference introduced in the first choice task induced a (persistent) starting point bias, we base the rest of the analysis on choice tasks 2-9.

<sup>78</sup> More extensive versions of the L-MNL model can additionally control for socio-economic characteristics and heterogeneity in risk perceptions in either the kernel density function, the utility function or both (see Koster and Koster, 2011).

**Table 6.2: Share of respondents selecting each alternative in the HSB and LSB subsamples**

Choice Task	Alt 1		Alt 2		SQ		Shares Test <sup>1</sup>		SQ Test <sup>2</sup>	
	HSB	LSB	HSB	LSB	HSB	LSB	$\chi^2$ -test statistic	p-value	$\chi^2$ -test statistic	p-value
1	54%	61%	25%	26%	21%	13%	5.52	0.06*	5.37	0.02**
2	41%	37%	45%	40%	15%	23%	5.63	0.06*	5.63	0.02**
3	44%	39%	41%	35%	15%	26%	9.07	0.01**	9.07	0.00***
4	38%	36%	43%	35%	19%	29%	6.02	0.05**	5.62	0.02**
5	42%	36%	38%	36%	20%	28%	4.35	0.11	4.20	0.04**
6	44%	35%	38%	38%	19%	27%	6.08	0.05**	4.72	0.03**
7	40%	37%	42%	38%	18%	25%	3.05	0.22	3.05	0.08*
8	42%	33%	43%	42%	15%	25%	8.94	0.01**	7.83	0.01**
9	40%	38%	36%	37%	24%	25%	0.26	0.88	0.11	0.74
Average Tasks 2-9	41%	36%	41%	38%	18%	26%	35.54	0.00***	34.89	0.00***

<sup>1</sup> Test for between sample differences in choice shares, 2 degrees of freedom

<sup>2</sup> Test for between sample differences in propensity to select the SQ option, 1 degree of freedom

\*[\*\*](\*\*\*) Denotes significance at the 10 [5] (1) % level

Averaged over choice tasks 2-9, alternatives 1 and 2 are both selected 41% of the time in the HSB sample. In the LSB sample this is respectively 36% and 38%. The within sample equivalence in choice shares for the unlabelled alternatives 1 and 2 is a direct consequence of alternating their order of appearance across versions of the design. We limit our discussion in this subsection to variations in the propensity to select the SQ-option within and between the two samples.<sup>79</sup>

Compared to the first choice task, the share of SQ responses doubles on average in the LSB sample to 26%, whereas this share decreases on average, from 21% to 18% in the HSB sample. Moreover, the share of SQ responses in the LSB sample is consistently higher than the same share in the HSB sample. Statistical support for the presence of a SPB between samples is provided by the  $\chi^2$ -test, which rejects the null hypothesis of an identical propensity to select the SQ option in both samples over choice cards 2-9 at the 1% level. Since the experimental design and respondent key characteristics are equivalent in both samples, we attribute this support for a starting point bias to the initial choice task.

A closer look at the dynamics of the choice shares over the choice sequence reveals that after the first choice task a jump occurs.<sup>80</sup> After this jump, choice shares vary around their

<sup>79</sup> Shares for alternatives 1 and 2 are barely informative, because of the alternate ordering and presenting each choice card multiple times at different moments in the choice sequence.

<sup>80</sup> This jump is primarily caused by the set-up of the survey. Instead of using the same choice card in subsequent choice tasks, the full design is used and presented at each moment in the choice sequence to the respondents.

mean levels, suggesting stability in preferences. Choice task nine stands out, revealing a relatively higher share of SQ responses in the HSB sample, which comes close to the level in the LSB sample. Indeed, the  $\chi^2$ -tests in the final columns of Table 6.2 reveal that only in the final choice task we cannot reject the null-hypothesis of an equivalent propensity to select the SQ option at the 10% level in both samples. A priori there is no reason why choice task nine should stand out. The same choice cards have been answered by similar respondents during the choice sequence. Hence, the choice shares suggest that the observed SQ share in choice task nine is somewhat ad hoc and structural preference dynamics between samples are absent. The choice shares also provide limited support for within sample preference dynamics. The only significant differences in the propensity to select the SQ option are identified in the HSB sample, where the share of the SQ option in choice tasks two, three and eight is significantly lower than the same share in choice task nine.<sup>81</sup> The choice shares are in line with Ariely et al.'s (2003) theory of CA supporting a persistent SPB in the propensity to select the SQ option. Anchoring respondents on a low price in the first choice task increases this propensity. In particular, the remarkable jump in choice shares in the HSB sample during the final choice task questions whether preferences have stabilised by the end of the choice sequence.

### **6.5.2 Preference dynamics according to choice task and sample specific models**

Choice shares do not reflect the extent to which alternatives and their attribute levels affect decisions and hence whether *welfare measures* are subject to a SPB and dynamics over the choice sequence. To this end we present a set of sixteen choice task and sample specific marginal WTP estimates in Table 6.3.<sup>82</sup> The high standard errors indicate that estimation at the choice task level is inefficient. Moreover, the 95% confidence intervals reveal that marginal WTP estimates for the probability and evacuation attributes are not significant in most models at the 5% level. Mean marginal WTP levels also reveal that preference dynamics follow a somewhat ad hoc pattern over the choice sequence. The latter is in line with earlier comments that under-smoothing may result in over-fitting due to random fluctuations in the data. In Table 6.4 we test whether anchoring on a lower price in the first choice task results in significantly lower marginal WTP estimates in the LSB sample. Although the high standard errors make it difficult to identify significant differences in WTP at the 5% level, Table 6.4 indicates that at the start of the survey respondents in the LSB sample have a higher tendency

---

<sup>81</sup> Results of the  $\chi^2$ -test for within sample differences in choice shares are available upon request.

<sup>82</sup> Standard deviations of the scale-free marginal WTP estimates are derived by means of the Krinsky and Robb method (Krinsky and Robb, 1986; 1990). The test statistic used in all methods to compare simulated marginal WTP distributions is based on the one-sided complete combinatorial approach described in Poe et al. (2005).



to select the SQ option.<sup>83</sup> This effect disappears after choice task three when significantly higher WTP estimates in the HSB are reported for the probability attribute, and the compensation attribute in respectively choice tasks five and six. In the final two choice tasks, marginal WTP for the evacuation attribute is lower in the LSB sample, but only at the 10% level. In contrast to the choice shares, these estimates only provide limited support for a (persistent) SPB in welfare estimates. The relatively small sample size and complexity of the test contribute to these results, but most remarkable is that differences in marginal WTP estimates between the two samples are not consistently found for the same attributes over the choice sequence. In choice tasks four and seven none of the reported welfare estimates of the two samples are significantly different from each other at the 10% level.

---

<sup>83</sup> An increase in the alternative specific constant and reduction in the cost coefficient both decrease the reported ratio and imply a higher probability to select one of the policy alternatives.

**Table 6.3: Marginal WTP estimates based on choice task specific models**

Sample HSB		ASC				PROB				COMP				EVAC			
Task	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	
2	118.20	32.80	65.00	194.20	8.50	5.50	-1.20	19.80	1.07	0.36	0.56	1.91	1.60	2.00	-2.30	5.50	
3	130.40	35.20	77.40	212.30	4.40	5.00	-5.30	14.50	0.96	0.34	0.46	1.75	2.10	2.10	-1.70	6.20	
4	83.60	22.90	39.50	129.40	3.60	3.90	-4.00	11.70	1.00	0.26	0.57	1.60	1.70	1.60	-1.30	5.00	
5	55.50	22.90	7.90	99.00	8.80	4.60	0.80	18.60	1.18	0.32	0.72	1.89	1.70	1.50	-1.30	4.80	
6	43.50	17.20	8.00	76.10	11.60	3.30	5.80	19.00	0.92	0.18	0.60	1.31	2.90	1.10	0.80	5.20	
7	75.90	20.10	38.50	117.50	9.60	3.80	2.60	17.50	0.81	0.20	0.46	1.26	1.60	1.40	-1.10	4.30	
8	90.80	14.70	63.30	121.40	2.00	2.50	-2.90	6.80	0.80	0.15	0.54	1.12	2.30	0.90	0.60	4.20	
9	40.60	19.20	-1.10	76.00	9.30	3.90	2.40	17.80	0.88	0.21	0.52	1.37	3.90	1.50	1.30	7.20	

Sample LSB		ASC				PROB				COMP				EVAC			
Task	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	
2	65.00	25.40	16.90	111.20	8.80	4.30	1.20	17.80	0.70	0.30	0.28	1.27	2.60	2.50	-0.70	6.50	
3	76.60	17.90	40.20	111.00	3.80	3.30	-2.80	10.30	0.43	0.18	0.09	0.81	2.80	1.40	0.10	5.60	
4	63.70	33.90	-3.20	120.90	6.50	6.10	-3.20	18.80	0.60	0.36	0.08	1.31	1.30	2.40	-3.20	6.10	
5	43.60	18.20	5.90	78.90	10.10	3.40	4.10	17.50	0.62	0.18	0.28	1.01	1.40	1.30	-1.10	4.00	
6	35.80	17.90	-1.60	70.40	3.30	3.20	-3.00	9.80	1.05	0.20	0.70	1.49	2.10	1.20	-0.30	4.50	
7	54.70	13.10	28.20	79.90	4.80	2.40	0.20	9.70	0.61	0.13	0.37	0.87	2.70	1.00	1.00	4.70	
8	67.90	16.50	35.30	100.50	4.70	2.90	-0.90	10.60	0.68	0.16	0.37	1.02	0.20	1.10	-2.10	2.40	
9	49.20	15.60	18.30	80.10	6.10	2.90	0.40	11.90	0.89	0.17	0.58	1.24	1.20	1.20	-1.00	3.70	

PROB – (€ per household per year for an extra 1,000 years in the denominator of the flood probability, from e.g. 1/4,000 → 1/5,000)

COMP – (€ per household per year for an extra percentage point of compensation)

EVAC – (€ per household per year for an extra hour of evacuation time)

**Table 6.4: p-values for higher marginal WTP values in the HSB sample using choice task specific models**

Task	ASC	PROB	COMP	EVAC
2	0.08*	0.53	0.17	0.65
3	0.05*	0.46	0.05*	0.62
4	0.30	0.66	0.13	0.44
5	0.33	0.61	0.04**	0.45
6	0.38	0.03***	0.69	0.30
7	0.19	0.13	0.19	0.75
8	0.14	0.76	0.29	0.07*
9	0.63	0.25	0.53	0.07*

\*[\*\*](\*\*\*) indicates significance at the 10[5](1)% level (one sided test)

Regarding within sample preference dynamics, choice tasks two and three in the HSB sample stand out and differ from most other choice tasks on the ratio of the alternative specific constant (ASC) over the cost coefficient (see Table 6.5). Encountering lower prices after the first choice task thus decreases the propensity to select the SQ option.<sup>84</sup> This anchoring effect becomes less strong and seems to wear out rapidly. The same ratio is significantly different between choice task three and choice tasks five and six in the LSB sample at the 10% level. However, a similar impact and decay of the initial choice task cannot be identified in this sample. Table 6.5 highlights that more significant within sample preference dynamics are found, but these are likely to be subject to under-smoothing effects. In particular, choice task eight in the HSB sample stands out revealing a low marginal WTP estimate for the probability attribute relative to choice tasks which are close by in the choice sequence. Within the LSB sample a consistent pattern also seems to be lacking. Choice task five reports a somewhat higher marginal WTP for the probability attribute at the 10% level compared to choice tasks three, six and seven. Similarly, marginal WTP for the evacuation attribute is a bit lower in choice task eight compared to choice tasks three and seven. Finally, marginal WTP for an additional percentage of compensation in choice task six seems to stand out. As such, the choice task and sample specific model results provide limited support for a (persistent) SPB and the presence of consistent within sample preference dynamics. Only within the HSB sample we are able to identify a change in the propensity to select the SQ option and a related decay of the impact of the initial choice task. However, by treating each of the sixteen models as independent, gradual changes in preferences are not captured and

<sup>84</sup> Such a status quo effect may be induced by what Loomes et al. (2009) label as taste uncertainty. Uncertain respondents may exhibit trade-off resistance. Presenting them with lower prices may alleviate such trade-off resistance. Balcombe and Fraser (2011) also look into the impacts of preference uncertainty on the propensity to select the “Don’t Know” option.

under-smoothing may have resulted in over-fitting due to the random fluctuations as revealed in Table 6.5.

**Table 6.5: Overview of significant within sample preference dynamics using choice task specific models**

Task 1	Task 2	HSB		LSB		COMP	EVAC
		ASC	PROB	ASC	PROB		
2	3						
2	4						
2	5	***					
2	6	***					
2	7						
2	8						
2	9	****					
3	4						
3	5	***		+	-		
3	6	****		+		****	
3	7	+					
3	8						+
3	9	****				**	
4	5						
4	6	+	-			-	
4	7						
4	8						
4	9	+					
5	6				+	-	
5	7				+		
5	8	-	+				
5	9						
6	7					***	
6	8	-	****	-		+	
6	9						
7	8		**				***
7	9	+				-	
8	9	***	**				

+ indicates the welfare measure is significantly higher in Task 1 compared to Task 2

- indicates the welfare measure is significantly lower in Task 1 compared to Task 2

\*[\*\*](\*\*\*\*) indicates significance at the 10[5](1)% level

### 6.5.3 Preference dynamics according to the L-MNL model

Setting both bandwidth parameters in the L-MNL model to zero results in the same set of parameter estimates as discussed in the previous section. The approximated number of parameters for this L-MNL specification, i.e. Model 1 in Table 6.6, is close to the number of

parameters underlying Table 6.3. The difference arises due to the use of robust standard errors.<sup>85</sup> The high number of parameters in Model 1 results in an improved model fit, but the *AICc* for this specification is lower than the *AICc* for a standard MNL model neglecting between and within preference dynamics (Model 2). Hence, treating the different samples and choice tasks as independent blocks of observations leads to a reduction in the information criterion and parameter efficiency. In Model 3, we have optimized the *AICc* by controlling for both within ( $h_1=0.43$ ) and between ( $h_2=0.20$ ) sample preferences dynamics, while in Model 4 the *AICc* is optimized only with respect to between sample preference variations setting  $h_1=1$ . The fit for Model 3 is significantly better relative to Model 4 supporting the notion that within sample preference dynamics are present over the choice sequence. Simultaneously, the bandwidth parameter controlling for between sample preference variation highlights that both samples should not be treated as being the same. Finally, since there is only a marginal increase of 0.01 in  $h_2$  between Model 3 and Model 4, the kernel densities operate in a relatively independent fashion.

**Table 6.6: Overview of the *AICc* for alternative L-MNL specifications**

Model	Description	Bandwidth	Bandwidth LL	Approx	<i>AICc</i>	
		$h_1$	$h_2$			# of pars
		(within)	(between)			
(1)	Within + between sample variation	0.00	0.00	-3660.78	79.74	1.9616
(2)	MNL	1.00	1.00	-3720.78	10.63	1.9559
(3)	Optimal bandwidth parameter	0.43	0.20	-3677.24	24.01	1.9402
(4)	Optimal between sample variation	1.00	0.19	-3700.79	15.46	1.9480

Table 6.7 reports the choice task and sample specific marginal WTP estimates obtained for L-MNL Model 3. The major advantage of the L-MNL model is its increase in efficiency illustrated by the reduction in standard errors compared to Table 6.5. Improvements up to 72% are found and standard errors of the marginal WTP estimates reduce by 54% on average. The 95% confidence intervals for the probability and evacuation attributes are now strictly positive. The smoothing procedure also has an impact on mean marginal WTP estimates. Table 6.8 reveals that the HSB sample has a higher marginal WTP for the compensation attribute until choice task six. This effect is only significant at the 5% level in choice task three. Similarly, we find a lower tendency to select the SQ option in choice tasks two and three in the HSB sample. The effect is significant at the 10% level and the L-MNL

<sup>85</sup> Without specifying robust standard errors, there would be five parameters in each MNL model resulting in a total of 80 parameters (sixteen independent models each consisting of five parameters).

model thereby provides limited support for a starting point bias due to anchoring on the price attribute of the first choice task. After five choice tasks welfare measures seem to converge between samples, only marginal WTP for the evacuation attribute is significantly higher in the HSB sample in the final choice task, but only at the 10% level. Clearly, smoothing parameter estimates results in a more consistent pattern of between sample preference dynamics compared to Table 6.4 as most random fluctuations are filtered out.

**Table 6.7: Marginal WTP estimates based on L-MNL Model 3**

Sample	HSB				PROB				COMP				EVAC			
	ASC															
Task	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%
2	99.28	13.93	73.33	128.11	6.85	2.38	2.32	11.68	0.93	0.16	0.65	1.28	1.89	0.98	-0.03	3.87
3	94.13	11.25	73.27	117.69	5.73	1.91	2.14	9.65	0.90	0.14	0.65	1.19	1.98	0.87	0.31	3.71
4	77.65	10.08	58.00	97.89	5.98	1.72	2.80	9.54	0.93	0.12	0.71	1.18	1.91	0.73	0.54	3.39
5	63.58	9.20	45.81	82.05	7.56	1.61	4.53	10.81	0.95	0.11	0.74	1.19	1.99	0.62	0.75	3.23
6	58.59	8.48	41.87	74.91	8.23	1.38	5.65	11.05	0.90	0.10	0.71	1.11	2.28	0.54	1.22	3.38
7	66.17	8.72	49.24	83.32	7.25	1.37	4.62	9.99	0.84	0.10	0.66	1.04	2.19	0.54	1.15	3.27
8	70.95	8.18	55.04	87.31	5.36	1.27	2.94	7.89	0.82	0.09	0.65	1.01	2.29	0.53	1.28	3.37
9	58.52	9.65	39.18	77.55	6.74	1.68	3.61	10.17	0.85	0.11	0.64	1.07	2.77	0.70	1.45	4.19

Sample	LSB				PROB				COMP				EVAC			
	ASC															
Task	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%	Mean	St. error	2.5%	97.5%
2	71.82	12.41	46.90	95.66	6.92	2.08	2.98	11.23	0.67	0.13	0.44	0.94	2.30	0.90	0.61	4.13
3	72.14	10.42	51.40	92.51	5.60	1.71	2.24	9.09	0.62	0.11	0.42	0.84	2.22	0.74	0.82	3.73
4	63.61	10.29	43.60	83.54	6.21	1.66	3.00	9.57	0.68	0.11	0.47	0.90	1.81	0.68	0.51	3.18
5	53.32	8.73	36.24	70.43	7.07	1.48	4.24	10.09	0.74	0.09	0.56	0.93	1.77	0.56	0.68	2.90
6	50.31	8.06	33.74	65.79	5.83	1.30	3.32	8.42	0.81	0.08	0.65	0.99	1.99	0.50	1.04	2.98
7	55.89	7.48	40.98	70.55	5.40	1.23	3.04	7.87	0.73	0.07	0.59	0.88	2.02	0.49	1.09	3.00
8	60.84	8.06	44.85	76.67	5.10	1.36	2.55	7.85	0.74	0.08	0.58	0.91	1.44	0.54	0.41	2.50
9	55.23	8.89	37.76	72.79	5.72	1.53	2.73	8.73	0.80	0.09	0.63	0.99	1.52	0.65	0.29	2.82

PROB – (€ per household per year for an extra 1,000 years in the denominator of the flood probability, from e.g. 1/4,000 → 1/5,000)

COMP – (€ per household per year for an extra percentage point of compensation)

EVAC – (€ per household per year for an extra hour of evacuation time)

**Table 6.8: p-values for higher marginal WTP in the HSB sample using L-MNL Model 3**

Task	ASC	PROB	COMP	EVAC
2	0.07*	0.51	0.10*	0.62
3	0.07*	0.48	0.05**	0.58
4	0.16	0.54	0.06*	0.46
5	0.21	0.41	0.07*	0.40
6	0.24	0.10	0.25	0.34
7	0.19	0.16	0.20	0.40
8	0.19	0.44	0.25	0.13
9	0.40	0.33	0.39	0.09*

\*[\*\*](\*\*\*) indicates significance at the 10[5](1)% level (one sided test)

The same observation can be made for within sample preference dynamics in Table 6.9. Within the HSB sample choice tasks two and three reveal a higher ratio of the ASC over the cost coefficient relative to choice tasks five to nine, which implies a lower tendency to select the SQ option at the start of the choice experiment. This effect is significant in all cases at the 5% level. In choice task four, this effect is still present compared to choice tasks six and nine and only at the 10% significance level. The L-MNL model thus supports a gradual decay of the impact of the initial choice task on the tendency to select the SQ option in the HSB sample. Regarding the other policy attributes only a significant difference at the 10% level is found for the probability attribute when comparing choice tasks six and eight in the HSB sample. Within the LSB sample, the ratio of the ASC over the cost coefficient in choice task two is significantly higher compared to choice task six, but only at the 10% level. The same ratio is also higher in choice task three relative to choice tasks five and six. This may indicate that respondents in general have a lower tendency to select the opt out at the start of the survey. This effect can be amplified by anchoring on respondents on high prices in the initial choice task. Finally, marginal WTP for the compensation attribute is lower in choice task three when contrasted against choice tasks six and nine.



**Table 6.9: Overview of significant within sample preference dynamics based on L-MNL Model 3**

Task1	Task2	HSB		LSB	
		ASC	PROB	ASC	COMP
2	3				
2	4				
2	5	***			
2	6	***		*	
2	7	**			
2	8	**			
2	9	***			
3	4				
3	5	**		*	
3	6	***		**	.*
3	7	**			
3	8	**			
3	9	***			.*
4	5				
4	6	*			
4	7				
4	8				
4	9	*			
5	6				
5	7				
5	8				
5	9				
6	7				
6	8		*		
6	9				
7	8				
7	9				
8	9				

\*[\*\*](\*\*\*) indicates significance at the 10[5](1)% level

In summary, primarily the HSB sample reveals an impact of the initial choice task on welfare estimates, but this effect gradually wears out after the third choice task. Combined with the lack of overwhelming support for a SPB and the comparable welfare estimates across samples after choice task five, the L-MNL model provides support for Plott's DPH. The extent to which preferences are stable after the fifth choice task remains, however, questionable. First, between sample differences in marginal WTP for the evacuation attribute are found in choice task nine. Second, within the HSB sample dynamics in marginal WTP for the probability attribute are also observed between choice tasks six and eight. Despite the

new estimation method, standard errors remain relatively high due to our limited sample size. Not only does this affect the efficiency of our estimates, it also affects the number of times a specific card is answered at each moment in the sequence and thereby the impact of individual respondent characteristics on choice task specific parameter estimates. Keeping this drawback in mind, we interpret our results as providing indicative support for the DPH.

