Using discrete choice experiments to inform environmental policy in a developing country context

Case studies from Ethiopia

Solomon Tarfasa Faro
Using discrete choice experiments to inform environmental policy in a developing country context: Case studies from Ethiopia

PhD Dissertation, Vrije Universiteit Amsterdam, The Netherlands

Solomon Tarfasa, Amsterdam, October 30, 2018

The research for this dissertation was carried out at the Institute for Environmental Studies (IVM), Department of Environmental Economics, Vrije Universiteit, Amsterdam.
Using discrete choice experiments to inform environmental policy in a developing country context

Case studies from Ethiopia
promotor: prof.dr. R. Brouwer
Table of Contents

Acknowledgements........................................................................................................ 7

Summary..................................................................................................................... 9

Abstract.................................................................................................................... 15

1. Introduction........................................................................................................... 21
   1.1 Background and problem statement............................................................... 21
   1.2 Main objective, research questions and novelty.......................................... 24
   1.3 Discrete choice experiments........................................................................... 26
   1.4 Random utility model....................................................................................... 30
   1.5 Outline of the PhD dissertation....................................................................... 36

2 Public preferences for improved urban waste management............................ 39
   2.1 Introduction....................................................................................................... 39
   2.2 The choice experiment...................................................................................... 42
   2.3 The choice model............................................................................................. 45
   2.4 Results.............................................................................................................. 48
   2.5 Discussion and conclusions............................................................................ 59

3 Estimation of the public benefits of urban water supply improvements.62
   3.1 Introduction...................................................................................................... 62
   3.2 The choice model............................................................................................ 65
   3.3 The experiment................................................................................................ 67
   3.4 Results.............................................................................................................. 70
   3.5 Conclusions..................................................................................................... 79
Informing water harvesting technology contract design using choice experiments

4.1 Introduction
4.2 Methodological design
4.3 Case study area and data collection
4.4 Results
4.5 Conclusions

5 Testing hypothetical bias in an agricultural field experiment

5.1 Introduction
5.2 Experimental design and data collection
5.3 Results
5.4 Conclusions and discussion

6 Conclusions and recommendations

6.1 Methodological advancements and future research
6.2 Policy recommendations

References
Acknowledgements

You can imagine the confidence it gives when an incidental contact in Addis Ababa ends up with a professor offering a PhD study under his supervision at a prestigious university in Europe. This is how my PhD study started at the Vrije Universiteit under the supervision of Professor Roy Brouwer. I would like to thank Professor Roy Brouwer for his excellent advice guiding me through the technically demanding and conceptually challenging PhD study from the beginning to the end. Professor Roy led me through the necessary critical thinking for doing a PhD, helped me improve my academic writing, effective communication, understanding, punctuality and systematic analysis. He covered my expenses from his own money during my PhD study since I had never been part of an official PhD project. Professor Roy is a great academic advisor. I will always admire your commitment to bring the best out of your students and you will remain my role model for the rest of my life. I have no doubt that I would not have come to this moment in my PhD study had professor Roy not been my advisor. I faced the most difficult moment in my life while doing this PhD study.

I would also like to thank all those in the Vrije Universiteit who supported me in one way or another during my PhD study under the supervision of Professor Roy. I am very grateful to Oleg Sheremet for his help with the experimental design of one of my PhD chapters and his willingness to be a co-author on that paper. This was a great opportunity to learn a lot from working with him.

I am also grateful to Dr. Jetske Bouma for her help in getting funding for my survey as part of the project WHaTeR funded under the European Union’s Seventh Framework Programme (grant agreement number 266360).
I would like to thank Marjolijn Staarink and Suzan Besuijen for their kindness and help in administrative matters during my stay at the Institute for Environmental Studies (IVM) in Amsterdam. I will never forget that Suzan gave me a winter jacket when I arrived in Amsterdam and had no winter clothes with me the first time I came to visit the Vrije Universiteit in Amsterdam.

I am grateful to the Eawag Fellowship Program for Developing Countries, and for the opportunity they provided me to work on one of my PhD chapters in Switzerland and covering all my expenses during my 4 months stay at Eawag.

I would like to acknowledge the moral and social support provided by all colleagues at various circles back home. The list is endless but I will always treasure every moment of support.