#### **6.5.4 Preference dynamics according to the Swait-Louviere test procedure**

The SL-test for between sample preference dynamics does not find a starting point bias between both samples. The preference structure in the HSB and LSB samples is found to be equivalent in all choice tasks except choice tasks three and seven in Table 6.10. Indeed, choice task three revealed a significant difference in the ASC over cost ratio and marginal WTP for the compensation attribute between the two samples in the L-MNL model. For choice task seven the L-MNL model did not detect a significant difference in between sample welfare estimates.<sup>86</sup> Also limited within sample preference dynamics are identified by the SL-test. The LSB sample does not reveal any differences in preference structure over the choice sequence, while in the HSB sample only significant differences are found between respectively choice tasks two, three and choice tasks six and nine (see Table 6.11). This pattern is consistent with our L-MNL estimates revealing that respondents in the HSB sample have a lower tendency to select the SQ option in choice tasks two and three. These results highlight the limitations of the SL test procedure. The independently estimated models result in inefficient parameter estimates, but these models are still used as inputs in the SL test. Hence, the test clearly supports smoothing the parameter estimates by fully combining most samples, which prevents us from testing for the presence of preference dynamics in the welfare measures of interest. Even if the preference structure is equivalent, then still differences in welfare measures can be present across samples (see for example choice task two in Tables 6.4 and 6.8). By over-smoothing, the SL-test is not as flexible as our L-MNL model, and can therefore not provide as much insight into patterns of preference dynamics over the choice sequence.

---

<sup>86</sup> Analyzing the parameters based on the SL-test in preference space reveals that a significant difference can be identified between the estimated parameters for the evacuation and cost attribute in this choice task. The properties of the SL test do not allow us to attribute these differences to variations in preferences or in scale.

**Table 6.10: Results for the between sample Swait and Louviere (1993) test**

Task	LL_HSB	LL_LSB	LL_SUM	LL_ Pooled_ scale	LL_ Pooled	LR-test1	p-value	LR-test2	p-value	scale ln(HSB / LSB)
2	-234.33	-232.03	-466.36	-468.94	-470.09	5.17	0.27	2.29	0.13	0.33
3	-238.44	-231.20	-469.64	-475.27	-476.07	11.27	0.02**	-	-	-
4	-238.93	-242.19	-481.11	-482.71	-486.01	3.20	0.53	6.59	0.01**	0.66
5	-236.19	-225.98	-462.17	-466.00	-466.28	7.66	0.10	0.55	0.46	0.16
6	-222.35	-219.96	-442.31	-445.93	-446.89	7.23	0.12	1.93	0.17	0.26
7	-232.57	-209.10	-441.67	-446.91	-447.26	10.48	0.03**	-	-	-
8	-217.26	-221.97	-439.23	-442.88	-444.74	7.30	0.12	3.71	0.05*	0.36
9	-241.60	-216.69	-458.29	-459.98	-460.21	3.38	0.50	0.47	0.49	-0.14

LR-test1 – Test for differences in the preference parameters, 4 degrees of freedom

LR-test2 – Test for differences in the scale parameter, 1 degree of freedom

\*(\*\*)[\*\*\*] indicates significance at the 10(5)[1]% level

The SL-test has the benefit of testing for differences in the scale of the utility function. Significant scale differences between the two samples are found in choice tasks four and eight, where the HSB sample is found to have a higher scale parameter. Our L-MNL model confirms choice task eight does not display a difference in welfare estimates between the two samples, but in choice task four marginal WTP for the compensation attribute is significantly lower in the LSB sample. Here, the inefficiency of the individual models is thus transferred into the scale parameter. The fact that the HSB sample reveals a higher scale parameter implies that respondents' choices exhibit less randomness in the LSB sample. It is hard to draw any conclusions on this regarding the degree of preference uncertainty between the two samples. Table 6.11 reveals significant within sample differences in the scale parameter at various stages of the choice sequence. Within the HSB sample choice task eight stands out by having a higher scale parameter relative to most preceding choice tasks. This learning effect, however, disappears rapidly since choice task nine has a lower scale parameter relative to the preceding choice tasks. Also within the LSB sample indications of learning effects are found. Choice task seven has a higher scale parameter relative to choice task two, three, four and five. Choice task four stands out by having the lowest scale parameter. The scale parameter stabilizes after choice task five, which indicates convergence of preferences. Given the limitations of the SL-test under the current sample size, it is, however, unclear whether the SL-test over-smoothes the within (and between) sample preference dynamics, an effect which may be consequently picked up by the scale parameter.

**Table 6.11: Within sample preference and scale dynamics in the Swait and Louviere (1993) test**

Task 1	Task 2	HSB		LSB			
		p-value equivalence of preferences	p-value equivalence of scale	scale ln( task2 / task1)	p-value equivalence of preferences	p-value equivalence of scale	scale ln( task2 / task1)
2	3	0.98	0.71	-0.07	0.62	0.99	0.00
2	4	0.56	0.67	-0.08	0.88	0.13	-0.42
2	5	0.38	0.85	-0.04	0.53	0.55	0.14
2	6	0.09*	-	-	0.23	0.21	0.28
2	7	0.52	0.95	0.01	0.45	0.02**	0.47
2	8	0.22	0.14	0.26	0.45	0.28	0.24
2	9	0.05*	-	-	0.56	0.13	0.32
3	4	0.61	0.94	-0.02	0.90	0.15	-0.40
3	5	0.26	0.86	0.03	0.35	0.47	0.17
3	6	0.07*	-	-	0.11	0.18	0.31
3	7	0.43	0.69	0.08	0.82	0.02**	0.47
3	8	0.44	0.07*	0.32	0.37	0.25	0.26
3	9	0.06*	-	-	0.20	0.11	0.35
4	5	0.86	0.80	0.05	0.96	0.03**	0.55
4	6	0.42	0.12	0.28	0.69	0.01***	0.70
4	7	0.79	0.64	0.09	0.97	0.00***	0.86
4	8	0.91	0.06*	0.34	0.98	0.01**	0.63
4	9	0.39	0.86	-0.04	0.90	0.00***	0.73
5	6	0.74	0.20	0.23	0.32	0.55	0.13
5	7	0.82	0.84	0.04	0.58	0.10*	0.32
5	8	0.34	0.09*	0.30	0.67	0.64	0.10
5	9	0.52	0.68	-0.08	0.65	0.38	0.18
6	7	0.80	0.30	-0.18	0.31	0.28	0.20
6	8	0.13	0.72	0.06	0.50	0.90	-0.02
6	9	0.93	0.09*	-0.30	0.87	0.80	0.05
7	8	0.49	0.14	0.26	0.43	0.21	-0.24
7	9	0.57	0.52	-0.12	0.36	0.40	-0.16
8	9	0.29	0.04**	-0.39	0.85	0.68	0.08

\*(\*\*)[\*\*\*] indicates significance at the 10(5)[1]% level

## 6.6 Summary

The existence of a set of well defined preferences in many environmental economic valuation studies can be questioned due to unfamiliarity and inexperience of respondents with the policy attributes. Plott's (1996) discovered preference hypothesis and Ariely et al.'s (2003) coherent arbitrariness are well known frameworks on this topic, providing contradicting hypotheses on the extent to which respondents cope with this preference uncertainty and how preferences evolve over a sequence of choices. In this chapter, we test for the presence of between and within sample preference dynamics in the face of an arbitrarily induced starting point bias in stated choice experiments. To this end, a uniquely designed choice experiment on flood risk valuation is applied in combination with a new econometric model, which is better suited to test for gradual changes in preferences over the choice sequence. The latter model is contrasted against the Swait and Louviere (1993) test procedure, the most common approach to test for preference dynamics. We argue that the latter test is not designed and suited to test for dynamics in welfare estimates, in particular if samples of smaller sizes are considered.

This chapter finds limited support for the existence of a (persistent) starting point bias in stated choice experiments. These results are in line with findings by Ladenburg and Olsen (2008) and support the discovered preference hypothesis. The sample provided with a higher bid vector at the start of the choice sequence has a lower tendency to select the status quo option in subsequent choice tasks and thereby reveals lower cost sensitivity. The impact of the initial choice task seems to gradually disappear after the third choice task, resulting in a set of stable marginal WTP estimates in both samples. More specifically, after the fifth choice task welfare estimates are not statistically different across the two samples in our novel L-MNL model at the 5% significance level. As such, we find evidence of preference dynamics in both welfare estimates and scale parameters induced by arbitrary framing effects. This is in contrast to the predictions of the standard micro-economic framework and the related hypotheses tested in this chapter.

Four implications follow from this chapter. First, researchers should be aware of potential dynamics in welfare estimates over the choice sequence and not only focus on inherent differences in preferences across respondents (e.g. Hess and Rose, 2009). Absence of stable welfare estimates in stated choice experiments would, however, complicate welfare analysis and policy recommendations as it becomes unclear which choice tasks should be used for welfare analysis and how many choices should be included in the experiment. Ideally, these stated preferences are compared to choices in a revealed preference setting.

Second, the Swait and Louviere (1993) test procedure has the tendency to over-smooth the data and thereby neglect underlying dynamics in preferences when the underlying models produce inefficient parameter estimates. The local MNL logit model is more suited for the purpose of testing for preference dynamics, offering improvements in flexibility and efficiency in terms of estimating choice task specific preference parameters. Significant reductions in standard errors are observed without the need to bundle observations from various choice tasks. As such, the model is able to control for gradual changes in preferences and prevents against over-identification due to random variations in the data by smoothing parameter estimates. It should be noted that applications of the local MNL model are not restricted to variations in preferences over time. Third, additional effort needs to be placed in the development of experimental set-ups in which sample sizes and the experimental design enabling researchers to estimate choice task specific choice models. Sample sizes used in this chapter are comparable to those used in other studies by Braga and Starmer (2005) and Ladenburg and Olsen (2008) who also use around 250-300 respondents per sample. Closely related, - and despite our careful study set-up -, individual respondents could have caused the observed dynamics in preferences, since at each moment in the sequence each choice card was answered by ten respondents on average. Finally, the sensitivity of marginal WTP estimates to arbitrary initial value cues asks for careful testing of the choice experiment and careful specification of the initial choice task. Looking beyond the scope of the current chapter, an alternative approach could be to present respondents with an overview of all possible attribute levels before introducing a specific instructional choice task. In that case, starting point biases may be circumvented by not presenting a single set of arbitrary value cues to the respondent (e.g. Bateman et al., 2004). The appropriateness of the levels included in the choice experiment, however, needs to be defined in pre-testing stages while taking into account the preference uncertainty of respondents also in those stages.

## Appendix 6.A - Optimal bandwidth parameters

Fosgerau (2007) and Fröhlich (2006) argue that the bandwidth parameter generally has a larger impact on model results than the shape of the continuous kernel density itself. They also note that there is not a single bandwidth selection method considered to be the best. A practical approach is to select the smallest possible bandwidth for which all local models converge. This approach seems to work well for large datasets. However, it is unknown in advance if this will result in under-smoothing. Additional criteria are needed in order to have the possibility to test the model against the standard MNL model.

Hurvich et al. (1998) propose a statistic based on the trade-off between model fit and the number of parameters in the model, which can be used to determine the optimal bandwidth and select the appropriate model. The number of parameters in the model can be approximated by evaluating the trace of the hat-matrix  $H$  (see below). If the bandwidth  $h$  of a categorical variable is low, the fit of the model will be better, but more parameters are needed, so the trace of the hat matrix  $tr(H)$  will be higher. Model evaluation criteria like the *Akaike Information criterion* (AIC) and *Bayesian Information Criterion* (BIC) can be used for selecting the optimal bandwidth. Hurvich et al. (1998) note that the AIC can lead to under-smoothing, while the BIC tends to support a high degree of smoothing. In this paper, we apply the corrected AIC (AICc) as model selection criterion

$$AICc = \frac{-2LL(\hat{\beta})}{I \cdot T} + \frac{2 \cdot tr(\hat{H}) + 1}{I \cdot T - tr(\hat{H}) - 2},$$
 introducing an additional penalty for additional

parameters in the model compared to the AIC. As a rule of thumb, models are considered significantly different if the difference between model criteria is larger than  $3/(I \cdot T)$  (Charlton and Fortheringham 2009).

As discussed in Koster and Koster (2011), this L-MNL method has its drawbacks if panel data are used. If one does not correct for the panel nature of the data, the local standard errors will be underestimated. Therefore, the trace of the hat-matrix becomes too low, which will result in an optimal bandwidth that is too low and therefore under-smoothing of the model. We correct for this by estimating robust standard errors clustered over respondents (Freedman 2006).

We follow Nagel and Hatzinger (1992) and Koster and Koster (2011) in deriving the hat-matrix for each of the  $I \cdot T$  locally estimated weighted MNL models. Let  $\Omega_l$  represent the  $k \cdot k$  (robust) covariance matrix of parameter estimates belonging to a specific locally estimated weighted MNL model  $l$ . Alternatively,  $\Omega_l$  can be specified as the inverse hessian

matrix  $\Omega_l=(X^*{}'V_l X^*)^{-1}$ , but using the covariance matrix reduces computation time.  $X^*$  is a transformation of the design matrix  $X$ , where each observation is multiplied by the square root of its own weight  $\sqrt{K_{it}}$ .<sup>87</sup>  $V_l$  represents the locally estimated covariance matrix of choice probabilities. Due to the IIA property of the (weighted) MNL model,  $V_l$  is a block diagonal matrix containing the observation specific covariance matrices of estimated choice probabilities  $V_{it}^l$  along the main diagonal:

$$V_{it}^l = \begin{pmatrix} \hat{P}_1(1-\hat{P}_1) & \dots & -\hat{P}_{J-1}\hat{P}_1 \\ \vdots & \ddots & \vdots \\ -\hat{P}_1\hat{P}_{J-1} & \dots & \hat{P}_{J-1}(1-\hat{P}_{J-1}) \end{pmatrix}$$

$$V^l = \begin{pmatrix} V_{11}^l & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & V_{nT}^l \end{pmatrix}$$

Nagel and Hatzinger (1992) define the hat-matrix for a standard MNL model by  $H=V^{1/2}X(X'VX)^{-1}X'V^{1/2}$ . We use this specification to construct the hat-matrix for the locally estimated weighted MNL model  $l$ . Rewriting  $X^*{}'V_l X^*=X^*{}'V_l^{1/2}V_l^{1/2}X^*$  and noting the similarity between this and the specification by Nagel and Hatzinger (1992), we can specify the local Hat-matrix in the following way:  $H_l=V_l^{1/2}X^*(X^*{}'V_l X^*)^{-1}X^*{}'V_l^{1/2}$ . The specification can be further simplified by replacing the middle statement by the local covariance matrix.  $H_l=V_l^{1/2}X^*\Omega_l X^*{}'V_l^{1/2}$ . Note that for each local point a local Hat-matrix needs to be derived.

Using properties of linear algebra, we can rewrite the trace of the local Hat-matrix by  $tr(H_l)=tr(X^*\Omega_l X^*{}'V_l)$ , which saves substantial computation time. As mentioned in Section 4, the trace of the Hat-matrix approximates the number of parameters in the local model. In the eventual comparison of alternative bandwidth parameters, only the trace elements of the local hat matrix belonging to the local point are used and summed. More specifically, for the first choice card, which contains three alternatives in our case, the first two trace elements of the local hat matrix are stored. For local point two, the elements three and four form its own local hat-matrix. In order to reduce computation time, specific elements  $c$  on the trace of the local Hat-matrix can be obtained by calculating  $X^*(c,:) \Omega_l X^*{}'V_l(:,c)$ , picking the  $c$ -th row of  $X^*$  and

<sup>87</sup> More formally,  $X$  is a  $I \cdot T \cdot (J-1)$  by  $k$  matrix describing the characteristics of each alternative adjusted for a reference alternative (in our case the status quo option). Additionally, it also includes additional explanatory variables in the model. Hence, each observation is described by  $(J-1)$  rows in  $X$ .



the  $c$ -th column of  $V_l$ . The number of parameters related to a specific bandwidth parameter is approximated by summing the stored trace elements over all local models. Clearly, under uniform weights the hat-matrix reduces to the MNL hat-matrix in which the trace sums to the exact number of parameters in the model.

## Chapter 7: Implicit and explicit modelling of preference uncertainty<sup>88</sup>

This chapter addresses research question three by explaining preference uncertainty and controlling for its impacts in discrete choice models. A hybrid choice model is developed which combines the implicit and explicit measurement of preference uncertainty, as described in Chapter 3, in a unified framework. Implicit measurement of preference uncertainty refers to tracing the impact of preference uncertainty in the choice model by means of variations in the scale parameter (or other parameters of the utility function) across respondents and choice tasks. Identification is solely based on responses to the choice tasks. Explicit measurement of preference uncertainty relies on the use of choice certainty follow-up questions directly after each choice task. Since responses to these follow-up questions are hypothesized to be an indicator of actual preference uncertainty, we expect them to be of value in explaining causes of preference uncertainty. Moreover, they can help in improving identification of preference uncertainty and its impacts in the choice model when analyzed in a unified framework. Accordingly, the hybrid choice model developed in this chapter may improve the validity and reliability of willingness-to-pay estimates. In the proposed model specification, responses to the choice task and the follow-up question are treated as a simultaneous decision both affected by an unobserved (or latent) variable preference uncertainty. Section 7.1 provides an introduction into the relevant literature. Section 7.2 develops the model. Section 7.3 describes the dataset and Section 7.4 summarizes the results and Section 7.5 concludes.

### 7.1 Introduction

Within the Random Utility Model uncertain respondents are generally hypothesized to make choices that appear more random to an analyst, i.e. affecting the variance of the estimated utility function (Li and Mattsson, 1995). Such randomness in decisions may result in violations of internal consistency over a sequence of choice tasks. In Section 3.5, we have discussed the close connection between the complexity of choice tasks and preference uncertainty. The easier a choice task, the more certain respondents are likely to be about their decisions. Moreover, we have discussed a set of models, including the heteroskedastic MNL (HMNL) model, that implicitly account for preference uncertainty and choice task complexity by allowing the scale parameter to vary across respondents and the choice sequence. Clearly, most of the papers adopting an HMNL or comparable model only learn

---

<sup>88</sup> This chapter is based on joint work with Stephane Hess ([s.hess@its.leeds.ac.uk](mailto:s.hess@its.leeds.ac.uk)) and under review at the Journal of Applied Econometrics.

about scale heterogeneity (possibly interpreted as the degree of preference uncertainty) by means of observed responses to the choice tasks, and characteristics of these choice tasks and respondents.

Alternative hypotheses exist arguing that uncertain respondents adopt simplifying choice heuristics affecting the structural part of the utility function. For example, Loomes et al. (2009) develop a theoretical model in which uncertain respondents are more likely to pick the status quo alternative than is the case for certain respondents. This heuristic also finds empirical support by Balcombe and Fraser (2011) and Swait and Adamowicz (2001b). Our case in point is that irrespective of the way preference uncertainty affects decisions in the stated choice experiment, researchers only learn about preference uncertainty and its impact by observing a sequence of choices while more information is usually at hand. This chapter will contrast these alternative hypotheses regarding the impact of preference uncertainty on decisions in stated choice experiments. The main emphasis of this chapter is, however, on developing adequate ways of controlling for preference uncertainty when analyzing choice data where additional information on respondent's self-reported choice certainty levels is also available.

Researchers in stated choice experiments have incorporated post-decisional choice certainty questions in their surveys, similar to those applied in the contingent valuation literature (e.g. Brouwer et al., 2010; Hensher et al., 2011; Hensher and Rose, 2011; Lundhede et al., 2009; Olsen et al., 2011).<sup>89</sup> On the basis of respondent characteristics and the choice tasks, these papers have sought to find explanations for the degree of self-reported choice certainty. In Section 3.4, we have labelled this as explicit modelling of preference uncertainty. By analyzing these responses independently from the choice data, these explicit models only learn about preference uncertainty through the responses to the follow-up questions. For example, Olsen et al. (2011) first estimate a choice model and calculate the expected utility level for each alternative. Implicitly, they thereby assume that preference uncertainty does not affect the utility function. They then derive a deterministic measure of utility difference between the alternatives in the choice set in order to explain self-reported choice certainty. They show that self-reported choice certainty is found to increase when the utility difference becomes larger between the chosen and second best alternative.<sup>90</sup>

---

<sup>89</sup> Fenichel et al. (2009) and Balcombe and Fraser (2011) are exceptions to this rule in that they present a "Don't Know" option to respondents in addition to a status quo option.

<sup>90</sup> Comparable results are found by Brouwer et al. (2010) in a similar two-stage modelling procedure. Balcombe and Fraser (2011) observe a similar tendency to select the "Don't Know" option when alternatives in the choice set become more similar in terms of their choice probabilities. As utility differences decrease the random

The implicit and explicit modelling strategies approach the issue of preference uncertainty from a different angle. The former accounts for the impact of preference uncertainty on the choice model, while the latter seeks to explain self-reported choice certainty. If self-reported choice certainty is indeed a proper reflection of actual preference uncertainty, analyzing the implicit and explicit modelling strategies in a unified framework may result in informational gains regarding the impact of preference uncertainty on decision strategies and welfare estimates.

In this chapter, we propose a simultaneous decision framework. The need for such a simultaneous perspective is embedded within the literature discussed in Chapter 3. Preference uncertainty, and thereby the scale of the utility function, is affected by the complexity of the choice task. The characteristics of the choice task are therefore likely to affect both the decision in the choice task and the response to the choice certainty follow-up question. Alternatively, it can be argued that the two decisions are made in a sequential fashion, because the follow-up questions is answered directly after the choice task. Not accounting for preference uncertainty in the first part of the sequential model, i.e. the choice model, is likely to result in biased explanatory factors of preference uncertainty. The explicit models of preference uncertainty discussed in Brouwer et al. (2010) and Olsen et al. (2011) are an example of such a sequential modelling procedure where the impact of preference uncertainty is not taken into account in the choice model. Note that their two-stage modelling procedure also does not analyze the choice model and self-reported choice certainty model in a unified framework. Transforming the two-stage model into a unified sequential framework can easily be achieved using a Full Information Maximum Likelihood (FIML) approach. In Section 7.4, we contrast our simultaneous decision model with a sequential decision model using a FIML approach.

Albeit that Brouwer et al. (2010) and Olsen et al. (2011) fail to make the appropriate bi-directional (simultaneous) link between the choice task and self-reported choice certainty, their results are not affected by a different shortcoming which can be levelled as a criticism of other work. There exist a growing number of studies in which self-reported choice certainty is used as an explanatory variable in the specification of the utility function, and the scale parameter in particular. As an example, Lundhede et al. (2009) and Beck et al. (2011) present a model where the response to the follow-up question is directly included as an explanatory factor of the scale parameter. Similarly, Hensher and Rose (2011) weight each choice in the

---

component of the utility function becomes more dominant and assigns more equal probabilities to the alternatives in the choice task.

likelihood function by using the response to the follow-up question. Beck et al. (2011) adopt a recalibration approach conditional on the self-reported choice certainty levels comparable to approaches adopted in the CVM literature. The key shortcoming of such approaches is that they fail to recognise the likely correlation between self-reported choice certainty and unmodelled factors that enter into the random component of utility, thus putting the analyst at risk of endogeneity bias. These papers also treat the self-reported choice certainty responses as an error free measure of actual preference uncertainty, rather than recognising that self-reported choice certainty is merely a function of the true underlying degree of preference uncertainty. From this point of view, self-reported choice certainty must thus be treated as a *dependent* variable rather than as an *independent* (explanatory) variable.

## 7.2 The model

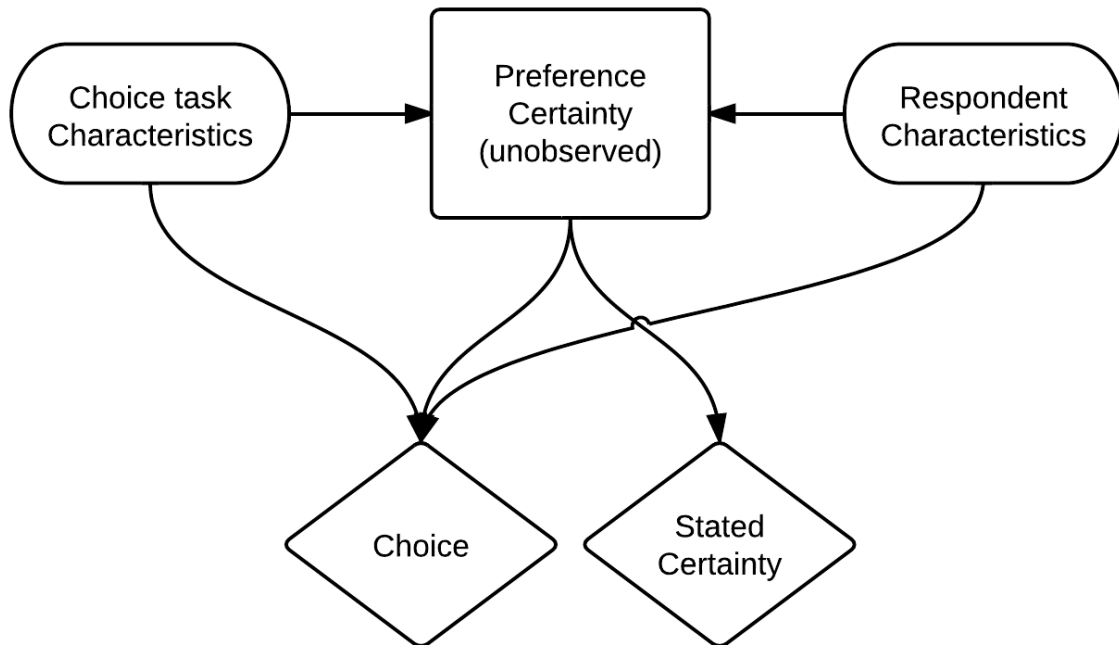
The model developed in this section addresses the two shortcomings mentioned in the previous section. First, self-reported choice certainty is treated as a dependent variable for which we seek explanatory variables in a similar vein as in Olsen et al. (2011) and related papers. Second, the choice model and the model characterizing the self-reported choice certainty responses are linked in an integrated model by means of the unobserved (or latent) variable labelled ‘preference certainty’. Estimations are conducted in a Bayesian and a FIML framework.

The proposed model treats preference certainty as a latent variable which is a function of respondent characteristics and the characteristics of the choice task at hand. By treating preference certainty as a latent variable we work around potential endogeneity issues, i.e. we can use the *latent* certainty as an explanatory variable, whereas the use of *observed* self-reported choice certainty would have created such problems. Preference certainty on itself simultaneously affects the responses to the actual choice and the follow-up question on choice certainty. Treating the responses to these two components as a simultaneous decision is not only more intuitive, it also allows us to better identify factors affecting preference (un)certainty by learning from responses to both the choice sequence and the follow-up questions. More specifically, what we labelled above as implicit and explicit modelling of preference uncertainty is now combined in a single model. Figure 7.1 describes this simultaneous modelling approach to account for preference certainty.<sup>91</sup>

---

<sup>91</sup> Note that in an alternative sequential specification, the choice model will not be affected by latent preference certainty. In that case the results from the choice model are used as an explanatory variable of latent preference certainty, which then still affects the self-reported choice certainty (see Section 7.2.5).

Figure 7.1: A latent variable approach to preference (un)certainty



### 7.2.1 Latent preference certainty

Denote preference certainty  $C_{it}$  for respondent  $i$  in choice task  $t$  as a function of respondent characteristics  $R_i$  and a set of choice task specific characteristics  $W_{it}$ . The respondent characteristics remain constant over the choice sequence, while the latter set of variables is subject to change over the choice sequence and possibly respondents.  $W_{it}$  captures, amongst other things, choice task complexity and possible learning and fatigue effects. In our simultaneous model, we learn about  $C_{it}$  by means of responses  $y_i = [y_{i1}, y_{i2}, \dots, y_{iT}]$  to a set of choice tasks  $t=1, 2, \dots, T$  and  $I_i = [I_{i1}, I_{i2}, \dots, I_{iT}]$  a set of choice task specific self-reported choice certainty responses. The latter is labelled as the indicator variable for preference certainty in Figure 7.1. More formally, we describe  $C_{it}$  in (7.1) by a linear function where the parameters  $\delta$  and  $\zeta$  measure the structural variation in preference certainty. The variate  $\xi_{it}$  is a standard i.i.d. normally distributed zero mean stochastic term with variance  $\gamma^2$  capturing measurement error.

$$(7.1) \quad C_{it} = R_i\delta + W_{it}\zeta + \rho_i + \xi_{it}$$

Recent applications of latent variable models, or hybrid choice models, in the choice modelling literature (e.g. Abou-Zeid et al., 2011; Daly et al., 2011a; Daziano and Bolduc,

2011; Hess and Beharry-Borg, 2011; Yanez et al., 2010) have focussed on underlying attitudes, and have used attitudinal questions at the level of the respondent as an indicator of these latent attitudes. Note that these responses have been captured only once per respondent. In contrast with this, we obtain a degree of self-reported choice certainty after each choice task. Accordingly, self-reported choice certainty responses are likely to be correlated at the level of the individual. This is captured by  $\rho_i$ , which represents a zero mean normally distributed stochastic term with variance  $\sigma_\rho^2$ .

### 7.2.2 The choice model

A linear utility function  $U_{ijt}$  in preference space for alternative  $j \in D_{it}$  is described by (7.2). With this notation,  $D_{it}$  gives the choice set,  $Z_{ijt}$  is a set of explanatory variables with fixed effects, and  $X_{ijt}$  is a set of explanatory variables of which the impact varies across respondents. As such,  $\beta_f$  traces the impact of  $Z_{ijt}$  on the utility and  $\beta_i$  represents the vector of marginal utility parameters of individual  $i$  for specific policy attributes. Finally,  $\varepsilon_{ijt}$  represents an i.i.d. zero mean Type 1 extreme-value distribution of which the variance is normalised to  $\pi^2/6$  for identification purposes.

$$(7.2) \quad U_{ijt} = \lambda_{it} (Z_{ijt} \beta_f + X_{ijt} \beta_i) + \varepsilon_{ijt}$$

Following Hess and Stathopoulos (2011) and in line with the main stream CVM and SCE literature (e.g. Hanemann and Kristrom, 1995; Li and Mattsson, 1995), we assume that latent preference certainty affects the scale parameter  $\lambda_{it}$  by specifying  $\lambda_{it} = \exp(\tau C_{it})$ . Scale thereby follows a lognormal distribution and  $\tau$  measures the impact of the latent variable  $C_{it}$  on the scale parameter. Conditional on  $\lambda_{it}$ ,  $\beta_i$ ,  $\beta_f$  and by taking into account the proper identification requirements, the multinomial logit (MNL) choice probability of observing a choice  $y_{it}$  for alternative  $j$  is described in (7.3).<sup>92</sup>

$$(7.3) \quad P(y_{it} = j | Z, X, \beta_i, \beta_f, \lambda_{it}) = \frac{\exp(\lambda_{it} (Z_{ijt} \beta_f + X_{ijt} \beta_i))}{\sum_{k \in D_{it}} \exp(\lambda_{it} (Z_{ikt} \beta_f + X_{ikt} \beta_i))}$$

---

<sup>92</sup> Since  $C_{it}$  is modelled as a random variable we adopt the normalisation structure as proposed by Fiebig et al. (2010) and Greene and Hensher (2010). See the estimation appendix for more details.

The alternative hypothesis that preference uncertainty affects the tendency to select the SQ option is modelled by interacting  $C_{it}$  only with the alternative specific constant (ASC) by specifying  $(\tau C_{it} + \beta^{ASC}_i) \cdot ASC_{ijt}$ . The parameter  $\tau$  then measures the impact of the latent variable  $C_{it}$  on the tendency to select the SQ. By combining (7.1) and (7.3) we can trace the impact of respondent and choice task specific characteristics on preference certainty and hence also on choices.

### 7.2.3 The self-reported choice certainty model

Simultaneously,  $C_{it}$  also has an impact on the response to the task specific follow-up question  $I_{it}$ . The response format of the applied self-reported choice certainty question in this chapter is a *five-point* Likert scale ranging from “very uncertain” to “very certain”. Despite the ordered nature of these outcomes, the majority of the literature has relied on a continuous specification for the measurement model, which we label as the self-reported choice certainty model. Daly et al (2011) recently put forward the use of an ordered logit model as a more appropriate specification. In the present chapter, we rely on an ordered probit approach to explain responses  $I_{it}$  since it facilitates model estimation in a Bayesian framework (e.g. Koop, 2003). Let  $I_{it}^*$  represent a mapping of  $I_{it}$  on a continuous scale, such that a respondent will select  $I_{it}=g$  if  $I_{it}^*$  falls between thresholds  $\psi_{g-1}$  and  $\psi_g$ . (7.4) then describes the impact of  $C_{it}$  on this part of the model and adds the zero mean i.i.d. normally distributed stochastic term  $v_{it}$  with standard deviation  $\sigma_v$  to obtain an ordered probit specification. Accordingly, the probability that the respondent will indicate the degree of choice certainty  $g$  is represented by (7.5), where  $\phi$  is the standard normal density function following from the distribution imposed on  $v_{it}$ , and where  $\Phi$  is its cumulative equivalent.

$$(7.4) \quad I_{it}^* = C_{it} + v_{it}$$

$$(7.5) \quad P(I_{it} = g | C_{it}) = \int_{\psi_{g-1}}^{\psi_g} \phi\left(\frac{I_{it}^* - C_{it}}{\sigma_v}\right) dI_{it}^* = \Phi\left(\frac{\psi_g - C_{it}}{\sigma_v}\right) - \Phi\left(\frac{\psi_{g-1} - C_{it}}{\sigma_v}\right)$$

Ben-Akiva et al. (1999) and Bolduc et al. (2005) highlight that normalisation is needed to estimate integrated choice and latent variable models (or hybrid choice models), and put forward two different (but equivalent) normalisations, recently contrasted by Daly et al (2011). Here, we discuss a set of normalisations required in the self-reported choice certainty model. Suppose there are  $G$  response categories to  $I_{it}$ , then only  $G-1$  threshold



parameters, i.e.  $\psi$ 's, can be estimated. Since ordered responses are independent of level and scale, we also need to normalise the overall variance of  $I_{it}^*$  to 1. This requires that  $\sigma_v^2 = 1 - \gamma^2$ , since both  $C_{it}$  and  $I_{it}^*$  have a random component at the cross-sectional level. In addition, we impose  $\psi_g > \psi_{g-1}$  and set  $\psi_0 = -\infty$  and  $\psi_G = \infty$ .

By estimating a joint model in which latent preference certainty simultaneously affects the responses to the choice task *and* the self-reported choice certainty indicator, we learn about the structural parameters defining preference (un)certainty while simultaneously controlling for the impact of preference (un)certainty on parameter estimates in the choice model. (7.6) describes the associated likelihood function. Here,  $j$  represents the chosen alternative in choice task  $it$ , i.e.  $y_{it} = j$ , and  $I_{it}$  the responses to the choice certainty follow up question.  $y$  and  $I$  are the associated vectors containing all responses across respondents and choice tasks. For notational convenience we have dropped the conditionality on the model parameters and explanatory variables in the specification of the likelihood function.

$$(7.6) \quad L(y, I) = \prod_{i=1}^n \int_{\beta_i, \rho_i} \prod_{t=1}^T \int_{\xi_{it}} P(y_{it} = j | \beta_i, C_{it}) P(I_{it} = g | C_{it}) h(C_{it} | \Omega) d\xi_{it} f(\beta_i | \theta) d\beta_i, \rho_i$$

#### 7.2.4 Entropy as a measure of choice task complexity

The elements comprised in  $W_{it}$  in (7.1) measure the complexity of each choice task. In order to maintain sufficient degrees of freedom during estimation, we use a single measure of complexity rooted in information theory, namely the Shannon entropy measure (Shannon, 1948). Entropy  $H_{it}$  is defined in (7.7) and is increasing in the informational content of the choice task, for example, when more alternatives are added to the choice set. Swait and Adamowicz (2001) note that the entropy measure reaches its minimum 0 when an alternative has a probability of one of being chosen and its maximum of  $-\ln(1/J)$  when all alternatives are equally likely to be chosen.<sup>93</sup> Accordingly, we work with the hypothesis that preference certainty is decreasing in entropy.

$$(7.7) \quad H_{it} = - \sum_{j \in D_{it}} \hat{P}_{ijt} \ln(\hat{P}_{ijt})$$

---

<sup>93</sup> If  $P_{ijt} = 0$ , then  $0 \cdot \ln(0) = 0$ .

In our model, entropy affects preference (un)certainty and in line the scale of the utility function and the response to the follow-up question. Thereby, entropy influences choice probabilities, while it is simultaneously a function of the same choice probabilities. Including entropy as an integral part of the likelihood function would therefore introduce endogeneity. We work around this issue by approximating  $P_{ijt}$ , the choice probabilities feeding into the entropy measure, by  $\hat{P}_{ijt}$  which is based on parameters obtained from a basic utility function estimated on a similar dataset.<sup>94</sup> Since entropy is calculated using the parameters from an external MNL model, its values vary across choice tasks, but not across respondents (for an identical choice task). The latter is taken into account by  $\rho_i$ , i.e. the respondent specific random effect on preference certainty in (7.1).

### 7.2.5 Sequential modelling

In the sequential model specification, there is no longer a need to approximate the entropy measure using external data sources. Conditional on the parameters of the choice model, the researcher can derive the choice probabilities directly and calculate an observation specific entropy measure. Entropy is still used as an explanatory variable of latent preference certainty, which is fed into the self-reported choice certainty model. In this specification, the error terms in the latent variable model  $\zeta_{it}$  and the self-reported choice certainty model  $v_{it}$  are perfectly confounded, since we only have a single indicator variable in our empirical model. For identification purposes we therefore remove  $\zeta_{it}$  from (7.1). Consequently, we can no longer estimate  $\gamma$ , but which simplifies the likelihood function in (7.8). Note that the sequential specification will still be estimated in an integrated framework, where the link between the choice and self-reported choice certainty model is made by making  $C_{it}$  dependent on the choice probabilities. However, the latent variable  $C_{it}$  is no longer used as an explanatory variable in the choice model.

$$(7.8) \quad L(y_{it}, I_{it}) = \prod_{i=1}^n \int \prod_{t=1}^T P(y_{it} = j | \beta_i) P(I_{it} = g | C_{it}) h(C_{it} | \Omega, \beta_i) f(\beta_i | \theta) d\beta_i, \rho_i$$

---

<sup>94</sup> We use the HSB and LSB subsamples of the survey for this purpose, but in general one can also use results from an earlier pre-test stage.

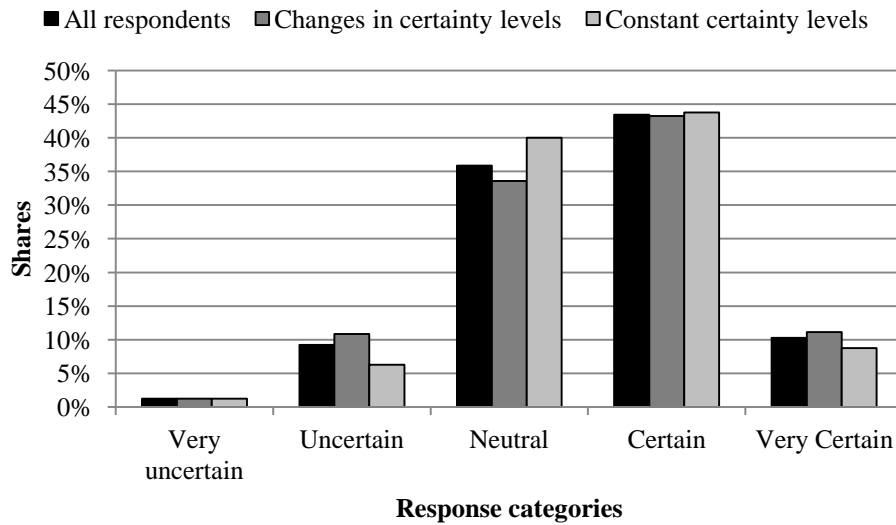
### 7.3 The dataset

The data used in this chapter are based on the CCF sample, as described in Chapter 4. The sample is characterized by adding a choice certainty follow-up question directly after each choice task. Respondents are able to express their level of preference certainty on a 5 point Likert scale ranging from “*very uncertain*” to “*very certain*”. Figure 7.2 presents an overview of the 1,792 responses to the follow-up questions for the 224 respondents in the sample. Despite their lack of experience with floods, most respondents report a relatively high level of choice certainty.<sup>95</sup> Choice certainty responses concentrate around the “neutral” and “certain” levels, respectively 36% and 43%. In approximately 10% of the observations respondents reported that they were “very certain” and in another 10% of the observations they were “uncertain” about their responses. Only in 1% of the observations, respondents did select the “very uncertain” response option. Closer investigation of dynamics in these response patterns over the choice sequence reveals that 80 respondents (36%) did not alter their certainty responses. Figure 7.2 reveals that their distribution over the response categories provides a relatively close match to responses provided by respondents that did change their responses during the choice sequence. The response pattern of these 80 respondents could be mainly related to the attitudes towards the choice tasks in general, i.e. choice heuristics and lack of engagement or disinterest in the topic. Since these respondents do not provide any information regarding dynamics in choice certainty over the choice sequence, nor about the impact of choice task complexity on choice certainty, we decided to exclude them from further analysis in this subsample. As such, a final sample of 1,152 observations from 144 respondents was retained.

---

<sup>95</sup> Only 28 respondents had experienced a flood during their lifetime.

**Figure 7.2: Distribution of responses to the choice certainty follow-up question across groups of respondents**



These 144 respondents reveal substantial dynamics in self-reported choice certainty levels over the choice sequence in Table 7.1. Three respondents changed their certainty response in almost every choice task during the choice sequence (seven out of eight times) and nineteen did this only once. As revealed by columns three and four in Table 7.1, clear learning and fatigue effects cannot be identified. Most respondents revealed both increases and decreases in choice certainty during the choice sequence. Only one respondent experienced five times an increase in choice certainty. This suggests that choice certainty is perhaps more likely to be affected by the complexity of the choice task, rather than the position in the choice sequence. This picture is further confirmed by the final three columns of Table 7.1, where the number of (directional) changes in self-reported choice certainty varies over the choice sequence, but again no systematic pattern can be identified. Average choice certainty peaks in choice task 4 at 3.64 (closest to the ‘certain’ level) and shows a gradual decline afterwards to 3.43 in the ninth choice task, but the  $\chi^2$ -test fails to reject the null hypothesis of identical distribution of response patterns over the choice sequence at the 5% significance level.

**Table 7.1: Changes in self-reported choice certainty within respondents and over the choice sequence**

# of response changes	Within respondent changes in stated certainty levels: <sup>a</sup>			Choice Task <sup>c</sup>	Average Stated Certainty	Number of respondents showing a: <sup>b</sup>		
	All changes	Positive changes	Negative changes			Increase in certainty level	Decrease in certainty level	Constant certainty level
0x	0	11	9	2	3.53	0	0	0
1x	19	63	60	3	3.50	34	31	79
2x	39	45	50	4	3.64	23	39	82
3x	21	24	21	5	3.57	37	33	74
4x	31	0	4	6	3.53	37	32	75
5x	22	1	0	7	3.50	35	36	73
6x	9	0	0	8	3.48	40	31	73
7x	3	0	0	9	3.43	33	28	83

<sup>a</sup>- Overall changes in stated certainty over the choice sequence. Vertical summation gives 144 respondents

<sup>b</sup>- Changes in stated certainty relative to previous choice task. Horizontal summation gives 144 respondents

<sup>c</sup>- The first (instructional) choice task is not included in the analysis

## 7.4 Results

We now proceed with a discussion of the estimation results. Section 7.4.1 presents the results from our simultaneous model specification where we contrast the alternative hypotheses on whether preference (un)certainty affects the scale parameter or the tendency to select the status quo option in the choice model. Section 7.4.2 compares the simultaneous model to the sequential model. Finally, in Section 7.4.3 we trace the impact of preference (un)certainty on WTP estimates across the various model specifications, while section 7.4.4 provides some further analysis of self-reported choice certainty responses.

### 7.4.1. Simultaneous modelling

Based on the discussion of self-reported choice certainty responses in Table 7.1, we do not expect to find strong learning and fatigue effects over the choice sequence. For this reason, we present the results for our simultaneous latent variable model using only our measure of choice task complexity, i.e. entropy, as an explanatory variable of choice certainty. Table 7.2 presents the results for three alternative models estimated using Bayesian methods, which facilitates estimation given the complex and highly non-linear likelihood function (e.g. Daziano and Bolduc, 2011).<sup>96</sup> Model 1 combines an independently estimated attributes-only choice model not affected by preference certainty, with an independently estimated random effects ordered probit model, where entropy  $H_{it}$  and  $\rho_i$  are used as explanatory variables for

<sup>96</sup> Bayesian estimations are conducted in Matlab and the Gibbs Sampler is specified in the estimation appendix.

self-reported choice certainty. Accordingly,  $\tau$  and  $\gamma$  are not estimated in this specification. In Model 2, latent preference certainty is explained by entropy  $H_{it}$  and  $\rho_i$ , and it affects the utility function through the scale parameter. The impact on the self-reported choice certainty model is in line with (7.4), and in this specification  $\tau$  and  $\gamma$  are estimated (and identified) due to the inclusion of the latent variable in the choice model and the self-reported choice certainty model. Model 3 is comparable to Model 2, but preference certainty affects the choice model by interacting  $C_{it}$  with the alternative specific constant (ASC), as opposed to the scale parameter. Hence, Models 2 and 3 can be used to test whether increasing preference certainty induces less random decision making or reduces the tendency to select the SQ option. In the choice model, all policy attributes, except evacuation, are assigned a lognormal distribution ensuring a positive (negative for the cost attribute) impact on utility. Both the ASC and evacuation attribute follow a normal distribution.<sup>97</sup>

---

<sup>97</sup> Model results are most stable using a normal distribution for the evacuation attribute, but our overall conclusions also apply to other specifications we tested, including one with a fixed parameter for evacuation.