Lastly, many thanks to my closest family, my wife Kidist, my daughters Hara and Ella, my brother Guta and Talile for bearing with me during the time that I could not be with you. I promise to spend more time with my family after this.
Summary

In order to progress towards achieving the sustainable development goals articulated most recently by the United Nations in 2015, it is paramount to improve the knowledge and information base for integrated environmental-economic policy and decision-making. The integration of socio-economic knowledge and information into decision-making surrounding environmental policies, programs or projects is expected to benefit in particular developing countries where there is a strong correlation between poverty, economic development and environmental degradation. Because of the public good nature of many if not most environment and development policies, programs and projects, information about their socio-economic impacts on different groups in society for inclusion in cost-benefit analysis will in most cases not be available by investigating traditional economic markets. Appropriate alternative non-market valuation methods are therefore needed instead. Stated preference approaches that directly ask potential beneficiaries of these policies, programs or projects in public surveys for the value they attach to the impacts on their own or community livelihoods, typically measured through their willingness to pay (WTP) in a simulated market, are the most widely applied non-market valuation method. Within the group of stated preference approaches, discrete choice experiments (DCEs) are nowadays most popular due to their convenient ability of eliciting public preferences for policy alternatives and their key characteristics, including the price of implementation. However, reviews of and guidelines for DCEs in developing countries seems to have followed a one-size-fits-all approach since they are often instigated by international donors and higher educated local students in developed countries from onlookers’ point of view, and homegrown expertise and institutional capacities are still limited.
The main objective of this PhD dissertation is to develop, apply and test the usefulness of tailored DCEs in informing environmental policy and decision-making in a developing country context in Ethiopia. Specific methodological issues are tested, in particular the role of income, literacy and communication in public understanding and learning in DCEs where survey participants are typically asked to answer hypothetical questions for more or less familiar goods and services in a series of choice tasks. The research is based on four case studies in Ethiopia, in contexts ranging from urban water supply to solid waste collection and the marketing of organic beans in a rural market.

The first case study in this dissertation attempted to elicit public preferences for improvements in solid waste services across different socio-economic zones in the city of Hawassa. Most attempts to improve solid waste management in cities in Ethiopia have focused on the technical supply aspects of the services, mostly ignoring the relevant demand side of the solid waste management services. This includes the ability to pay for the services in order for the service to be provided in a financially viable way based on full cost recovery and without relying on child labor. The results show that there exists substantial public WTP to improve solid waste management in the city Hawassa in terms of more frequent collection times and recyclable waste separation through the provision of waste bins. New in this study is the focus on child labor in the waste management sector. Interestingly, households are willing to pay a significant premium over and above existing service charges to abolish child labor in the waste sector, especially households with female household heads and higher income levels. As expected, respondents living in wealthier neighborhoods are more likely to be willing and able to pay higher service charges. These results provide important information to municipality officials and the waste management sector as these aspects can be targeted more specifically to generate
additional revenues and design appropriate strategies to improve existing solid waste management services.

In the second case study in this PhD dissertation, households’ WTP for improved urban drinking water supply services in the city of Hawassa are examined. In Ethiopia, drinking water supply coverage is among the lowest in the Sub-Saharan Africa countries. Cost recovery rates of these services are typically low, while demand for more reliable services is high and rapidly growing. This case study employs DCEs to estimate the non-market value of specific water supply service improvements related to water safety and more frequent water supply. The findings from this case study indicate that households are willing to pay extra for improved levels of water supply over and above their current water bill, in particular households living in the poorest neighborhoods of the city with the lowest service levels. Women value the improvement of water quality most. Respondents who spend on average more money on bottled water value water quality improvements more than those who spend less on bottled water. New in this study from a methodological point of view is the preference learning detected in survey participants’ choice behavior when going through a sequence of hypothetical choice tasks. The estimated economic values can be used in policy appraisals of improved drinking water supply investment decisions in the future.

The third case study in this PhD dissertation assesses through a contractual agreement what technical, institutional and economic conditions need to be in place to encourage farm households in Halaba, Ethiopia to invest in and maintain water ponds on their land and reduce their vulnerability to droughts. Previous attempts to build heavily subsidized water ponds in the area largely failed. The novelty of the study is that we use a DCE to identify relevant
contract characteristics, to which farmers attach most importance when aiming to improve water management in dryland agriculture, and support effective uptake of contractual agreements through adequate communication. The results in this study show that demand for the offered contractual agreements is high and farmers are willing to pay market interest rate levels to obtain the financial support (micro credit) for such an investment decision. Given the high illiteracy rate in the study area and the uncertainty whether or not the sampled households understand the specific contractual terms and conditions, visual aids were used in a split sample approach. The samples with and without the visual aids generated significantly different results. The sample who received the version of the CE with the visualizations made more stable and less random choices, highlighting the importance of how information is conveyed to survey participants in a developing country context. Moreover, the results indicate that some of the contract characteristics offered were considered more important than others, with the water harvesting technology such as pond lining, cover and capacity attributes valued the highest. In conclusion, the use of visual aids in CEs has a significant effect on choice behavior and ultimately on the economic values derived from the estimated choice models in this study carried out in a remote developing country context. The empirical evidence of such effects in the existing literature is very limited and deserves more attention in future stated preferences research.