**Table 7.2: Bayesian model results of the full latent variable model**

		Independent models		Latent variable model impact on SCALE		Latent variable model impact on ASC	
		(1)		(2)		(3)	
	Variable	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.	Post. Mean	Post. St. Dev.
<i>Choice model</i>	ASC	1.84	0.26	2.64	0.50	2.41	0.35
	PROB	-1.82	0.21	-1.50	0.22	-1.82	0.21
	COMP	-2.09	0.18	-1.85	0.21	-2.11	0.19
	EVAC	0.46	0.13	0.60	0.17	0.45	0.13
	COST	-1.66	0.12	-1.45	0.16	-1.71	0.13
	Std. ASC	1.40	0.29	1.88	0.46	1.59	0.32
	Std. PROB	0.97	0.20	0.94	0.20	0.95	0.20
	Std. COMP	1.04	0.18	1.06	0.19	1.03	0.19
	Std. EVAC	0.81	0.18	0.84	0.24	0.79	0.18
	Std. COST	0.87	0.11	0.99	0.13	0.91	0.12
	$\tau$	-		0.48	0.11	0.75	0.20
<i>Preference Certainty</i>	Entropy	-1.38	0.25	-1.37	0.25	-1.36	0.25
	$\sigma_\rho^2$ (variance of $\rho$ )	0.92	0.14	0.91	0.14	0.91	0.14
	$\gamma$	-		0.71	0.01	0.71	0.01
<i>Stated Choice Certainty Model</i>	$\psi_1$	-3.65	0.19	-3.64	0.19	-3.64	0.19
	$\psi_2$	-2.07	0.13	-2.06	0.13	-2.06	0.13
	$\psi_3$	-0.55	0.12	-0.55	0.12	-0.54	0.12
	$\psi_4$	1.26	0.13	1.25	0.12	1.26	0.12
<i>Model Fit</i>	LL* Choice	-905.17		-905.93		-904.97	
	LL Stated	-1274.49		-1261.71		-1270.86	
	LL Overall	-2179.66		-2167.64		-2175.82	

\* All LL values are evaluated at the posterior mean, not by using the Bayesian concept of marginal likelihood

Model 3 suggests that controlling for preference certainty in the choice model only has a limited impact on the estimated preference parameters. All parameters in the choice model, except the ASC, are stable in comparison to Model 1. The ASC increases, which simultaneously comes at the cost of an increase in its posterior standard deviation. In Model 2, the posterior means show more variation relative to Model 1, but again come at the cost of increased posterior standard deviations. In fact, the log-likelihood values (evaluated at the posterior means) for the choice model part highlight that controlling for latent preference certainty does not offer a better explanation for the observed choices. The lack of explanatory power of preference certainty in the choice model is supported by a set of independently

estimated choice models (not presented here) in which entropy is directly interacted with either the scale parameter or the ASC. In both models, entropy is not found to have an impact on utility.

Nevertheless,  $\tau$  is positive in Models 2 and 3. In the case of Model 2, this confirms our hypothesis that respondents with higher levels of preference certainty exhibit a higher scale parameter and thereby make less random decision than uncertain respondents. In the case of Model 3, we also note that increased preference certainty leads to respondents being less inclined to select the Status Quo option. In both models, preference certainty is decreasing in the entropy measure implying that increased complexity of the choice task reduces preference certainty. Since preference certainty barely has an impact on the fit of the choice model, the inclusion of  $\tau$  (i.e. making the link between the different model components) contributes to explaining the responses to the self-reported choice certainty questions. The latter would imply that the implicit preference certainty information contained in the choice model is in line with the explicit measure of preference certainty, i.e. the level of self-reported choice certainty. Model 2 reveals a substantial improvement in model fit for the self-reported choice certainty component when compared to Models 1 and 3. However, the parameter estimates for the self-reported choice certainty model are remarkably stable across model specifications. The threshold parameters and coefficient on the entropy measure (including their standard errors) are virtually the same, which should therefore not result in such a large improvement in model fit in the self-reported choice certainty model.

We suspect that these results are driven by the lack of impact of latent preference certainty on the choice model. First, the lack of impact explains the similarity in parameter estimates for the self-reported choice certainty model in Models 2 and 3 compared to Model 1. Second, the low posterior standard deviation on the parameter estimate for  $\gamma$  suggests an empirical identification issue. Since we do not learn about latent preference certainty through the choice model, the observation specific error term included in  $C_{it}$  cannot be separated from the error term in the probit model. Although this is not very problematic for estimation in a Bayesian framework, it does have an impact on the simulated likelihood values. The error term in  $C_{it}$  requires simulation at the observation and individual level, which may have introduced additional simulation errors. To this end, we present a set of results in Table 7.3 for the same set of models, but without including  $\zeta_{it}$  in the specification of latent preference certainty. The likelihood function now only includes a single integral at the individual level. Entropy and  $\rho_i$  still affect both the choice and the self-reported choice certainty model. Given the reduced complexity of the likelihood function, these models are estimated in a classical



framework using 1,000 Halton draws to simulate the likelihood function, and using the SQP algorithm reduces the probability of being trapped in a local optimum. Estimations are conducted in Ox (Doornik, 2007).

**Table 7.3: Results from a classical estimation simultaneous impact of entropy and random effect**

		Independent models		Latent variable model impact on SCALE		Latent variable model impact on ASC	
		(4)		(5)		(6)	
	Variable	Coeff.	St. error	Coeff.	St. error	Coeff.	St. error
<i>Choice</i>	ASC	1.85	0.27	2.23	0.37	2.00	0.30
<i>Model</i>	PROB	-1.78	0.20	-1.60	0.21	-1.78	0.20
	COMP	-2.03	0.16	-1.88	0.22	-2.10	0.20
	EVAC	0.47	0.13	0.58	0.16	0.47	0.13
	COST	-1.68	0.12	-1.47	0.14	-1.63	0.12
	Std. ASC	1.46	0.31	1.50	0.30	1.38	0.28
	Std. PROB	0.94	0.21	0.90	0.19	0.91	0.19
	Std. COMP	0.91	0.10	1.17	0.21	1.27	0.18
	Std. EVAC	0.76	0.19	0.78	0.26	0.72	0.20
	Std. COST	0.86	0.10	0.83	0.08	0.86	0.10
	$\tau$			0.31	0.10	0.27	0.18
<i>Preference</i>	Entropy	-1.37	0.25	-1.38	0.25	-1.37	0.25
<i>certainty</i>	$\sigma_p$	0.94	0.07	0.93	0.07	0.92	0.06
<i>Stated</i>	$\psi_1$	-3.63	0.19	-3.66	0.19	-3.66	0.19
<i>Choice</i>	$\psi_2$	-2.06	0.13	-2.08	0.13	-2.07	0.13
<i>Certainty</i>	$\psi_3$	-0.55	0.12	-0.57	0.12	-0.56	0.12
<i>Model</i>	$\psi_4$	1.26	0.12	1.24	0.12	1.24	0.12
<i>Model</i>	LL Choice	-904.67		-906.65		-906.90	
<i>Fit</i>	LL Stated	-1274.47		-1269.37		-1272.96	
	LL Total	-2179.14		-2176.02		-2179.86	

Conclusions regarding the impact of preference certainty on the choice model remain the same, since we do not observe an improvement in model fit for the choice model part in Models 5 and 6 relative to Model 4. The parameter  $\tau$  still has a positive and significant impact in Model 5, but no longer in 6. An improvement in model fit for the self-reported choice certainty model is still observed for Model 5 and only to a limited extent for Model 6. The improvement in fit for Model 5 is mainly the result of slight variations in the threshold

parameters and in the estimated standard deviation of  $\rho_i$ . Finally, the overall model fit reveals that only a minor improvement in model fit can be achieved by controlling for implicit and explicit preference certainty in a simultaneous fashion.

Based on these results, we conclude that the additional information obtained by asking a post-decisional choice certainty question does not help us in better explaining observed choices. Our results do provide more support for the hypothesis that latent preference certainty affects the utility function through the scale parameter, rather than the tendency to select the status quo option. However, the latter conclusion is only indicative due to the lack of impact of preference certainty on the fit of the choice model in both specifications. These conclusions are not affected by controlling for additional variables affecting latent preference certainty. Finally, our Bayesian model pointed out that empirical identification issues are likely to arise by including an integral at the level of observations. This is in line with previous results by Hess and Train (2011).

#### **7.4.2 Sequential modelling**

Section 7.4.1 suggests that there are hardly any benefits in simultaneously analyzing the choice model and the self-reported choice certainty model in the present case study. If any benefits arise, our results point towards an impact of the choice model in helping to provide a better explanation for the self-reported choice certainty responses. To this end, we alter our model set-up and estimate the choice and self-reported choice certainty models in a sequential fashion using a FIML approach. The parameters of the utility function are used to obtain individual and choice task specific choice probabilities, which are used to derive the entropy measure. Entropy is then fed into the self-reported choice certainty model after which we optimize the joint likelihood function. Results for three alternative specifications are reported in Table 7.4.

Model 7 presents the results for a two-stage model where the choice model is estimated independently in the first stage and its mean parameter estimates are used in the second stage to simulate the individual and choice task specific entropy measure in the self-reported choice certainty model. Model 8 estimates these models in an integrated fashion by calculating entropy already within the simulated likelihood function. Increases in efficiency are expected for this specification. Model 9 is included to contrast the results against the

results in Table 7.3.<sup>98</sup> We do this by feeding a measure of expected entropy into the self-reported choice certainty model during the integrated optimization. This measure of expected entropy should be similar across respondents in our simulation procedure, such that we can contrast the simultaneous and sequential modelling approach.

Model 7 shows an improvement in overall model fit compared to all previous models. The FIML approach applied in Model 8 reveals an additional increase in model fit. However, there is only marginal benefit in estimating the two models jointly. As expected, standard errors decrease in Model 8, which mainly benefits the choice model, but marginal WTP estimates are not significantly different across models 7 and 8 (results reported in Section 7.4.3). Part of the improvement in overall fit relative to Model 4 can be a result of calculating entropy within the integral of the likelihood function, making it a random variable on its own and hence offering additional flexibility. To test if sequential modelling in itself improves the model fit we feed a measure of expected entropy, which should be similar across respondents in our simulation procedure, into the self-reported choice certainty model in Model 9. The results for Model 9 display a decrease in overall model fit also relative to Table 7.3. Although this supports our notion that respondents decide simultaneously about their choice and the response to the follow-up question, the arguments for using a simultaneous modelling approach are not very strong. Both the simultaneous and sequential models reveal only minor variations in the fit of the choice model, while the sequential approach reveals that self-reported choice certainty responses can best be explained by varying entropy across respondents and choice tasks. Given the latter, the researcher seems just as well off by estimating the choice model and the self-reported choice certainty model in an independent two-stage fashion. Not only does this reduce the complexity of estimation, but it also alleviates the requirement of approximating entropy using an external dataset.

---

<sup>98</sup> Our previous measure of entropy only varied across choice tasks due to the use of an MNL model in approximating entropy.

**Table 7.4: Results for a set of sequential models**

		Sequential - separate models (7)		Sequential – FIML (8)		Sequential – expected entropy (9)	
Variable		Coeff.	St. error	Coeff.	St. error	Coeff.	St. error
<i>Choice</i>	ASC	1.85	0.27	1.73	0.20	1.79	0.24
<i>Model</i>	PROB	-1.78	0.20	-2.29	0.22	-1.97	0.24
	COMP	-2.03	0.16	-2.05	0.13	-2.00	0.17
	EVAC	0.47	0.13	0.52	0.11	0.57	0.13
	COST	-1.68	0.12	-1.67	0.10	-1.62	0.11
	Std. ASC	1.46	0.31	1.51	0.19	1.19	0.28
	Std. PROB	0.94	0.21	1.37	0.15	1.01	0.20
	Std. COMP	0.91	0.10	-0.87	0.07	0.99	0.35
	Std. EVAC	0.76	0.19	0.80	0.13	0.73	0.19
	Std. COST	0.86	0.10	0.75	0.05	0.74	0.08
<i>Preference</i>	Entropy	-2.39	0.25	-2.67	0.28	-1.10	0.43
<i>Certainty</i>	$\sigma_p$	0.93	0.07	0.95	0.07	0.91	0.07
<i>Stated</i>	$\psi_1$	-5.03	0.28	-5.34	0.33	-3.83	0.32
<i>Choice</i>	$\psi_2$	-3.28	0.22	-3.55	0.27	-2.27	0.29
<i>Certainty</i>	$\psi_3$	-1.60	0.18	-1.83	0.23	-0.78	0.28
<i>Model</i>	$\psi_4$	0.45	0.15	0.27	0.19	1.00	0.28
<i>Model</i>	LL Choice	-904.67		-908.804		-907.26	
<i>Fit</i>	LL Stated	-1239.33		-1229.36		-1286.45	
	LL Total	-2144.00		-2138.16		-2193.70	

### 7.4.3 WTP estimates

Table 7.5 provides an overview of the model specific marginal WTP estimates for the three non-price policy attributes in the stated choice experiment. Note that marginal WTP estimates for models 4 and 7 are equivalent and therefore only reported once, since the same choice model is estimated independently from the self-reported choice certainty model. WTP distributions for the ratio of two lognormal distributions follow a log-normal distribution, such that for the probability and compensation attributes, the WTP distribution can be obtained analytically. This is not the case for the WTP for evacuation time, where the numerator follows a normal distribution. In this case, the WTP is simulated by taking 10,000 draws from the estimated distributions, and one-sided complete combinatorial approach is applied to test for differences in marginal WTP across specifications (Poe et al., 2005).

Median WTP estimates are relatively stable across model 4, 5 and 6. This result is not surprising since we barely find an impact of our simultaneous model specification on the fit of the choice model. Mean marginal WTP estimates are a slightly more sensitive across these three models, in particular for the probability and compensation attributes. Even though parameter estimates do not vary much across models, the use of lognormal distributions amplifies these differences, which directly translates into variations in mean WTP. For example, the increase in mean marginal WTP for the compensation attribute in models 5 and 6 is a result of the increase in the standard deviation of the underlying normal distribution for the compensation attribute. The effect is less pronounced for the evacuation attribute, which is based on a normal distribution. Overall, statistical tests and simulated confidence intervals reveal no significant differences in the distribution of marginal WTP over the population of interest across model specifications. Median WTP estimates are, however, more variable in Models 8 and 9 relative to Model 7 for the probability and evacuation attributes. The decrease in model fit for the choice model reveals that the responses to the choice tasks help in explaining responses to the self-reported choice certainty questions, and not the other way around. The FIML approach in these sequential models also adjusts the parameters of the choice model to optimize the overall likelihood function. As such, the reliability of marginal WTP parameters may decrease. In our case, the impact seems to be minor. Thereby, the lack of impact of preference (un)certainly on marginal WTP estimates is a positive message for policy makers interested in obtaining welfare estimates for new policy plans.

**Table 7.5: Marginal WTP estimates for the non-cost policy attributes**

	PROB			COMP			EVAC		
	Mean	Median	Std. dev.	Mean	Median	Std. dev.	Mean	Median	Std. dev.
Model 4,7	20.37	9.05	28.65	1.07	0.70	1.21	3.71	1.92	9.13
Model 5	19.06	8.78	26.73	1.32	0.66	2.24	3.43	1.95	7.19
Model 6	18.38	8.61	25.91	1.40	0.63	2.84	3.45	1.84	8.13
Model 8	19.90	5.38	67.23	1.00	0.68	1.10	3.53	2.19	7.71
Model 9	15.55	7.05	27.58	1.12	0.68	1.47	3.78	2.37	6.92

Units:

All marginal WTP estimates are in € per household per year

Probability - Increase in the denominator of probability by 1,000 years; i.e. from (1/4,000 to 1/5,000)

Compensation - Additional percentage point of compensation, for example from 35% to 36%

Evacuation - Additional hour of available evacuation time

#### 7.4.4 Explaining self-reported choice certainty

Based on the conclusions from the previous subsections, we built upon Model 7 and add additional control variables to provide a better explanation of self-reported choice certainty. We expect self-reported choice certainty to be related to respondent characteristics and the complexity of the choice task itself. Starting with the latter, the 144 respondents answered the three blocks of the design respectively 38, 56 and 50 times. That is, choice cards 1-8 from the design are answered 38 times, cards 9-16 56 times and cards 17-24 50 times. The Mann-Whitney test reveals that there is a significant difference in self-reported choice certainty between responses to blocks 2 and 3 at the 1% level. Respondents in the third block report significantly lower certainty levels. A more detailed analysis at the choice card specific levels reveals that block 2 includes two relatively easy choice tasks for which respondents report significantly higher certainty levels in comparison to all other choice cards in the design. In Figure 7.3 we plot the relationship between average self-reported choice certainty per choice card on the x-axis and the card specific entropy measure on the y-axis. The figure clearly reveals that the informational content of a choice card is comparable between blocks 1 and 3. Indeed, block 2 has some choice cards with low informational content, i.e. low entropy, which is also the case for card 3 in block 1. Nevertheless, within each block of choice cards, a negative correlation between entropy and average choice certainty is observed.

Figure 7.3: Scatter plot of choice card specific entropy vs. average self-reported choice certainty

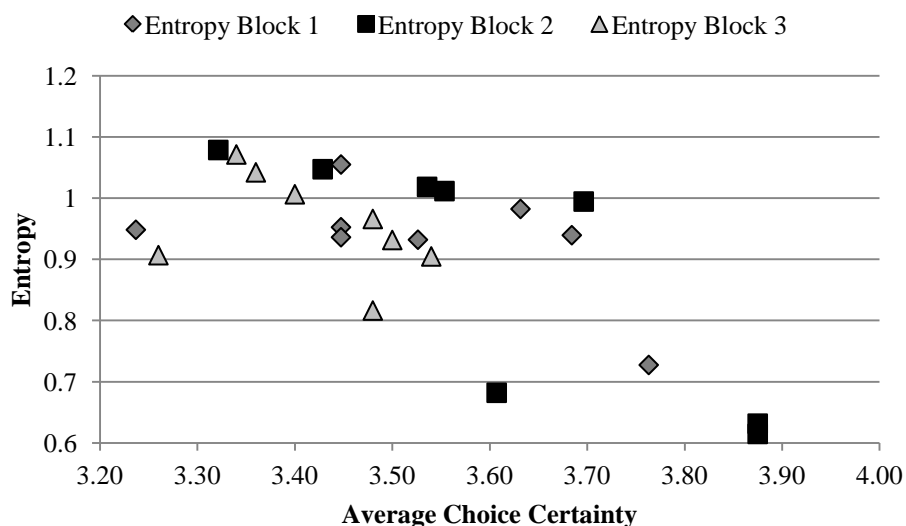


Table 7.6 reports the results for the extended latent variable model with additional control variables for self-reported choice certainty.<sup>99</sup> It captures dynamics in self-reported choice certainty over the choice sequence using a linear effect.<sup>100</sup> As expected, self-reported choice certainty is decreasing in the entropy measure and there is substantial heterogeneity in certainty across respondents, but the standard deviation is smaller due to controlling for various respondent characteristics. The block of the design assigned to a respondent does not have a significant impact on choice certainty itself. The sign of the linear effect accounting for the position in the choice task shows that certainty gradually decreases over the choice sequence suggesting possible fatigue effects. Like Olsen et al. (2011), we barely find significant impacts of respondent characteristics on self-reported choice certainty due to the inclusion of  $\rho_i$ . Respondents who believe that the proposed policies are credible tend to report higher levels of choice certainty, a finding also reported in Brouwer et al. (2010). Both Olsen et al. (2011) and Brouwer et al. (2010) find that males state that they are more certain than females. Our data do not support this notion, but find that highly educated respondents are less certain at the 5% significance level.

---

<sup>99</sup> Following Swait and Adamowicz (2001), we also estimated various models controlling for cumulative entropy, but the results were not significant.

<sup>100</sup> We also tested for a quadratic effect, but again this effect was not significant.

**Table 7.6: Explaining self-reported choice certainty (random effects ordered probit model)**

	(10)		
	Coefficient	St. Error	Significance
<i>Choice task characteristics:</i>			
Entropy	-2.39	0.25	***
Block 1	-0.30	0.20	NS
Block 3	-0.27	0.18	NS
<i>Sequence dynamics:</i>			
Card number	-0.03	0.02	**
<i>Respondent characteristics:</i>			
Gender	0.18	0.16	NS
Highly educated	-0.36	0.16	**
Flood experience	-0.13	0.26	NS
Flood expectance	0.24	0.22	NS
Credibility	0.28	0.09	***
<i>Thresholds:</i>			
U1	-5.42	0.33	***
U2	-3.66	0.27	***
U3	-1.97	0.24	***
U4	0.08	0.21	
<i>Random effect (<math>\rho_i</math>):</i>			
$\sigma_\rho$	0.88	0.07	***
<i>Model Fit:</i>			
LL Stated	-1228.49		
LL Overall	-2133.16		

## 7.5 Summary

In this chapter we have put forward a modelling approach towards accounting for preference uncertainty in stated choice experiments. The proposed model simultaneously takes into account the implicit and explicit modelling strategies of preference uncertainty adopted in, for example, Brouwer et al. (2010) and Lundhede et al. (2009). By treating preference certainty as a latent variable which simultaneously affects the stated choices and responses to the follow-up choice certainty questions, our model works around endogeneity issues and measurement error that are likely to arise. By treating the self-reported choice certainty responses in a unified framework with the responses to the choice task, this chapter addressed research question three and its related hypothesis formulated in Chapter 4.

We have contrasted two sets of hypotheses. First, we tested whether latent preference certainty is more likely to affect the scale of the utility function, i.e. the degree of random decision making, or the tendency to select the status quo option, i.e. inducing non-trading



behaviour. Although the impact of preference certainty on the choice model is limited and does not affect marginal willingness-to-pay estimates, we find evidence of both effects with somewhat more support for an impact on the scale of the utility function. In both specifications, we find that preference certainty increases when respondents are presented with easier choice tasks. Choice task complexity was quantified through the Shannon (1948) entropy measure. By not revealing an impact of certainty on willingness-to-pay estimates, our model results provide a positive message to policy makers interested in using welfare estimates from stated choice approaches similar to those investigated here.

Second, we contrasted our simultaneous model approach against alternative model specifications, including an independent and sequential modelling of both response formats. An improvement in model fit is not found by our simultaneous modelling approach. Hence, responses to the choice tasks and the follow-up questions can best be estimated in an independent or two stage fashion. An advantage of this independence is that measures of choice task complexity, like the entropy measure, or utility differences can be obtained from the choice model and then used in explaining responses to the follow-up questions without excessive concern about endogeneity. Since results by Brouwer et al. (2010) and Olsen et al. (2011) rely on this sequential modelling approach, their results are thus not put in doubt. Our work, however, does not provide any conclusions on the extent to which the responses to the self-reported choice certainty questions can help in explaining the choice sequence as proposed by Hensher and Rose (2011) and Lundhede et al. (2009) – notwithstanding the endogeneity issues arising in such work. Indeed, we only find indications that the implicit modelling of preference certainty in the choice model can help in explaining the responses to the latter, but not the other way around. All our estimations indicate a reduction in model fit for the choice model when controlling for both types of responses in either a simultaneous or sequential fashion. This is not surprising, since the follow-up question is placed after the choice task. More work is needed to establish whether similar results are obtained with other datasets.

In perspective of the research question and the related hypothesis, we do acknowledge that self-reported choice certainty measures may form a useful tool in explaining the degree of preference uncertainty, and tracing its impact on response patterns. In particular, because by linking both responses preference uncertainty can be separated from other sources of heterogeneity in the choice model. We do not find any evidence of the self-reported choice certainty responses not being in line with the response patterns in the choice model. However, since we primarily find that responses to the choice tasks help in explaining responses to the

follow-up question in our dataset, we cannot conclude that such follow-up questions are a useful tool for improving the validity and reliability of willingness-to-pay estimates.

## Appendix 7.A – Bayesian estimation of the latent variable model

In this estimation appendix we describe how the simultaneous latent variable model is estimated in a Bayesian framework. Daziano and Bolduc (2011) highlight that the rationale for using Bayesian estimation methods is given by the complexity of the likelihood function. Classical estimation methods are more likely to be caught in a local optimum, due to the non-linearity of the likelihood function. Since Bayesian methods do not rely on optimization, but simulate from a set of conditional posterior densities, these estimation issues are less likely to occur. Section 7.A.1 describes the set of conditional posteriors for our model specification. Section 7.A.2 discusses the setup of the Gibbs Sampler and Section 7.A.3 finalizes by showing the applied values for the prior parameters in the Gibbs Sampler.

### 7.A.1 - The conditional posterior distributions

Latent preference certainty  $C_{it}$  affects the choice model through the scale parameter  $\lambda_{it}$  in the following form  $\lambda_{it} = \exp(\tau C_{it})$ . Our hypothesis is that uncertain respondents display a lower scale on their utility function than certain respondents, i.e. they make more random decisions. Accordingly, we expect  $\tau$  to be positive. The impact of  $\tau$  on choice probabilities is denoted below, where  $-\frac{(\tau\gamma)^2}{2}$  is included for normalisation purposes in line with Fiebig et al. (2010) and Greene and Hensher (2010). The role of  $\gamma$  was discussed in Section 7.2.1 and represents the standard deviation on latent preference certainty.

$$P(y_{it} = j | \beta_i, \beta_f, C_{it}, \tau, \gamma) = \frac{\exp\left(V_{ijt} \cdot \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}{\sum_{k \in D_{it}} \exp\left(V_{ikt} \cdot \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}$$

Bayesian estimation of the parameters in the choice model is done conditional on the value of the (augmented) preference certainty variable  $C_{it}$ . As such, the choice model reduces to a standard Bayesian multinomial choice model with data augmentation as described in Train (2009, Chapter 12). By treating the latent variable  $C_{it}$  and preference parameters  $\beta_i$  as known parameters, data augmentation provides a convenient way to work around the integrals in the likelihood function and thereby provides significant reductions in estimation time. The parameters only appearing in the choice model are  $\beta_f$ ,  $\beta_i$ , and  $\tau$ . The selected

(log)normal mixing densities for  $\beta_i$  imply that each random parameter has an assigned mean  $\mu$  and variance  $\sigma^2$  as its hyper-parameters. These are assigned respectively a normal and an inverse gamma prior. Koop (2003) and Train (2009) show that analytical posteriors can be derived for these hyper-parameters of respectively a normal and inverse-gamma form.

The conditional posterior on  $\mu$  is of normal form with posterior mean  $\mu_1$  and variance  $\sigma_1^2$ . These parameters are defined by  $\sigma_1^2 = \left( \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$  and  $\mu_1 = \sigma_1^2 \left( \sum_{i=1}^n \frac{\omega_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right)$ , where the subscript 0 denotes prior knowledge. Prior values imposed in this chapter are declared in Appendix 7.A.3. The conditional posterior on  $\sigma^2$  follows an inverse gamma distribution with posterior shape parameter  $\eta_1 = \eta_0 + \sum_{i=1}^n \frac{(\omega_i - \mu)^2}{2}$  and scale parameter  $\kappa_1 = \kappa_0 + \frac{n}{2}$ . These posteriors also apply to the hyper-parameters of the lognormal coefficients in  $\beta_i$ .

The vector of fixed coefficients  $\mathcal{G}$ , comprising both  $\beta_f$  and  $\tau$ , affects the likelihood function directly and therefore requires a Metropolis-Hastings algorithm to generate draws from the joint conditional posterior distribution (see Train 2009, Chapter 9 and 12). To this end we assign a multivariate normal prior to the vector of fixed parameters  $\mathcal{G}$  with prior mean  $\mathcal{G}_0$  and covariance matrix  $V_0$  and apply a random walk chain to generate candidate draws from its conditional posterior density.

$$\begin{aligned}
p(\mathcal{G} | y, \beta, C) &\propto p(\mathcal{G}) \prod_{i=1}^n \prod_{t=1}^T P(y_{it} = j | \mathcal{G}, \beta, C) \\
&\propto \exp\left(-\frac{1}{2}(\mathcal{G} - \mathcal{G}_0)'(V_0)^{-1}(\mathcal{G} - \mathcal{G}_0)\right) \prod_{i=1}^n \prod_{t=1}^T \frac{\exp\left(\left(Z_{ijt}\beta_f + X_{ijt}\beta_i\right) \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}{\sum_{k=1}^J \exp\left(\left(Z_{ikt}\beta_f + X_{ikt}\beta_i\right) \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}
\end{aligned}$$

Similarly, the posterior on the augmented variables  $\beta_i$  requires a M-H algorithm. The prior in this case is the mixing density  $f(\mu, \sigma^2)$  and the parameters of individual  $i$  only affect the choices by that individual. Hence, the individual specific conditional posterior for  $\beta_i$  is

described below. In case of a lognormal distribution  $\beta_i$  changes into  $\exp(\beta_i)$  in the MNL probability part.

$$p(\beta_i | y, C_i, \beta_f, \tau) \propto p(\beta_i | \mu, \sigma^2) \prod_{i=1}^T P(y_{it} = j | \beta_f, \beta_i, C_i, \tau) \\ \propto \exp\left(-\frac{(\beta_i - \mu)^2}{2\sigma^2}\right) \prod_{i=1}^T \frac{\exp\left(\left(Z_{ijt}\beta_f + X_{ijt}\beta_i\right) \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}{\sum_{k \in D_{it}} \exp\left(\left(Z_{ikt}\beta_f + X_{ikt}\beta_i\right) \exp\left(\tau C_{it} - \frac{(\tau\gamma)^2}{2}\right)\right)}$$

*The self-reported choice certainty model:*

The self-reported choice certainty model captures the impact of latent preference certainty  $C_{it}$  on the response to the self-reported choice certainty question after each choice task  $I_{it}$ . Due to the ordered character of the response format, we apply an ordered probit model for the measurement model. The required normalisations are discussed in Section 7.2.3. Each of the threshold parameters  $\psi_g$  is assigned a non-informative normal prior resulting in the following conditional posterior distribution. Note that  $\psi_g > \psi_{g-1}$  and  $\psi_0 = -\infty$  and  $\psi_G = \infty$ , which is monitored in the required M-H algorithm.

$$p(\psi_g | C) \propto \exp\left(-\frac{(\psi_g - \psi_{g0})^2}{2\sigma_{g0}^2}\right) \prod_{g=1}^G \Phi\left(\frac{\psi_g - C_{it}}{\sqrt{1 - \gamma^2}}\right) - \Phi\left(\frac{\psi_{g-1} - C_{it}}{\sqrt{1 - \gamma^2}}\right)$$

*The latent variable ‘‘Preference Certainty’’:*

The latent variable  $C_{it}$  was already present in the choice model and in the measurement model. The functional form defines that  $C_{it}$  on itself follows a normal distribution with mean  $\delta R_i + \zeta W_{it} + \rho_i$  and variance  $\gamma^2$ . Like the hyper-parameters for the mixing densities in the choice model, we can assign a conjugate multivariate normal prior to the vector  $(\delta, \zeta)$ , resulting in a multivariate normal conditional posterior with its variance defined by

$$V_{\delta\zeta 1} = \left( (V_{\delta\zeta 0})^{-1} + \frac{[R, W]': [R, W]}{\gamma^2} \right)^{-1}, \text{ where } R \text{ and } W \text{ are transformations of } R_i \text{ and } W_{it} \text{ to get a}$$

matrix of size observation by  $ks$ , where  $ks$  is the number of structural variables characterizing the latent variable. The mean of the conditional posterior is defined by

$\mu_{\delta_{\zeta 1}} = (V_{\delta_{\zeta 1}})^{-1} \left( \frac{\mu_{\delta_{\zeta 0}} + [R, W]' \cdot (C_{it} - \rho_i)}{\gamma^2} \right)$ . A conditional posterior of similar structure can be

derived for  $\rho_i$ , with variance  $\sigma_{\rho_1}^2 = \left( \frac{1}{\sigma_{\rho}^2} + \frac{T}{\gamma^2} \right)^{-1}$  and mean  $\mu_{\rho_1} = \sigma_{\rho_1}^2 \left( \frac{C_{it} - R_i \delta - W_{it} \zeta}{\gamma^2} \right)$ . The

variance term  $\sigma_{\rho}^2$  is assigned an inverse gamma prior, which results again in an inverse

gamma posterior with shape parameter  $\eta_{\rho_1} = \eta_{\rho_0} + \sum_{i=1}^n \frac{(\rho_i)^2}{2}$  and scale parameter

$$\kappa_{\rho_1} = \kappa_{\rho_0} + \frac{n}{2}.$$

The final two parameters in the latent variable model  $C_{it}$  and  $\gamma$ , influence multiple parts of the model, i.e. also the measurement model and (or) the choice model. It is not hard to see that the conditional posteriors for these parameters are highly complex and require a M-H algorithm to generate draws from these densities in the Gibbs Sampler. The distribution on  $C_{it}$  described above serves as a prior resulting in the following conditional posterior density

$$p(C_{it} |) \propto \exp \left( - \frac{(C_{it} - R_i \delta - W_{it} \zeta - \rho_i)^2}{2\gamma^2} \right) \cdot \left[ \frac{\exp \left( (Z_{ijt} \beta_f + X_{ijt} \beta_i) \exp \left( \tau C_{it} - \frac{(\tau \gamma)^2}{2} \right) \right)}{\sum_{k \in D_{it}} \exp \left( (Z_{ikt} \beta_f + X_{ikt} \beta_i) \exp \left( \tau C_{it} - \frac{(\tau \gamma)^2}{2} \right) \right)} \right] \exp \left( - \frac{(I_{it}^* - C_{it})^2}{2(1-\gamma^2)} \right)$$

Note that in the above conditional posterior we augment the categorical scale response  $I_{it}$  to a continuous variable  $I_{it}^*$  drawn from a truncated normal density. The normality assumption is a result of the selected ordered probit model and the truncation requires that the value of  $I_{it}^*$  falls in between the thresholds for the level of the respective response category, i.e.  $I_{it}=g$ . This data augmentation step facilitates the estimation procedure.

In order to restrict the overall variance of the measurement model to 1, we imposed the normalisation  $\sigma_v^2 = 1 - \gamma^2$ . This requires  $\gamma$  to fall between 0 and 1, which we adopt by

imposing a uniform prior over this interval. Consequently, the conditional posterior is of the following shape, but still requires a M-H algorithm.

$$p(\gamma | I) \propto \frac{\exp\left((Z_{ij} \beta_f + X_{ij} \beta_i) \exp\left(\tau C_{it} - \frac{(\tau \gamma)^2}{2}\right)\right)}{\sum_{k \in D_n} \exp\left((Z_{ik} \beta_f + X_{ik} \beta_i) \exp\left(\tau C_{it} - \frac{(\tau \gamma)^2}{2}\right)\right)} \exp\left(-\frac{(I_{it}^* - C_{it})^2}{2(1-\gamma^2)}\right)$$

### 7.A.2 - The Gibbs Sampler

Model estimation proceeds by sequentially drawing from the set of conditional posteriors. The sequence is repeated for a number of times. The draws from the conditional posterior will eventually converge to a stable level and represent draws from the joint posterior of model parameters. These converged draws provide information on the parameters of interest. The sequence of draws proceeds in the following order:

1. Assign starting values to all model parameters.
2. Update the fixed coefficients of the choice model. M-H required Feed into 3.
3. Update the hyper-parameters of the mixing densities choice model. Drawing from a normal and inverse gamma density. Feed into 4.
4. Update the augmented (individual specific) utility parameters  $\beta_i$ . M-H required.
5. Update the threshold parameters of the measurement model. M-H required.
6. Augment the response variable  $I_{it}^*$ . Draw from truncated normal density using 5.
7. Update  $\delta$  and  $\zeta$ , draw from a normal density.
8. Update  $\rho_i$ , draw from a normal density.
9. Update  $\sigma_\rho^2$ , draw from an inverse gamma density.
10. Update  $C_{it}$ , M-H required.
11. Update  $\gamma$ , M-H required.
12. Repeat steps 2-11 a large number of times and check convergence of the draws by examining posterior plots and using Geweke's convergence diagnostics (Geweke, 1992).
13. Store a set of converged draws and describe the mean and st. error and other elements of interest. These draws serve as informational content for posterior analysis.



### 7.A.3 - Applied Prior values:

**Table 7A.3.1 – Prior parameters in the choice model:**

Prior Parameter	Description	Affects parameter	Size	Value
$\mu_0$	Prior mean for the mean of the normal mixing density in the choice model	$\mu$	scalar	0
$\sigma_0^2$	Prior variance for the mean of the normal mixing density in the choice model	$\mu$	scalar	1000
$\eta_0$	Prior shape parameter for the inverse gamma prior density on $\sigma^2$	$\sigma^2$	scalar	0.5
$\kappa_0$	Prior scale parameter for the inverse gamma prior density on $\sigma^2$	$\sigma^2$	scalar	0.5
$\vartheta_0$	Prior mean for the fixed parameters $\beta_f$ and $\tau$ in the choice model	$\beta_f$ and $\tau$	$(kf + 1)$ by 1	0
$V_0$	Prior variance for the fixed parameters $\beta_f$ and $\tau$ in the choice model	$\beta_f$ and $\tau$	$(kf + 1)$ by $(kf + 1)$	$1000 \cdot I_{kf}$

**Table 7.A.3.2 – Prior parameters in the self-reported choice certainty model**

Prior Parameter	Description	Affects parameter	Size	Value
$\psi_{g0}$	Prior mean for the threshold parameter in the measurement model	$\psi_g$	scalar	0
$\sigma_{g0}^2$	Prior variance for the threshold parameter in the measurement model	$\psi_g$	scalar	1000

**Table 7.A.3.3 Prior parameters in the latent variable model**

Prior Parameter	Description	Affects parameter	Size	Value
$\mu_{\delta\zeta_0}$	Prior mean for the parameters describing the latent variable	$\delta$ and $\zeta$	$ks$ by 1 vector	0
$V_{\delta\zeta_0}$	Prior variance for parameters describing the latent variable	$\delta$ and $\zeta$	$ks$ by $ks$ matrix	1000
$\eta_{\rho_0}$	Prior shape parameter for the inverse gamma prior density on $\sigma_\rho^2$	$\sigma_\rho^2$	scalar	0.5
$\kappa_{\rho_0}$	Prior scale parameter for the inverse gamma prior density on $\sigma_\rho^2$	$\sigma_\rho^2$	scalar	0.5

## Chapter 8: Conclusions and discussion

This thesis addressed the extent to which individuals are willing to pay for public safety programs reducing the risks of natural hazards. Our interest was not so much in the actual level of willingness-to-pay, but more in the development of discrete choice models to take the impact of preference uncertainty, and related preference dynamics in response patterns across respondents and over the choice sequence into account. In the context of natural hazards, the common lack of experience with the natural hazard itself and unfamiliarity with making trade-offs between safety and income is likely to induce aspects of preference uncertainty. Being able to control for preference uncertainty in discrete choice models may help in improving the validity and reliability of willingness-to-pay estimates obtained from stated choice experiments where respondents lack experience with the presented trade-offs. This chapter summarizes the main contributions and conclusions of this thesis. Section 8.1 starts with an overview of the different parts of the thesis and provides an answer to the main research questions in the context of a stated choice experiment on flood risk valuation in the Netherlands in the face of climate change. Section 8.2 provides an integrated discussion regarding the contributions, findings and implications of this thesis for researchers and policy makers. It also comprises recommendations for future research. Section 8.3 finalizes with a view on the current literature on discrete choice modelling in environmental economics.

### **8.1 Overview and answers to the research questions**

From a theoretical perspective, as discussed in Chapter 2, stated choice experiments form a suitable tool to elicit individual willingness-to-pay estimates for public safety programs (or alternative non-market goods and services) as long as respondents know their preferences. The application of stated choice experiments to (public) safety programs is, however, not straightforward. The complexity of natural hazards and the frequency at which they occur are likely to be associated with aspects of ‘unfamiliarity’ and ‘limited experience’ at the level of the respondent. These aspects are not necessarily related to the properties of the natural hazards, but also to the related trade-offs between safety and income presented to respondents. Consequently, respondents may experience preference uncertainty about appropriate risk reducing policies when being interviewed in a stated choice experiment, or when making actual trade-offs in real-world markets. The degree of preference uncertainty may vary across respondents and different respondents find different ways to deal with such

preference uncertainties. Therefore, this thesis analyzed the impact of preference uncertainty and more general, heterogeneity in response patterns across respondents and over the choice sequence on willingness-to-pay estimates.

In Chapter 3, the concept of preference uncertainty was defined and put in the context of stated choice experiments. By simultaneously making trade-offs between attributes and alternatives in a stated choice experiment, preference uncertainty may also arise at both levels. As such, two complementary forms of preference uncertainty were defined, but Chapter 3 also highlighted that the empirical identification of these separate forms of preference uncertainty is hampered by the limited informational content of discrete choice type data. The empirical part of this thesis therefore focused on the identification of dynamics in preference and scale parameters in general and puts those into the perspective of preference uncertainty, but alternative causes of these dynamics may exist. Chapter 3 therefore also provided an extensive overview of the current state of the empirical literature and models accounting for the impact of preference uncertainty and related preference dynamics (mainly in the scale parameter) on marginal WTP estimates. This formed the basis for the development of the proposed local multinomial logit model in Chapter 6 and combining the implicit and explicit measurement of preference uncertainty in the hybrid choice model in Chapter 7.

Given the interest in preference uncertainty and dynamics in response patterns, the stated choice experiment on flood risk valuation in the Netherlands was specifically designed to test various hypotheses related to the three research questions defined in Chapter 1. The characteristics and development of the stated preference survey were described in detail in Chapter 4. We deliberately chose an area characterized by low probability – high impact flood risks where respondents are likely to experience preference uncertainty. The careful development and testing of risk communication measures helped in bringing about the required information to respondents, but is unlikely to have entirely eliminated respondents' preference uncertainty entirely. Three alternative versions of the experimental design were generated to test for the presence of a starting point bias (Chapter 6) and the relationship between self-reported choice certainty measures and the implicit measurement of preference uncertainty in the random utility model (Chapter 7). Chapter 5 combined all three samples to test for heterogeneity in response patterns across respondents. We now turn to a discussion of the results of these empirical chapters and the answers they provided regarding the three research questions.

*Q1. How can we account for heterogeneity in response patterns across respondents in discrete choice models in a flexible and behaviourally relevant way, without increasing the complexity of estimation? How sensitive are willingness-to-pay estimates to alternative specifications of preference heterogeneity in discrete choice models?*

In Chapter 5 alternative mixing densities characterizing heterogeneity in response patterns within the population of interest have been tested in the mixed multinomial logit model (MMNL). Mixing densities are required since individual-level parameters cannot be estimated accurately enough given the limited number of choice tasks (generally) presented to respondents in a stated choice experiment. On the one hand, the mixing density thereby measures uncertainty regarding individual-level parameter estimates. On the other hand, it accounts for heterogeneity in response patterns across respondents. Hence, the mixing density is merely an approximation of the true distribution of response patterns (or marginal WTP) across the population of interest. This implies that there is a limit to the flexibility of the mixing density specified in the MMNL model. Generally, the more flexible the distribution, the more parameters are required to characterize the shape, support and location of the mixing density. Not only does this increase the requirements on the data to empirically identify these parameters, also the likelihood function becomes more complex to estimate using classical estimation methods. Bayesian estimation methods can overcome the latter issue by not relying on optimization algorithms.<sup>101</sup> The latter method and more complex mixing densities in general are, however, not included in standard econometric software packages like NLOGIT, BIOGEME or STATA, complicating estimation by standard practitioners. Most researchers have therefore applied standard distributional specifications like the normal and lognormal distribution each having its specific drawbacks in terms of behavioural relevance. That is, respondents either can have both positive and negative marginal WTP levels for policy attributes in the case of a normal distribution or they have an incredibly high marginal WTP for a lognormal distribution.

To this end, we proposed the asymmetric triangular distribution as a new mixing density in the MMNL model in Chapter 5. Its properties can overcome the drawbacks of more common distributions by having a bounded support and ability to control for skewed distribution of marginal WTP over the population of interest. As such, the proposed distribution is flexible in its form, results in behaviourally relevant marginal WTP estimates

---

<sup>101</sup> Bayesian methods also facilitate non-nested model comparison, an important reason for us to estimate the MMNL models in Chapter 5 in a Bayesian framework.

and only slightly increases the complexity of estimation, particularly in a Bayesian framework.<sup>102</sup> Naturally, the triangular distribution also has its own drawbacks due to its relatively simple form. Indeed, our results reveal that all three continuous mixing densities mentioned above have difficulties approximating the ‘true distribution’ of preferences over the population of interest, including the asymmetric triangular distribution. A more in depth analysis using Latent Class (LC) modelling revealed that not all respondents take all policy attributes, including the cost of the policy alternative, into account when making decisions. Consequently, some respondents either have a zero marginal WTP for a specific attribute while for others marginal WTP approaches infinity. It is this aspect of behaviour that lies at the heart of the observed shape inconsistencies in our case study. The properties of the lognormal distribution can best accommodate these aspects, but can for example not take into account zero WTP levels. Moreover, its shape is still relatively inflexible resulting in potential empirical identification issues as revealed by high correlations between its underlying hyper-parameters.

The empirical analysis in Chapter 5 showed that heterogeneity in response patterns across respondents is important and should be taken into account. Researchers therefore need to carefully select the shape of the mixing density when applying the MMNL model. Not unexpectedly, the alternative properties of the used mixing densities have a substantial impact on marginal WTP estimates derived through the MMNL model. Appropriate selection of the mixing density is therefore quintessential to derive valid and reliable marginal WTP estimates. Given the empirical limitations to derive individual-level parameter estimates and therefore deduction of the true mixing density, proper sensitivity analysis seems to be required to define whether marginal WTP estimates are sensitive to the chosen mixing density. We find that the LC model is particularly useful in providing insight in underlying behavioural patterns. The issue of heterogeneity in response patterns is put into a broader perspective in Section 8.2.

*Q2. To what extent is respondents’ choice behaviour subject to preference dynamics, i.e. learning and fatigue effects, over the choice sequence? And does this have an impact on willingness-to-pay estimates?*

---

<sup>102</sup> Note that the Johnson SB distribution also combines a flexible shape with a bounded support, but estimation frequently results in a set of highly correlated hyper-parameters (e.g. Train and Sonnier, 2005). Estimation using the asymmetric triangular distribution requires the identification of three instead of four parameters.

Chapter 6 addressed preference dynamics over the choice sequence. Preference dynamics imply that response patterns vary over the choice sequence, which the researcher observes by obtaining variations in preference and (or) scale parameters. The discovered preference hypothesis and the theory of coherent arbitrariness attribute such dynamics respectively to aspects of institutional and value learning, and an internal drive for consistency. As such, both theories predict that at the start of the choice sequence individual preferences are ill-defined, but that by the end of the choice sequence these preferences have (gradually) evolved and stabilize at a specific level. Naturally, alternative explanations for preference dynamics exist, for example randomness in decision making due to preference uncertainty, learning effects, fatigue or survey engagement in general. Although the first and final type of preference dynamics are likely to manifest themselves in an ad hoc fashion during the stated choice experiment, preference dynamics are observed by the researcher through changes in preference and scale parameters. The researcher is therefore faced with a challenge to elicit choice task specific preference parameters or welfare measures as accurately as possible.

First, this thesis reduces the standard errors, i.e. increases efficiency, of choice task specific parameter estimates by rotating the order of appearance of choice cards across versions of the design. As such, at each moment during the choice sequence all choice cards included in the experimental design are answered by multiple respondents. By evaluating all trade-offs in the design, the impact of specific choice cards on choice task specific preference parameters is reduced. Second, Chapter 6 argues that the most commonly applied econometric test to identify preference dynamics, i.e. the Swait and Louviere (1993) test procedure, has its limitations in testing for what we label as within and between sample preference dynamics. Its main drawback is that the obtained parameter estimates may be subject to under- and over-smoothing. That is, the Swait and Louviere (1993) test procedure either fully combines choice task specific models or analyzes them in a separate fashion, thereby neglecting that preferences may evolve gradually over the choice sequence. Choice task specific preference parameters are inefficient given the limited sample sizes frequently used, including our case study. Combining choice tasks in a single model implies welfare measures are identical and subtle differences in preference patterns may be neglected. Chapter 6 proposes a semi-parametric model, labelled as the Local Multinomial Logit (L-MNL) model, which does take into account these possible and gradual changes in preferences by smoothing parameter estimates. In fact, the model provides an intermediate solution to fully combining responses to different choice tasks or analyzing them in a completely

independent fashion. As such, efficiency of choice task specific preference parameters increases on average by 54 percent in our case study.

In the empirical application, an attempt was made to induce a starting point bias using a split sample approach. By anchoring two groups of respondents on respectively high and low price levels for exactly the same policies in the initial choice task, we contrasted the predictions of the discovered preference hypothesis and theory of coherent arbitrariness. Results from alternative model specifications revealed that the initial choice task had an impact on the sample presented with the high price levels. In choice tasks two and three, these respondents revealed a lower tendency to select the status quo relative to subsequent choice tasks. Encountering lower prices directly after the initial choice task may have caused this effect. Within the sample presented with low prices such preference dynamics are not as pronounced. Moreover, results from the L-MNL model suggest that apart from the impact of the initial choice task on the tendency to select the status quo option, limited within and between sample dynamics in willingness-to-pay estimates are present in our case study. In particular, the observed starting point bias between the two samples regarding the status quo wears out after a couple of choice tasks. On top of that, respondents within the high starting bid sample are willing to pay more for an additional percentage of compensation at the start of the survey compared to the other sample. Again, this effect is no longer present after the fifth choice task. Finally, within sample preference dynamics are barely identified after the fourth choice task. The results from the proposed L-MNL model provide a consistent pattern of preference dynamics over the choice sequence due to the underlying smoothing procedure. The observed patterns are confirmed by a set of independently estimated choice task specific models, which however pick up more random fluctuations in preference parameters. It is therefore not surprising that the Swait and Louviere (1993) test procedure has a tendency to combine choice task specific models and attribute some subtle differences in preferences over the choice sequence to variations in the scale parameter.

Overall, the results of the proposed L-MNL model suggest that preferences converge between both samples already half-way the survey and they stabilize within each sample. Thereby, we find limited support for the discovered preference hypothesis. Given our research question, we conclude that in our case study preference dynamics are only present at the start of the survey and primarily have an impact on the tendency to select the status quo, not so much marginal willingness-to-pay estimates. Some dynamics are, however, found in welfare estimates. Since these dynamics gradually disappear, researchers may neglect responses to the first choice tasks in order to obtain stable welfare estimates. In our case

study, this would imply the first four choice tasks, but the exact number remains an empirical issue.

*Q3. Do self-reported choice certainty follow-up questions offer a useful tool to improve willingness-to-pay estimates derived from stated choice experiments?*

In this thesis implicit and explicit measurement of preference uncertainty has been discussed extensively. Implicit measurement accounts for the impact of preference uncertainty on the scale of the utility function. Certain respondents are assumed to make better informed and more consistent decisions, which are observed by the researcher in the form of lower variance, or higher scales, in the utility functions. Discrete choice models, like the heteroskedastic MNL model (HMNL) are at the disposal of the researcher in order to account for variations in (relative) scale across respondents possibly caused by preference uncertainty and the closely related concept of choice task complexity. Frequently, researchers also have additional information available regarding the degree of preference uncertainty experienced by respondents in stated choice experiments. This information is obtained from follow-up questions directly after each choice task requesting respondents to state their level of choice certainty. Explicit measurement attempts to explain these self-reported choice certainty measures based on respondent and choice task characteristics. Chapter 7 explored whether the information regarding preference uncertainty contained in these self-reported choice certainty questions can be used to better trace the impact of preference uncertainty on choices in stated choice experiments and consequently control for the impact on willingness-to-pay estimates.

Thus far, treatment of self-reported choice certainty responses in stated choice experiments has been limited from a methodological perspective. First, existing approaches fail to make the connection that choices in the experiment and responses to the follow-up question are both possibly affected by preference uncertainty. Second, several papers have interpreted self-reported choice certainty as a direct measure of preference uncertainty. Without controlling for possible measurement errors they have directly incorporated responses to the follow-up questions in the discrete choice model. The latter may also be associated with endogeneity bias, since choice certainty is likely to be correlated with other un-modelled factors that enter in the random component of the utility function. Chapter 7 takes both shortcomings into account by developing a hybrid choice model in which latent



preference (un)certainty simultaneously affects the decision in the choice task and the response to the follow-up question.

Results show that a correlation exist between the implicit and explicit measurement of preference uncertainty. Respondents who reveal a higher scale parameter also state they are more certain about their choices in the follow-up questions. Moreover, preference uncertainty is increasing in the complexity of the choice task. However, we do not find an improvement in model fit when accounting for preference uncertainty in the choice model using the proposed hybrid simultaneous choice model. Most important, marginal WTP estimates are not affected by preference uncertainty when modelling the impact of preference uncertainty either through the scale parameter or by accounting for alternative choice heuristics. Our results do indicate that the responses to the choice model may help in better explaining responses to the follow-up questions. This suggests that a sequential modelling approach is more appropriate, but even in this case the use of a hybrid sequential modelling approach does not result in an improvement in model fit compared to existing approaches. Given the complexity of the model structure, we therefore conclude that responses to the choice model and follow-up questions can best be analyzed in an independent fashion or two-stage model. The latter approach seems most promising, since results from the choice model turned out to be particularly useful in explaining self-reported choice certainty.

## **8.2 Discussion and directions for future research**

A natural question that follows from the analysis presented in this thesis is whether preference uncertainty should be of concern to researchers and policy makers when conducting a stated choice experiment. The answer to this question is twofold. First, the presented case study provides a positive message for policy makers, because it suggests that welfare measures of interest are relatively stable over the choice sequence. Preference uncertainty is not likely to induce dynamics in response patterns over the choice sequence. Second, even though preference dynamics may be absent, which makes it obsolete to take possible learning and fatigue effects into account during the analysis, preference uncertainty may still induce significant heterogeneity in response patterns across respondents. In identifying and explaining heterogeneity of response patterns across respondents, preference uncertainty may still play a significant role. Therefore, preference uncertainty should be of concern to researchers and policy makers. The former group can use this information in the development and analysis of the survey to reduce the impact of preference uncertainty on welfare measures, while the latter group can evaluate the extent to which stated choice

experiments can be considered a valid and reliable input for cost-benefit analysis. Remarks to these conclusions are provided below.

Chapter 6 revealed that dynamics in preference and scale parameters over the choice sequence are only observed to a limited extent and that when such dynamics are present they hardly affect willingness-to-pay estimates. Hess et al. (2012) confirm our finding that the role of constants becomes less important as respondents progress through a sequence of choices. Moreover, Hess et al. (2012) also show that controlling for scale heterogeneity over the choice sequence barely has an impact on substantive model results. These results are in line with results from Chapter 7, where marginal willingness-to-pay estimates are not affected by controlling for the impact of preference uncertainty on the choice model. As discussed in Chapter 3, stronger evidence of scale dynamics over the choice sequence is found in various empirical papers (e.g. Brown et al., 2008; DeShazo and Fermo, 2002; Holmes and Boyle, 2005; Ladenburg and Olsen, 2008). These results should not be neglected given the limited preference and scale dynamics observed in our (single) case study. The discussions in Chapters 3 and 6, however, do point out that scale dynamics may be a reflection of (or are confounded with) underlying preference dynamics. That is, since identification of choice task specific preference parameters can be problematic given the experimental design and the sample size, some of the preference dynamics are captured by the scale parameter. Indeed, Hess et al. (2012) merely find evidence of preference dynamics over the choice sequence. The methods proposed in this thesis, being the set-up of the experimental design and the L-MNL model, provide a good starting point for future research on the issue of preference dynamics.

The absence of preference and (or) scale dynamics over the choice sequence does not remove the potential impact of preference uncertainty on response patterns. Preference uncertainty may manifest itself, for example, in the adoption of simplifying choice heuristics. This thesis has put a lot of emphasis on the identification of heterogeneity in response patterns (see for example Chapter 5). We believe identification is a first important step, because if our models do not allow to capture simplifying choice heuristics, like attribute non-attendance (e.g. Campbell, Hutchinson and Scarpa, 2008), there is no ground to find explanations. Finding explanations and causes for the observed response patterns is, however, critical for interpreting the observed distribution of marginal willingness-to-pay over the population of interest. The same holds for preference uncertainty as an explanatory factor of response patterns. Choice certainty follow-up questions, as applied in Chapter 7, or related certainty and attitudinal statements in general, can play an important role in explaining the

observed levels of response heterogeneity. In particular, because Brouwer et al. (2010) showed that choice certainty follow-up questions do not affect the responses to the choice task itself. Together with the proposed hybrid choice model in Chapter 7, such follow-up questions allow to simultaneously trace the impact of preference uncertainty on the choice model and provide explanations for the level of preference uncertainty experienced by the respondents. A clear topic for future research is to find sources of preference uncertainty in hybrid choice models, possibly controlling for dynamic response patterns but also at the level of the respondent in general. This provides valuable information for researchers and policy makers as they may be able to adjust the development of their surveys accordingly to reduce the level of preference uncertainty experienced by respondents.

### **8.3 Final remarks**

The methodological contributions presented in this thesis, and summarized in Section 1.6, are part of a broader and quickly advancing literature exploring the bounds on the informational content of discrete choice type data. The complexity of the applied econometric models is increasing rapidly, in particular with flexible software programs as python BIOGEME becoming available to researchers. On the one hand, such deep analysis of the data provides good insight in the possibilities and limitations of particular datasets and the underlying behavioural patterns. From this perspective, research should indeed pursue in further exploring identification of heterogeneity in responses across respondents and over the choice sequence. Simulation exercises seem most helpful in these circumstances to clarify the limitations of the underlying experimental design and econometric models before applying them to real world datasets. Policy makers, however, should be provided with clear and simple information. A clear evaluation of the extent to which such complex models can provide the required information seems necessary. As noted above, we believe the L-MNL model, as an intermediate model, can provide these insights and deserves further exploration in future research not only to account for preference dynamics over the choice sequence, but also across respondents. In perspective of the insights and results provided by this thesis, controlling for heterogeneity in response patterns across respondents has a higher priority than controlling for heterogeneity in responses over the choice sequence.

## Samenvatting (Summary in Dutch)

Dit proefschrift heeft als uitgangspunt om de individuele betalingsbereidheid voor het reduceren van overstromingsrisico's in Nederland te bestuderen. De focus van het onderzoek ligt op de ontwikkeling van discrete keuze modellen die kunnen controleren voor de invloed van preferentieonzekerheid op het keuzegedrag in een discreet keuze experiment betreffende het reduceren van overstromingsrisico's. Het proefschrift gaat in op drie onderzoeksvragen gerelateerd aan preferentieonzekerheid en mogelijk bijkomende leereffecten. Iedere onderzoeksvraag behandelt een van de volgende onderwerpen: i) heterogeniteit in antwoordpatronen tussen respondenten; ii) veranderingen in antwoordpatronen gedurende een discreet keuze experiment; en iii) het gebruik van vervolgvragen om preferentieonzekerheid te meten. De methodologische bijdragen van dit proefschrift zijn toepasbaar in meerdere contexten die gebruik maken van discrete keuze experimenten en kunnen de validiteit en betrouwbaarheid van zogeheten willingness-to-pay (WTP) schattingen verhogen wanneer er mogelijk sprake is van preferentieonzekerheid. In de hierop volgende uiteenzetting wordt een samenvatting gegeven van het proefschrift en bovenstaande terminologie toegelicht.

### **Waardering van overstromingsrisico's en preferentie onzekerheid**

De voorspelde stijging van de zeespiegel als gevolg van klimaatverandering zal de Nederlandse blootstelling aan overstromingsrisico's verhogen. Beschermingsmaatregelen moeten mogelijk worden aangepast, maar de economische haalbaarheid van dergelijke projecten hangt af van hoe de bijbehorende monetaire opbrengsten in verhouding staan tot de kosten. Hoofdstuk 2 laat zien dat vanuit theoretisch oogpunt discrete keuze experimenten een geschikte methode kunnen zijn om een afname in overstromingsrisico's uit te drukken in monetaire termen zolang respondenten op de hoogte zijn van hun eigen preferenties.

De basis van de methode is gebaseerd op het feit dat individuen in diverse beslissingen hun persoonlijke veiligheid in acht nemen, zo ook in hun blootstelling aan overstromingsrisico's. Zo zijn ze (mogelijk) bereid om meer te betalen voor een huis wat minder blootgesteld wordt aan een dergelijk risico. Door mensen een aantal hypothetische keuzes voor te leggen tussen hun persoonlijke veiligheid en de daarmee gemoeid gaande kosten kan de mate waarin men bereid is om meer te betalen voor een verlaging van het overstromingsrisico bepaald worden. Overstromingsrisico's zijn echter een complex onderwerp waarmee mensen weinig affiniteit hebben, omdat ze zelden voorkomen en mensen daarmee een gebrek aan ervaring hebben. Een direct gevolg hiervan is dat mensen ook geen

of beperkte ervaring hebben in het maken van gerelateerde veiligheidsbeslissingen en de daarmee gemoeid gaande kosten. Het feit dat bescherming tegen overstromingen in Nederland een volledig publiek goed is, onderstreept dit gebrek aan ervaring met bovengenoemde afwegingen. Dit kan een bron vormen van preferentieonzekerheid in het discrete keuze experiment wat ten grondslag ligt aan dit proefschrift. De mate van preferentieonzekerheid kan verschillen tussen individuen, zo ook de wijze waarop zij hier mee omgaan in het maken van keuzes. Naast preferentieonzekerheid analyseert dit proefschrift dan ook de invloed van heterogeniteit in preferenties tussen respondenten op hun betalingsbereidheid voor een afname in overstromingsrisico's.

Hoofdstuk 3 bestudeert het theoretische concept van preferentieonzekerheid en plaatst het in de context van discrete keuze experimenten. Respondenten van een dergelijk experiment worden gevraagd uit een beperkte set aan alternatieven, hier gepresenteerd als beleidsplannen die overstromingsrisico's reduceren, het beste alternatief uit te kiezen. Hierbij maken ze gelijktijdig afwegingen tussen de verschillende alternatieven, maar ook de individuele karakteristieken (of attributen) die deze alternatieven beschrijven. Preferentieonzekerheid kan daarmee ook op beide niveaus plaatsvinden. Zo worden twee complementaire vormen van preferentieonzekerheid gedefinieerd. Empirische identificatie van beide vormen wordt echter beperkt door de beperkte informatie hoeveelheid besloten in data verkregen via discrete keuze experimenten. Om deze reden wordt er in de Hoofdstukken 5 t/m 7 primair gekeken naar veranderingen in preferentie en schaal parameters en worden deze veranderingen in de context van preferentieonzekerheid geplaatst. Hoofdstuk 3 sluit af met een uitgebreid literatuur overzicht van de toegepaste econometrische modellen en resultaten betreffende het effect van preferentieonzekerheid en preferentieveranderingen op WTP parameters. De beperkingen in de bestaande de modellen vormden de basis voor de ontwikkeling van twee nieuwe discrete keuze modellen als gepresenteerd in de Hoofdstukken 6 en 7.

### **Het discrete keuze experiment**

Hoofdstuk 4 beschrijft het discrete keuze experiment waar de analyses in de daaropvolgende hoofdstukken gebaseerd zijn. In het experiment worden ongeveer 1,000 bewoners van dijkkringen 13 en 14, voornamelijk gelegen in de provincies Noord- en Zuid-Holland, middels een internetenquête gepresenteerd met tien keuzekaarten waarin zij telkens twee overheidsplannen te zien krijgen. Ieder plan varieert in de mate waarin i) de kans op overstromingen wordt gereduceerd; ii) de opgelopen schade wordt gecompenseerd in geval

van een overstroming; iii) evacuatie tijd beschikbaar is voorafgaand aan een overstroming; en iv) een verhoging van de jaarlijkse waterschapsbelasting per huishouden noodzakelijk is. Iedere keuzekaart vraagt het individu om het beste plan te identificeren. Men kan ook besluiten dat de huidige beschermingsmaatregelen afdoende zijn. In dat geval zal er geen verhoging van de waterschapsbelasting plaatsvinden.

Het studiegebied is geselecteerd, omdat er een lage kans op overstromingen is van officieel eens per 10,000 jaar, en omdat de gevolgen van een overstroming groot zullen zijn in dit gebied vanwege de dichtbevolktheid. Dit in combinatie met de hoge regulering van overstromingsrisico's in Nederland vanuit de overheid zal er in dit gebied vermoedelijk sprake zijn van preferentieonzekerheid. De mate van preferentieonzekerheid wordt op twee manieren gemeten. Aan de ene kant wordt deze impliciet afgeleid uit het keuzegedrag van respondenten. Aan de andere kant wordt er in een van de versies na iedere keuzekaart gevraagd in hoeverre respondenten zeker zijn van hun gemaakte keuze. In totaal zijn vier verschillende versies van de enquête uitgezet om de drie onderzoeksvragen zoals gedefinieerd in Hoofdstuk 1 te kunnen beantwoorden.

### **Controleren voor verschillen in preferenties tussen respondenten**

Hoofdstuk 5 gaat in op de eerste onderzoeksvraag die een bredere insteek heeft dan alleen preferentieonzekerheid, namelijk het controleren voor verschillen in preferenties tussen respondenten. Het mixed multinomial logit (MMNL) model vormt de basis voor dit hoofdstuk en geldt momenteel als populairste econometrisch model in de discrete keuze literatuur. Door middel van een kansverdeling wordt er in deze modellen gecontroleerd voor het feit dat niet iedereen dezelfde voorkeuren heeft voor bepaalde beleidskarakteristieken. In het bepalen van de vorm van deze kansverdeling dienen echter afwegingen gemaakt te worden tussen de flexibiliteit van de vorm van de distributie, schattingsgemak en gedragsimplicaties. Het is bijvoorbeeld niet te verwachten dat iemand een positief effect ondervindt van hogere waterschapsbelastingen terwijl hier geen verlaging van het overstromingsrisico tegenover staat. Flexibiliteit van en restricties op de verdeling van preferenties tussen respondenten verhoogt echter de schattingscomplexiteit en veelal ervaren onderzoekers identificatie problemen in het schatten van dergelijke modellen. In dit hoofdstuk wordt een nieuwe kansverdeling toegepast, de asymmetrische driehoeksverdeling. De simpele, maar relatief flexibele vorm van de distributie in combinatie met de mogelijkheid om het domein te beperken bieden theoretisch een aantal voordelen ten opzichte van distributies die op dit moment vooral worden toegepast, zoals de normale en de log-normale

verdeling. Bayesiaanse schattingsmethodieken worden toegepast om deze nieuwe model vorm te contrasteren met gangbaardere modelspecificaties.

De resultaten laten zien dat alle gekozen distributies moeite hebben met het benaderen van de daadwerkelijke verdeling van preferenties tussen de respondenten. In een uitgebreide analyse op basis van latent class modellen wordt duidelijk dat niet alle respondenten alle beleidsattributen in hun afwegingen meenemen, inclusief de verhoging van belastingen. Een direct gevolg hiervan is dat sommige respondenten niet bereid zijn om te betalen voor een verlaging van het overstromingsrisico, terwijl een aantal anderen oneindig veel willen betalen. De lognormale verdeling blijkt het beste in staat te zijn om te controleren voor deze gedragspatronen, terwijl de nieuwe driehoeksverdeling een beter beschrijving van de data geeft dan de normale verdeling. Toepassingen van de nieuwe distributie lijken echter een beperkt perspectief te hebben. Desalniettemin laat Hoofdstuk 5 zien dat heterogeniteit in preferenties tussen respondenten een belangrijke rol speelt en dat hiervoor gecontroleerd dient te worden in de gebruikte modellen. Het bepalen van de meest geschikte kansverdeling is essentieel en kan grote invloed hebben op de WTP waarden. Sensitiviteitsanalyses worden dan ook aanbevolen.

### **Veranderingen in keuzegedrag**

Hoofdstuk 6 analyseert of er structurele veranderingen in gedragspatronen plaatsvinden, die mogelijk het gevolg zijn van leer- en vermoeidheidseffecten, gedurende het maken van een discreet keuze experiment. Deze effecten kunnen achterhaald worden door te kijken naar variaties in preferentie dan wel schaalparameters gedurende het keuze experiment. In dit hoofdstuk worden twee contrasterende theorieën vergeleken en empirisch getest. Zowel de theorie van ‘coherent arbitrariness’ als de ‘discovered preference hypothesis’ voorspellen dat mensen onzeker zijn over hun preferenties aan het begin van het keuze experiment, maar dat deze zich geleidelijk vormen (of duidelijk worden) na het maken van meerdere keuzes. Dit proces verschilt ergens tussen de beide theorieën.

Dit hoofdstuk maakt gebruik van twee versies van het keuze experiment die een verschillend startpunt hebben, maar verder op geen enkel punt verschillen. Bovendien staat het gekozen experimentele ontwerp ons toe om kaart specifieke modellen nauwkeuriger te schatten dan voorheen het geval was. Hoofdstuk 6 laat ook zien dat de tot nu toe meest gebruikte methode om te controleren voor veranderingen in keuzegedrag, de Swait en Louviere (1993) test, zijn beperkingen heeft en observaties als volledig gelijk of totaal verschillend ziet terwijl er met leereffecten juist verwacht kan worden dat preferenties zich

geleidelijk ontwikkelen. Om deze reden wordt een semi-parametrische schattingsmethodiek ontwikkeld in Hoofdstuk 6, het local multinomial logit model, waarin keuzekaart specifieke parameters met 54% meer precisie kunnen worden geschat door informatie van vergelijkbare observaties te gebruiken. De resultaten laten zien dat de eerste keuzekaart van invloed is op de daaropvolgende keuzes van respondenten, maar dat dit effect al snel verdwijnt. Daarnaast worden veranderingen in keuzegedrag slechts beperkt waargenomen en tegen het einde van de keuzeserie resulteren beide versies van het experiment in vergelijkbare preferenties en geschatte WTP waarden. Deze bevindingen komen overeen met de voorspellingen van de ‘discovered preference hypothesis’ en ondersteunen dat heterogeniteit in preferenties tussen respondenten een grotere rol speelt dan veranderingen in gedragspatronen van individuele respondenten gedurende een serie aan keuzes.

### **Impliciet en expliciet meten van preferentieonzekerheid**

De beperkte invloed van structurele veranderingen in preferenties over de keuzeserie op het keuzegedrag, zoals geïdentificeerd in Hoofdstuk 6, stelt ons in staat om de invloed van preferentieonzekerheid gedurende de keuzeserie te identificeren. Preferentieonzekerheid wordt op twee manieren gemeten in Hoofdstuk 7. In de eerste plaats impliciet door middel van de schaalparameter in het Random Utility Model. Hierbij wordt uitgegaan dat de storingsterm in het model voor onzekere respondenten een grotere variantie heeft dan voor zekere respondenten. Eerst genoemden hebben dan ook een grotere kans om inconsistente keuzes te maken. De tweede methode meet preferentieonzekerheid expliciet door het stellen van een vervolgvraag na iedere keuzekaart die achterhaalt hoe zeker de respondent van zijn of haar keuze is. Hoofdstuk 7 onderzoekt of de verkregen informatie betreffende preferentieonzekerheid middels vervolgvragen gebruikt kan worden om de invloed van preferentieonzekerheid op het keuzegedrag en WTP schattingen te identificeren.

Hoofdstuk 7 stelt dat het huidige gebruik van vervolgvragen methodologische beperkingen heeft. Zo wordt er meestal niet erkend dat zowel de antwoorden in het keuze experiment als de vervolgvragen worden beïnvloed door preferentieonzekerheid. Bovendien wordt de antwoorden op de vervolgvraag regelmatig geïnterpreteerd als een perfecte maatstaf van preferentieonzekerheid. Zonder te controleren voor mogelijke meetfouten worden de antwoorden op de vervolgvragen direct opgenomen in het keuze model. Dit kan leiden tot een endogeniteitsprobleem. Deze beperkingen worden in Hoofdstuk 7 opgelost door een model te presenteren waarin de antwoorden op het keuze experiment en de vervolgvragen gelijktijdig gegeven worden. Hierbij wordt preferentieonzekerheid gemodelleerd als een latente



variabele. De resultaten laten zien dat de impliciete en expliciete maatstaven van preferentieonzekerheid positief met elkaar gecorreleerd zijn. Preferentieonzekerheid neemt toe als de gepresenteerde keuzes moeilijker worden. Het gepresenteerde model beschrijft de data echter niet beter dan bestaande methodes. De schattingen laten zien dat antwoorden op het keuze experiment kunnen helpen in het verklaren van antwoorden op de vervolgvragen. WTP schattingen veranderen nauwelijks door gelijktijdig zowel impliciet als expliciet te controleren voor preferentieonzekerheid. Mede gegeven de complexiteit van het ontwikkelde model, concluderen we dat antwoorden op de keuzekaarten en de vervolgvragen het beste in twee onafhankelijke modellen of in een sequentieel model geanalyseerd kunnen worden.

### **Aanbevelingen**

De resultaten van het keuze experiment laten zien dat veranderingen in antwoordpatronen gedurende het experiment als gevolg van preferentieonzekerheid niet overduidelijk aanwezig zijn. Het onderwerp van veranderingen in preferenties gedurende een keuze experiment is tot nu toe echter onderbelicht gebleven in de discrete keuze literatuur in vergelijking tot heterogeniteit in preferenties tussen respondenten. Dit proefschrift heeft een bijdrage geleverd aan de empirische identificatie van dergelijke veranderingen en biedt daarmee een uitgangspunt voor vervolg onderzoek, met name in situaties waar mensen meermaals met de vergelijkbare keuzesituaties geconfronteerd worden.

Het ontbreken van dynamiek in de preferentiestructuur betekent niet dat preferentieonzekerheid geen invloed kan hebben op de antwoordpatronen van respondenten. Zo kunnen onzekere respondent nog steeds andere (makkelijkere) beslissingsregels gebruiken dan zekere respondenten. Identificatie van een dergelijke heterogeniteit, zoals beschreven in Hoofdstuk 5, blijft een speerpunt van vervolgonderzoek, maar ook de verklaring van de geobserveerde heterogeniteit dient niet onderbelicht te blijven. Hierin is een mogelijke rol weggelegd voor de expliciete preferentieonzekerheid vervolgvragen, aangezien deze weinig invloed lijken te hebben op het keuzegedrag zelf.

Tot slot dienen beleidsmakers en onderzoekers nog steeds rekening te houden met preferentieonzekerheid in discrete keuze experimenten. Onderzoekers kunnen in het ontwikkelen van het experiment rekening houden met preferentieonzekerheid door de inhoud en presentatie aan te passen op basis van de commentaren van respondenten. Beleidsmakers dienen informatie over preferentieonzekerheid, opgesloten in bijvoorbeeld vervolgvragen, te gebruiken om te evalueren of het gepresenteerde keuze experiment kan leiden tot betrouwbare welvaartsmaatstaven die gebruikt kunnen worden in kosten-baten analyses.

## References

- Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C., and Hess, S. (2011), "Attitudes and value of time heterogeneity," in *Applied Transport Economics - A management and policy perspective*, De Boeck Publishing, ed., pp. 525-548
- Adamowicz W., Louviere J. and Williams M. (1994), "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities", *Journal of Environmental Economics and Management*, 26 (3), pp. 271-292
- Aerts, J. C. J. H., Sprong, T., & Bannink, B. A. (2008), Aandacht voor veiligheid, *KvR rapport 009/2008*, <http://www.deltacommissie.com/doc/Aandacht%20voor%20veiligheid%20.pdf>
- Akter S., Bennett J. and Akhter S. (2008), "Preference uncertainty in contingent valuation", *Ecological Economics*, 67 (3), pp. 345-351
- Alberini A., Boyle K. and Welsh M. (2003), "Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty", *Journal of Environmental Economics and Management*, 45 (1), pp. 40-62
- Allenby G. M. and Rossi P. E. (1998), "Marketing models of consumer heterogeneity", *Journal of Econometrics*, 89 (1-2), pp. 57-78
- Arentze T., Borgers A., Timmermans H. and 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), pp. 229-244
- Ariely D., Loewenstein G. and Prelec D. (2003), "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences", *Quarterly Journal of Economics*, 118 (1), pp. 73-105
- Ariely D., Loewenstein G. and Prelec D. (2006), "Tom Sawyer and the construction of value", *Journal of Economic Behavior & Organization*, 60 (1), pp. 1-10
- Arrow, K., Solow, R., Portney, P., Learner, E., Radner, R., & Schuman, H. (1993), Report of the NOAA Panel on Contingent Valuation, *Federal Register* 58: 4601-4614
- Balcombe K., Bailey A., Chalak A. and Fraser I. (2007), "Bayesian Estimation of Willingness-to-pay Where Respondents Mis-report Their Preferences\*", *Oxford Bulletin of Economics and Statistics*, 69 (3), pp. 413-438
- Balcombe K., Chalak A. and Fraser I. (2009), "Model selection for the mixed logit with Bayesian estimation", *Journal of Environmental Economics and Management*, 57 (2), pp. 226-237
- Balcombe K. and Fraser I. (2011), "A general treatment of don't know responses from choice experiments", *European Review of Agricultural Economics*, 38 (2), pp. 171-191
- Bateman I. J., Burgess D., Hutchinson W. G. and Matthews D. I. (2008), "Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness", *Journal of Environmental Economics and Management*, 55 (2), pp. 127-141
- Bateman I. J., Cole M., Cooper P., Georgiou S., Hadley D. and Poe G. L. (2004), "On visible choice sets and scope sensitivity", *Journal of Environmental Economics and Management*, 47 (1), pp. 71-93
- Bech M., Kjaer T. and Lauridsen J. (2011), "Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment", *Health Economics*, 20 (3), pp. 273-286

- Beck, M., Rose, J. M., & Hensher, D. (2011), "Consistently inconsistent: The role of certainty, acceptability and scale in automobile choice", presented at *2nd International Choice Modelling Conference*, Leeds, UK
- Ben-Akiva M., Bolduc D. and Bradley M. (1993), "Estimation of travel choice models with randomly distribute values of time", *Transportation Research Record: Journal of the Transportation Research Board*, 1413, pp. 88-97
- Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis: Theory and application to travel demand*, MIT Press
- Ben-Akiva, M., Walker, J. L., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (1999), *Integration of choice and latent variable models*, Massachusetts Institute of Technology, Cambridge, MA,
- Bhat C. R. (1998), "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling", *Transportation Research Part A: Policy and Practice*, 32 (7), pp. 495-507
- Blomquist G., Blumenschein K. and Johannesson M. (2009), "Eliciting Willingness to Pay without Bias using Follow-up Certainty Statements: Comparisons between Probably/Definitely and a 10-point Certainty Scale", *Environmental and Resource Economics*, 43 (4), pp. 473-502
- Bockarjova, M., Rietveld, P., & Verhoef, E. (2010), "Immaterial damage valuation in flooding: value of statistical life, value of evacuation and value of injury", presented at *Delta's in Time of Climate Change*, Rotterdam, Netherlands
- Bolduc, D., Ben-Akiva, M., Walker, J. L., and Michaud, A. (2005), "Hybrid choice models with logit kernel: applicability to large scale models," in *Integrated Land-Use and Transportation Model: Behavioural Foundations*, M. Lee-Gosselin & S. Doherty, eds., Elsevier, Oxford, pp. 275-302
- Botzen, W. J. W. (2010), *The economics of insurance against climate change*, *Doctoral Thesis*, VU University Amsterdam
- Botzen W. J. W. and Van Den Bergh J. C. J. M. (2008), "Insurance Against Climate Change and Flooding in the Netherlands: Present, Future, and Comparison with Other Countries", *Risk Analysis*, 28 (2), pp. 413-426
- Botzen W. J. W. and Van Den Bergh J. C. J. M. (2009), "Bounded Rationality, Climate Risks, and Insurance: Is There a Market for Natural Disasters?", *Land Economics*, 85 (2), pp. 265-278
- Bouwer, L. M. and Vellinga, P. (2007), "On the Flood Risk in the Netherlands," in *Flood Risk Management in Europe*, 25 edn, S. Begum, M. J. F. Stive, & J. W. Hall, eds., Springer Netherlands, pp. 469-484
- Boyle K., Holmes T. P., Teisl M. F. and Roe B. (2001), "A Comparison of Conjoint Analysis Response Formats", *American Journal of Agricultural Economics*, 83 (2), pp. 441-454
- Bradley M. and Daly A. (1994), "Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data", *Transportation*, 21 (2), pp. 167-184
- Braga J. and Starmer C. (2005), "Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis", *Environmental and Resource Economics*, 32 (1), pp. 55-89

- Brazell, J. D. & Louviere, J. (1998), Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden, *Department of Marketing, The University of Sydney, July 1998*
- Breffe W. S. and Morey E. R. (2000), "Investigating Preference Heterogeneity in a Repeated Discrete-Choice Recreation Demand Model of Atlantic Salmon Fishing", *Marine Resource Economics*, 15 (1), pp. 1-20
- Brouwer R. (2011), "A mixed approach to payment certainty calibration in discrete choice welfare estimation", *Applied Economics*, 43 (17), pp. 2129-2142
- Brouwer R. and Akter S. (2010), "Informing micro insurance contract design to mitigate climate change catastrophe risks using choice experiments", *Environmental Hazards*, 9 (1), pp. 74-88
- Brouwer R., Dekker T., Rolfe J. and Windle J. (2010), "Choice Certainty and Consistency in Repeated Choice Experiments", *Environmental and Resource Economics*, 46 (1), pp. 93-109
- Brouwer, R. and Kind, J. (2005), "The costs and benefits of flood control policy in the Netherlands," in *Cost-Benefit Analysis and Water Resources Management*, R. Brouwer & D. W. Pearce, eds., Edward Elgar Publishing, Cheltenham, UK
- Brouwer, R. and Schaafsma, M. (2011), "The economics of flood disaster management in the Netherlands," in *Natural disasters: do they cost the earth?*, D. Guha Sapir & I. Santos, eds., Oxford University Press, Oxford
- Brouwer R. and van Ek R. (2004), "Integrated ecological, economic and social impact assessment of alternative flood control policies in the Netherlands", *Ecological Economics*, 50 (1-2), pp. 1-21
- Brown T. C., Kingsley D., Peterson G. L., Flores N. E., Clarke A. and Birjulin A. (2008), "Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement", *Journal of Public Economics*, 92 (7), pp. 1595-1606
- Brownstone D., Bunch D. S. and Train K. (2000), "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles", *Transportation Research Part B: Methodological*, 34 (5), pp. 315-338
- Brownstone D. and Train K. (1998), "Forecasting new product penetration with flexible substitution patterns", *Journal of Econometrics*, 89 (1-2), pp. 109-129
- Burda M., Harding M. and Hausman J. (2008), "A Bayesian mixed logit-probit model for multinomial choice", *Journal of Econometrics*, 147 (2), pp. 232-246
- Cameron T. A. and DeShazo J. R. (2010), "Differential attention to attributes in utility-theoretic choice models", *Journal of Choice Modelling*, 3 (3), pp. 73-115
- Cameron T. A., Poe G. L., Ethier R. G. and Schulze W. D. (2002), "Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?", *Journal of Environmental Economics and Management*, 44 (3), pp. 391-425
- Campbell, D., Doherty, E., Hynes, S., & van Rensburg, T. (2010a), "Combining discrete and continuous mixing approaches to accommodate heterogeneity in price sensitivities in environmental choice analysis", presented at *84th Annual Conference of the Agricultural Economics Society*, Edingburgh, Scotland

- Campbell D., Hutchinson W. and Scarpa R. (2008), "Incorporating Discontinuous Preferences into the Analysis of Discrete Choice Experiments", *Environmental and Resource Economics*, 41 (3), pp. 401-417
- Campbell, D., Lorimer, V., Aravena, C., & Hutchinson, W. G. (2010b), "Attribute processing in environmental choice analysis: implications for willingness to pay", presented at *84th Annual Conference of the Agricultural Economics Society*, Edinburgh, Scotland
- Carlsson F. and Martinsson P. (2008), "How Much is Too Much?", *Environmental and Resource Economics*, 40 (2), pp. 165-176
- Carlsson, F., Morkbak, M. R., & Olsen, S. B. (2010), The first time is the hardest: a test for ordering effects in choice experiments, *Gotenburg University, Working papers in Economics #470*
- Carson R. and Groves T. (2007), "Incentive and informational properties of preference questions", *Environmental and Resource Economics*, 37 (1), pp. 181-210
- Carson R. T., Flores N. E. and Meade N. F. (2001), "Contingent Valuation: Controversies and Evidence", *Environmental and Resource Economics*, 19 (2), pp. 173-210
- Caussade S., Ortuzar J. d. D., Rizzi L. I. and Hensher D. A. (2005), "Assessing the influence of design dimensions on stated choice experiment estimates", *Transportation Research Part B: Methodological*, 39 (7), pp. 621-640
- Champ P. A., Bishop R. C., Brown T. C. and McCollum D. W. (1997), "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods", *Journal of Environmental Economics and Management*, 33 (2), pp. 151-162
- Champ, P. A., Boyle, K. J., & Brown, T. C. (2005), *A primer on nonmarket valuation: The economics of non-market goods and resources*, Kluwer Academic Publishers
- Charlton, M. & Fortheringham, S. (2009), Geographically weighted regression: a white paper, [http://ncg.nuim.ie/ncg/GWR/GWR\\_WhitePaper.pdf](http://ncg.nuim.ie/ncg/GWR/GWR_WhitePaper.pdf)
- Chib S. and Greenberg E. (1995), "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, 49 (4), pp. 327-335
- Chien Y. L., Huang C. J. and Shaw D. (2005), "A general model of starting point bias in double-bounded dichotomous contingent valuation surveys", *Journal of Environmental Economics and Management*, 50 (2), pp. 362-377
- Collins J. P. and Vossler C. A. (2009), "Incentive compatibility tests of choice experiment value elicitation questions", *Journal of Environmental Economics and Management*, 58 (2), pp. 226-235
- Colombo S., Calatrava-Requena J. and Hanley N. (2007), "Testing Choice Experiment for Benefit Transfer with Preference Heterogeneity", *American Journal of Agricultural Economics*, 89 (1), pp. 135-151
- Corso P. S., Hammitt J. K. and Graham J. D. (2001), "Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation", *Journal of Risk and Uncertainty*, 23 (2), pp. 165-184
- CPB (2008), Economische effecten van regulering en subsidiëring van de huurwoningmarkt, *CPB Document 165 (in Dutch)*

- Daly A., Hess S., Patruni B., Potoglou D. and Rohr C. (2011a), "Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour", *Transportation*, 39 (2), pp. 267-297
- Daly A., Hess S. and Train K. (2011b), "Assuring finite moments for willingness to pay in random coefficient models", *Transportation*, 39 (1), pp. 19-31
- Daniel V. E., Florax R. J. G. M. and Rietveld P. (2009), "Flooding risk and housing values: An economic assessment of environmental hazard", *Ecological Economics*, 69 (2), pp. 355-365
- Dasgupta P. (2008), "Discounting climate change", *Journal of Risk and Uncertainty*, 37 (2), pp. 141-169
- Day B. and Pinto Prades J. L. (2010), "Ordering anomalies in choice experiments", *Journal of Environmental Economics and Management*, 59 (3), pp. 271-285
- Daziano R. A. and Bolduc D. (2011), "Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model", *Transportmetrica*, pp. 1-33
- de Blaeij, A. (2003), The value of a statistical life in road safety; Stated preference methodologies and empirical estimates for the Netherlands, *Doctoral Thesis*, VU University Amsterdam, Tinbergen Institute
- de Moel H., Aerts J. and Koomen E. (2011), "Development of flood exposure in the Netherlands during the 20th and 21st century", *Global Environmental Change*, 21 (2), pp. 620-627
- Dellaert B. G. C., Brazell J. D. and Louviere J. J. (1999), "The Effect of Attribute Variation on Consumer Choice Consistency", *Marketing Letters*, 10 (2), pp. 139-147
- DeShazo J. R. and Fermo G. (2002), "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency", *Journal of Environmental Economics and Management*, 44 (1), pp. 123-143
- Dubourg W. R., Jones-Lee M. W. and Loomes G. (1994), "Imprecise preferences and the WTP-WTA disparity", *Journal of Risk and Uncertainty*, 9 (2), pp. 115-133
- Elrod T. and Keane M. P. (1995), "A Factor-Analytic Probit Model for Repressing the Market Structure in Panel Data", *Journal of Marketing Research*, 32 (1), pp. 1-16
- Fan, J. & Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman & Hall, London
- Fan J., Heckman N. E. and Wand M. P. (1995), "Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions", *Journal of the American Statistical Association*, 90 (429), pp. 141-150
- Fenichel E. P., Lupi F., Hoehn J. P. and Kaplowitz M. D. (2009), "Split-Sample Tests of "No Opinion" Responses in an Attribute-Based Choice Model", *Land Economics*, 85 (2), pp. 348-362
- Ferrini S. and Scarpa R. (2007), "Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study", *Journal of Environmental Economics and Management*, 53 (3), pp. 342-363
- Fiebig D. G., Keane M. P., Louviere J. and Wasi N. (2010), "The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity", *Marketing Science*, 29 (3), pp. 393-421

- Fischhoff B., Welch N. and Frederick S. (1999), "Construal Processes in Preference Assessment", *Journal of Risk and Uncertainty*, 19 (1), pp. 139-164
- Flores, N. E. (2003), "Conceptual framework for nonmarket valuation," in *A primer on nonmarket valuation*, P. A. Champ, K. J. Boyle, & T. C. Brown, eds., Kluwer Academic Publishers, pp. 27-58
- Fosgerau M. (2006), "Investigating the distribution of the value of travel time savings", *Transportation Research Part B: Methodological*, 40 (8), pp. 688-707
- Fosgerau M. (2007), "Using nonparametrics to specify a model to measure the value of travel time", *Transportation Research Part A: Policy and Practice*, 41 (9), pp. 842-856
- Fosgerau M. and Bierlaire M. (2007), "A practical test for the choice of mixing distribution in discrete choice models", *Transportation Research Part B: Methodological*, 41 (7), pp. 784-794
- Fosgerau M. and Hess S. (2009), "Competing methods for representing random taste variation in discrete choice models", *European Transport*, 42, pp. 1-25
- Freeman, A. M. (2003), *The measurement of environmental and resource values: theory and methods*, 2 edn, Resources for the Future, Washington
- Gelfand A. E. and Dey D. K. (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations", *Journal of the Royal Statistical Society. Series B (Methodological)*, 56 (3), pp. 501-514
- Gelfand A. E., Smith A. F. M. and Lee T. M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling", *Journal of the American Statistical Association*, 87 (418), pp. 523-532
- Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics 4*, J. Bernardo, ed., Clarendon Press, Oxford, pp. 641-649
- Geweke, J. (2011), "Nonparametric Bayesian Modeling of Monotone Preferences for Discrete Choice Experiments", presented at *Econometrics Summer Conference*, Cowles Foundation for Research in Economics, Yale University
- Greene W. and Hensher D. (2010), "Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models", *Transportation*, 37 (3), pp. 413-428
- Greene W. H. and Hensher D. A. (2003), "A latent class model for discrete choice analysis: contrasts with mixed logit", *Transportation Research Part B: Methodological*, 37 (8), pp. 681-698
- Greene W. H., Hensher D. A. and Rose J. (2006), "Accounting for heterogeneity in the variance of unobserved effects in mixed logit models", *Transportation Research Part B: Methodological*, 40 (1), pp. 75-92
- Groeneveld R. (2010), "Framing and training to induce preference learning in choice experiments", *Marine Resource Economics*, 25, pp. 233-245
- Halton J. H. (1960), "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals", *Numerische Mathematik*, 2 (1), pp. 84-90
- Hanemann, M. and Kristrom, B. (1995), "Preference uncertainty, optimal designs and spikes," in *Current issues in environmental economics*, P. O. Johansson, B. Kristrom, & K. G. Maler, eds., Manchester University Press, pp. 58-77

- Hanley N., Adamowicz W. and Wright R. E. (2005), "Price vector effects in choice experiments: an empirical test", *Resource and Energy Economics*, 27 (3), pp. 227-234
- Hanley N., Mourato S. and Wright R. E. (2001), "Choice Modelling Approaches: A Superior Alternative for Environmental Valuation?", *Journal of Economic Surveys*, 15 (3), pp. 435-462
- Hanley N., Wright R. and Adamowicz V. (1998), "Using Choice Experiments to Value the Environment", *Environmental and Resource Economics*, 11 (3), pp. 413-428
- Hensher D. (2004), "Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings", *Journal of Transport Economics and Policy*, 38, pp. 425-446
- Hensher D. (2006), "Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment", *Environmental and Resource Economics*, 34 (1), pp. 7-44
- Hensher, D. (2010a), "Attribute processing, heuristics and preference construction in choice analysis," in *Choice Modelling: The State-of-the-Art and the State-of-Practice - Proceedings from the Inaugural International Choice Modelling Conference*, S. Hess & A. Daly, eds., Emerald Group Publishing Limited, pp. 35-69
- Hensher D. (2010b), "Hypothetical bias, choice experiments and willingness to pay", *Transportation Research Part B: Methodological*, 44 (6), pp. 735-752
- Hensher D. and Greene W. (2003), "The Mixed Logit model: The state of practice", *Transportation*, 30 (2), pp. 133-176
- Hensher D. and Greene W. (2010), "Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification", *Empirical Economics*, 39 (2), pp. 413-426
- Hensher, D., Rose, J. M., & Beck, M. (2011), "Are there specific design elements of choice experiments and types of people that influence choice response certainty?", presented at *2nd International Choice Modelling Conference*, Leeds, UK
- Hensher D. and Rose J. (2011), "The influence of alternative acceptability, attribute thresholds and choice response certainty on automobile purchase preferences", *Journal of Transport Economics and Policy*, forthcoming
- Hensher, D., Rose, J., & Greene, W. (2005), *Applied choice analysis: A primer*, Cambridge University Press, Cambridge
- Hess S. and Beharry-Borg N. (2011), "Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago", *Environmental and Resource Economics*, pp. 1-23
- Hess, S. & Daly, A. (2010), *Choice Modelling: The State-of-the-Art and the State-of-Practice - Proceedings from the Inaugural International Choice Modelling Conference*, 1 edn, Emerald Group Publishing Limited, UK
- Hess S., Hensher D. A. and Daly A. (2012), "Not bored yet – Revisiting respondent fatigue in stated choice experiments", *Transportation Research Part A: Policy and Practice*, 46 (3), pp. 626-644
- Hess, S. & Rose, J. (2011), "Can scale and coefficient heterogeneity be separated in random coefficient models?", *ITS working paper*



- Hess S. and Rose J. M. (2009), "Allowing for intra-respondent variations in coefficients estimated on repeated choice data", *Transportation Research Part B: Methodological*, 43 (6), pp. 708-719
- Hess, S. & Stathopoulos, A. (2011), Linking response quality to survey engagement: a combined random scale and latent variable approach, Institute for Transport Studies, University of Leeds, *ITS Working Paper*
- Hess S. and Train K. E. (2011), "Recovery of inter- and intra-personal heterogeneity using mixed logit models", *Transportation Research Part B: Methodological*, 45 (7), pp. 973-990
- Hess S., Train K. E. and Polak J. W. (2006), "On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice", *Transportation Research Part B: Methodological*, 40 (2), pp. 147-163
- Holmes T. P. and Boyle K. J. (2005), "Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions", *Land Economics*, 81 (1), pp. 114-126
- Huber J. and Zwerina K. (1996), "The Importance of Utility Balance in Efficient Choice Designs", *Journal of Marketing Research*, 33 (3), pp. 307-317
- Huber O. and Huber O. W. (2008), "Gambles vs. quasi-realistic scenarios: Expectations to find probability and risk-defusing information", *Acta Psychologica*, 127 (2), pp. 222-236
- Johnston R. (2007), "Choice experiments, site similarity and benefits transfer", *Environmental and Resource Economics*, 38 (3), pp. 331-351
- Jonkman S. N., Bockarjova M., Kok M. and Bernardini P. (2008), "Integrated hydrodynamic and economic modelling of flood damage in the Netherlands", *Ecological Economics*, 66 (1), pp. 77-90
- Kabat P., van Vierssen W., Veraart J., Vellinga P. and Aerts J. (2005), "Climate proofing the Netherlands", *Nature*, 438 (7066), pp. 283-284
- Kass R. E. and Raftery A. E. (1995), "Bayes Factors", *Journal of the American Statistical Association*, 90 (430), pp. 773-795
- Kessels R., Goos P. and Vandebroek M. (2006), "A comparison of criteria to design efficient choice experiments", *Journal of Marketing Research*, 43 (3), pp. 409-419
- Kingsley D. C. and Brown T. C. (2010), "Preference Uncertainty, Preference Learning, and Paired Comparison Experiments", *Land Economics*, 86 (3), pp. 530-544
- Kobayashi M., Moeltner K. and Rollins K. (2012), "Latent Thresholds Analysis of Choice Data under Value Uncertainty", *American Journal of Agricultural Economics*, 94 (1), pp. 189-208
- Koop, G. (2003), *Bayesian Econometrics*, Wiley, New York
- Kosenius, A. (2009), Causes of response uncertainty and its implications for WTP estimation in choice experiments, University of Helsinki - Department of economics and management, Helsinki, *Discussion paper no. 29*
- Koster, P. & Koster, H. (2011), "Analyzing observed heterogeneity in preferences: a semiparametric estimation approach", presented at *2nd International Choice Modelling Conference*, Leeds
- Krinsky I. and Robb A. L. (1986), "On Approximating the Statistical Properties of Elasticities", *The Review of Economics and Statistics*, 68 (4), pp. 715-719

- Krinsky I. and Robb A. L. (1990), "On Approximating the Statistical Properties of Elasticities: A Correction", *The Review of Economics and Statistics*, 72 (1), pp. 189-190
- Kunreuther H. and Pauly M. (2004), "Neglecting Disaster: Why Don't People Insure Against Large Losses?", *Journal of Risk and Uncertainty*, 28 (1), pp. 5-21
- Ladenburg J. and Olsen S. B. (2008), "Gender-specific starting point bias in choice experiments: Evidence from an empirical study", *Journal of Environmental Economics and Management*, 56 (3), pp. 275-285
- Lancaster K. J. (1966), "A New Approach to Consumer Theory", *The Journal of Political Economy*, 74 (2), pp. 132-157
- Li C. Z. and Mattsson L. (1995), "Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation", *Journal of Environmental Economics and Management*, 28 (2), pp. 256-269
- Lichtenstein, S. & Slovic, P. (2006), *The construction of preference*, Cambridge University Press, New York
- List J. A. (2003), "Does Market Experience Eliminate Market Anomalies?\*", *Quarterly Journal of Economics*, 118 (1), pp. 41-71
- Loomes G., Orr S. and Sugden R. (2009), "Taste uncertainty and status quo effects in consumer choice", *Journal of Risk and Uncertainty*, 39 (2), pp. 113-135
- Loomes G., Starmer C. and Sugden R. (2010), "Preference reversals and disparities between willingness to pay and willingness to accept in repeated markets", *Journal of Economic Psychology*, 31 (3), pp. 374-387
- Loomis J. and Ekstrand E. (1998), "Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican spotted owl", *Ecological Economics*, 27 (1), pp. 29-41
- Louviere, J. & Eagle, T. (2006), "Confound it! That pesky little scale constant messes up our convenient assumptions", presented at *Sawtooth Software Conference*, Delray Beach, Florida
- Louviere, J., Hensher, D., & Swait, J. D. (2003), *Stated choice methods: analysis and applications*, Cambridge University Press, Cambridge
- Louviere J., Street D., Burgess L., Wasi N., Islam T. and Marley A. (2008), "Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information", *Journal of Choice Modelling*, 1 (2), pp. 128-163
- Louviere J. and Swait J. (2010), "Commentary--Discussion of "Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test"", *Marketing Science*, 29 (1), pp. 18-22
- Lundhede T. H., Olsen S. B., Jacobsen J. B. and Thorsen B. J. (2009), "Handling respondent uncertainty in Choice Experiments: Evaluating recoding approaches against explicit modelling of uncertainty", *Journal of Choice Modelling*, 2 (2), pp. 118-147
- Lusk J. L., Roosen J. and Fox J. A. (2003), "Demand for Beef from Cattle Administered Growth Hormones or Fed Genetically Modified Corn: A Comparison of Consumers in France, Germany, the

United Kingdom, and the United States", *American Journal of Agricultural Economics*, 85 (1), pp. 16-29

Maaskant B., Jonkman S. N. and Bouwer L. M. (2009), "Future risk of flooding: an analysis of changes in potential loss of life in South Holland (The Netherlands)", *Environmental Science & Policy*, 12 (2), pp. 157-169

Manski C. F. (1977), "The structure of random utility models", *Theory and Decision*, 8 (3), pp. 229-254

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995), *Microeconomic Theory*, Oxford University Press, New York

McFadden D. (1999), "Rationality for Economists?", *Journal of Risk and Uncertainty*, 19 (1), pp. 73-105

McFadden D. (2001), "Economic Choices", *The American Economic Review*, 91 (3), pp. 351-378

McFadden D. and Train K. (2000), "Mixed MNL models for discrete response", *Journal of Applied Econometrics*, 15 (5), pp. 447-470

McIntosh E. and Ryan M. (2002), "Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences", *Journal of Economic Psychology*, 23 (3), pp. 367-382

Meijer E. and Rouwendal J. (2006), "Measuring welfare effects in models with random coefficients", *Journal of Applied Econometrics*, 21 (2), pp. 227-244

Meyerhoff J. and Liebe U. (2009), "Status Quo Effect in Choice Experiments: Empirical Evidence on Attitudes and Choice Task Complexity", *Land Economics*, 85 (3), pp. 515-528

Mitchell, R. C. & Carson, R. (1989), *Using surveys to value public goods: the contingent valuation method*, Resources for the future, Washington

Moore R., Bishop R. C., Provencher B. and Champ P. A. (2010), "Accounting for Respondent Uncertainty to Improve Willingness-to-Pay Estimates", *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 58 (3), pp. 381-401

Morkbak M., Christensen T. and Gyrd-Hansen D. (2010), "Choke Price Bias in Choice Experiments", *Environmental and Resource Economics*, 45 (4), pp. 537-551

Murphy J., Allen P., Stevens T. and Weatherhead D. (2005), "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation", *Environmental and Resource Economics*, 30 (3), pp. 313-325

NGENE (2010), *NGENE 1.0.2 User manual & reference guide*.

Ohler T., Le A., Louviere J. and Swait J. (2000), "Attribute Range Effects in Binary Response Tasks", *Marketing Letters*, 11 (3), pp. 249-260

Olsen S. B., Lundhede T., Jacobsen J. and Thorsen B. (2011), "Tough and Easy Choices: Testing the Influence of Utility Difference on Stated Certainty-in-Choice in Choice Experiments", *Environmental and Resource Economics*, 49 (4), pp. 491-510

Ortuzar, J. d. D., Roncagliolo, D. A., and Velarde, U. (2000), "Interactions and independence in stated preference modelling," in *Stated preference modelling techniques*, J. d. D. Ortuzar, ed., PTRC Education and research services Ltd., London, pp. 155-172

- Palma A. d., Gordon M. M. and Papageorgiou Y. Y. (1994), "Rational Choice Under an Imperfect Ability To Choose", *The American Economic Review*, 84 (3), pp. 419-440
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993), *The adaptive decision maker*, Cambridge University Press
- PBL (2010), Evaluatie van een IPCC-klimaatrapport: een analyse van conclusies over de mogelijke gevolgen van klimaatverandering, *Planbureau voor de Leefomgeving 500216001 (in Dutch)*
- Pearce, D. W. and Smale, R. (2005), "Appraising flood control investments in the UK," in *Cost-benefit analysis and water resource management*, R. Brouwer & D. W. Pearce, eds., Edward Elgar, Cheltenham, UK
- Perman, R., Common, M., Mcgilvray, J., & Ma, Y. (2003), *Natural Resource and Environmental Economics*, 3 edn, Addison Wesley
- Plott, C. R. (1996), "Rational individual behavior in markets and social choice processes: The discovered preference hypothesis," in *Rational foundations of economic behavior*, K. Arrow et al., eds., Macmillan and St. Martin's, London, pp. 225-250
- Poe G. L., Giraud K. L. and Loomis J. B. (2005), "Computational Methods for Measuring the Difference of Empirical Distributions", *American Journal of Agricultural Economics*, 87 (2), pp. 353-365
- Racine J. S., Hart J. and Li Q. (2006), "Testing the Significance of Categorical Predictor Variables in Nonparametric Regression Models", *Econometric Reviews*, 25 (4), pp. 523-544
- Ready R. C., Champ P. A. and Lawton J. L. (2010), "Using Respondent Uncertainty to Mitigate Hypothetical Bias in a Stated Choice Experiment", *Land Economics*, 86 (2), pp. 363-381
- Ready R. C., Navrud S. and Dubourg W. R. (2001), "How Do Respondents with Uncertain Willingness to Pay Answer Contingent Valuation Questions?", *Land Economics*, 77 (3), pp. 315-326
- Revelt D. and Train K. (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", *Review of Economics and Statistics*, 80 (4), pp. 647-657
- Rieskamp J., Busemeyer J. R. and Mellers B. A. (2006), "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice", *Journal of Economic Literature*, 44 (3), pp. 631-661
- Rigby D. and Burton M. (2006), "Modeling Disinterest and Dislike: A Bounded Bayesian Mixed Logit Model of the UK Market for GM Food", *Environmental and Resource Economics*, 33 (4), pp. 485-509
- Rijkswaterstaat (2005a), HIS - Schade en Slachtoffer Module Versie 2.1: Gebruikershandleiding, *DWW-2005-004 (in Dutch)*
- Rijkswaterstaat (2005b), Veiligheid Nederland in Kaart: Overstromingsrisico dijkkring 13 Noord-Holland, Rijkswaterstaat, *DWW-2006-023 (in Dutch)*
- Rijkswaterstaat (2005c), Veiligheid Nederland in Kaart: Overstromingsrisico dijkkring 14 Zuid-Holland, Rijkswaterstaat, *DWW-2006-023 (in Dutch)*
- Rijkswaterstaat (2008), Overstromingsscenario's voor rampenplannen: Westelijke kust, Rijkswaterstaat, *(in Dutch)*

Rose J. M. and Bliemer M. C. J. (2009), "Constructing Efficient Stated Choice Experimental Designs", *Transport Reviews: A Transnational Transdisciplinary Journal*, 29 (5), pp. 587-617

Rose J. M., Bliemer M. C. J., Hensher D. A. and Collins A. T. (2008), "Designing efficient stated choice experiments in the presence of reference alternatives", *Transportation Research Part B: Methodological*, 42 (4), pp. 395-406

Rose J. M., Hensher D. A., Caussade S., Ort-zar J. d. D. and Jou R. C. (2009), "Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis", *Journal of Transport Geography*, 17 (1), pp. 21-29

Rose, J. M., Scarpa, R., MacDonald, D. H., Morrison, M., & Boyle, K. (2011), Accounting for error heteroskedasticity in logit models for multi-attribute stated preference data.

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005), Bayesian statistics and marketing, John Wiley & Sons Ltd., Chichester, UK

Rouwendal J., de Blaeij A., Rietveld P. and Verhoef E. (2010), "The information content of a stated choice experiment: A new method and its application to the value of a statistical life", *Transportation Research Part B: Methodological*, 44 (1), pp. 136-151

Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008), Using Discrete Choice Experiments to Value Health and Health Care, Springer, Dordrecht

Salisbury L. C. and Feinberg F. M. (2010), "Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test", *Marketing Science*, 29 (1), pp. 1-17

Samnaliev M., Stevens T. H. and More T. (2006), "A comparison of alternative certainty calibration techniques in contingent valuation", *Ecological Economics*, 57 (3), pp. 507-519

Savage S. J. and Waldman D. M. (2008), "Learning and fatigue during choice experiments: a comparison of online and mail survey modes", *Journal of Applied Econometrics*, 23 (3), pp. 351-371

Scarpa, R., Ferrini, S., and Willis, K. (2005), "Performance of Error Component Models for Status-Quo Effects in Choice Experiments," in *Applications of Simulation Methods in Environmental and Resource Economics*, 6 edn, R. Scarpa & A. Alberini, eds., Springer Netherlands, pp. 247-273

Scarpa R. and Rose J. M. (2008), "Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why\*", *Australian Journal of Agricultural and Resource Economics*, 52 (3), pp. 253-282

Scarpa R., Thiene M. and Train K. (2008), "Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps", *American Journal of Agricultural Economics*, 90 (4), pp. 994-1010

Shaikh S. L., Sun L. and Cornelis van Kooten G. (2007), "Treating respondent uncertainty in contingent valuation: A comparison of empirical treatments", *Ecological Economics*, 62 (1), pp. 115-125

Shannon C. E. (1948), "A mathematical theory of communication", *Bell System Technical Journal*, 27, pp. 379-423

Shaw W. D. and Woodward R. T. (2008), "Why environmental and resource economists should care about non-expected utility models", *Resource and Energy Economics*, 30 (1), pp. 66-89

- Sjoberg L. (2000), "Consequences matter, risk is marginal", *Journal of Risk Research*, 3 (3), pp. 287-295
- Slovic P. (1987), "Perception of risk", *Science*, 236 (4799), pp. 280-285
- Slovic P. (1995), "The construction of preference", *American Psychologist*, 50 (5), pp. 364-371
- Stb (1996), Wet op de waterkering, *Staatsblad van het koninkrijk der Nederlanden 8 (in Dutch)*
- Stopher P. and Hensher D. (2000), "Are More Profiles Better Than Fewer?: Searching for Parsimony and Relevance in Stated Choice Experiments", *Transportation Research Record: Journal of the Transportation Research Board*, 1719 (-1), pp. 165-174
- Sun L. and van Kooten G. (2009), "Comparing Fuzzy and Probabilistic Approaches to Preference Uncertainty in Non-Market Valuation", *Environmental and Resource Economics*, 42 (4), pp. 471-489
- Swait J. and Adamowicz W. (2001a), "Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice", *Organizational Behavior and Human Decision Processes*, 86 (2), pp. 141-167
- Swait J. and Adamowicz W. (2001b), "The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching", *The Journal of Consumer Research*, 28 (1), pp. 135-148
- Swait J. and Louviere J. (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models", *Journal of Marketing Research*, 30 (3), pp. 305-314
- Train K. E. (2008), "EM-algorithms for nonparametric estimation of mixing distributions", *Journal of Choice Modelling*, 1 (1), pp. 40-69
- Train, K. E. (2009), *Discrete choice models with simulation*, 2nd Edition edn, Cambridge University Press
- Train, K. and Sonnier, G. (2005), "Mixed Logit with Bounded Distributions of Correlated Partworths," in *Applications of Simulation Methods in Environmental and Resource Economics*, 6 edn, R. Scarpa & A. Alberini, eds., Springer Netherlands, pp. 117-134
- Train, K. and Weeks, M. (2005), "Discrete Choice Models in Preference Space and Willingness-to-Pay Space," in *Applications of Simulation Methods in Environmental and Resource Economics*, 6 edn, R. Scarpa & A. Alberini, eds., Springer Netherlands, pp. 1-16
- Tseng Y. Y. and Verhoef E. T. (2008), "Value of time by time of day: A stated-preference study", *Transportation Research Part B: Methodological*, 42 (7-8), pp. 607-618
- van den Berg, V. A., Kroes, E., & Verhoef, E. T. (2010), Biases in Willingness-to-Pay Measures from Multinomial Logit Estimates Due to Unobserved Heterogeneity, SSRN, *Tinbergen Institute Discussion Paper # 10-014/3*
- van Dyk D. A. and Meng X. L. (2001), "The Art of Data Augmentation", *Journal of Computational and Graphical Statistics*, 10 (1), pp. 1-50
- van Kooten G. C., Krcmar E. and Bulte E. H. (2001a), "Preference Uncertainty in Non-Market Valuation: A Fuzzy Approach", *American Journal of Agricultural Economics*, 83 (3), pp. 487-500
- van Kooten G. C., Krcmar E. and Bulte E. H. (2001b), "Preference Uncertainty in Non-Market Valuation: A Fuzzy Approach", *American Journal of Agricultural Economics*, 83 (3), pp. 487-500

- van Kooten G. C., Krcmar E. and Bulte E. H. (2001c), "Preference Uncertainty in Non-Market Valuation: A Fuzzy Approach", *American Journal of Agricultural Economics*, 83 (3), pp. 487-500
- Vazquez M. X., Arana J. E. and Leon C. J. (2006), "Economic evaluation of health effects with preference imprecision", *Health Economics*, 15 (4), pp. 403-417
- Vermeulen B., Goos P., Scarpa R. and Vandebroek M. (2011), "Bayesian Conjoint Choice Designs for Measuring Willingness to Pay", *Environmental and Resource Economics*, 48 (1), pp. 129-149
- Wang H. (1997), "Treatment of "Don't-Know" Responses in Contingent Valuation Surveys: A Random Valuation Model", *Journal of Environmental Economics and Management*, 32 (2), pp. 219-232
- Welsh M. P. and Poe G. L. (1998), "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach", *Journal of Environmental Economics and Management*, 36 (2), pp. 170-185
- Whitehead J. C., Blomquist G. C., Hoban T. J. and Clifford W. B. (1995), "Assessing the Validity and Reliability of Contingent Values: A Comparison of On-Site Users, Off-Site Users, and Non-users", *Journal of Environmental Economics and Management*, 29 (2), pp. 238-251
- Yanez M. F., Raveau S. and Ortuzar J. d. D. (2010), "Inclusion of latent variables in Mixed Logit models: Modelling and forecasting", *Transportation Research Part A: Policy and Practice*, 44 (9), pp. 744-753