The fourth and final case study in this PhD dissertation tests hypothetical bias by comparing preferences for attributes describing red haricot beans, such as whether the beans are produced conventionally or organically, newly introduced or local, cultivated through monoculture or mixed with other crops, using hypothetical and real choices on a local agricultural market in Ethiopia. New in this study is the inter-respondent and intra-respondent comparison of choice
behavior. The former comparison consists of a control group of buyers who are only asked to hypothetically choose in a DCE between conventional and organic haricot beans, and an experimental group of buyers who are financially endowed to actually purchase the same conventional and organic haricot beans using an identical experimental design. In the latter comparison, a random sample of market visitors is first asked in a hypothetical discrete choice experiment to choose between conventional and organic red haricot beans given different price levels, and then endowed with a lump sum of money to actually purchase the same conventional and organic red haricot beans using exactly the same experimental design. Differences in choices, preference parameters and WTP are tested. In addition, factors explaining hypothetical bias are examined. A key outcome of this study is that there is a negative hypothetical bias, i.e. actual WTP appears to be higher than hypothetical WTP, but differences between hypothetical and real food purchases are only significant in the case of the inter-respondent comparison, not in the intra-respondent comparison. This study is the first to find evidence of negative hypothetical bias in this specific field of application when comparing the collected hypothetical and real choice data. This puts the general validity and reliability of results from previous tests of hypothetical bias applying inter-respondent comparisons into question. Possible explanations for the negative hypothetical bias found in this study are mainly speculative, as underlying drivers were not further investigated here.

A key outcome of the DCEs developed and applied in this PhD study is the generation of benefit estimates for use in cost-benefit analysis of public and private investment decisions in urban waste collection and drinking water supply in the city Hawassa, water harvesting ponds on private land and organic beans sold on a rural agricultural market in the Southern Nations, Nationalities and Peoples’ Region in Ethiopia, of which Hawassa is the capital city. Moreover,
the studies investigating public WTP for urban waste collection and drinking water supply inform policy and decision-makers at the same time about the expected financial cost recovery of these investment decisions. The practical relevance of the results from the water harvesting survey is that they help regional policymakers identify how water harvesting can be rolled out as a farm household climate risk adaptation and mitigation strategy, while the results from the haricot beans study could guide rural agricultural policy by providing new insights into what matters in consumers’ decisions with respect to the crop supplementing efforts by existing government research centers in rural agricultural development programs.

In summary, DCEs are applied across many fields in developed countries and could play a similar role in developing countries. To this end, this dissertation explored the applicability of DCEs in a series of case studies in Ethiopia. The fact that the case studies in this dissertation are all from Ethiopia and within specific fields of environmental and agricultural economics limits the generalization of the case study findings to other sub-Saharan countries and other fields of application. More studies in other African countries to assess the representativeness of the outcomes presented here are needed, as well as a broadening of the scope of applications in order to be able to generalize the present case study findings to other areas and developing countries.
Abstract

The desirability of quantifying the often intangible impacts of projects, programs or policies on the well-being of the public in monetary terms, where relevant and practicable, is an emerging theme in the Developing World. However, many of these impacts concern public goods and services which are not traded in traditional markets. This means that the value the public places on these impacts cannot simply be observed with the help of market information, such as price and consumption levels. This has given rise to the proliferation of methods that have sought to uncover, in a variety of ways, the value of these nonmarket goods and services.

One of the most prominent non-market valuation techniques are stated preference approaches that directly ask participants in public surveys to state their willingness to pay (WTP) or willingness to accept (WTA) compensation for an environmental change in a hypothetical or simulated market scenario. Stated preference approaches are survey-based and elicit people’s intended future behavior, typically in currently non-existing or poorly developed markets. By means of an appropriately designed questionnaire, a hypothetical market is described where the good in question is traded. This hypothetical market defines the good itself, the institutional context in which it would be provided, and the way it would be financed.

The most widely used variant of stated preference methods nowadays are discrete choice experiments (DCEs). DCEs have the convenient property that they can elicit public preferences for policy alternatives and their key characteristics, including their price. This PhD dissertation presents the use of DCEs to inform environmental policy and decision-making in a developing country context, illustrating their challenges and usefulness in four different case studies.
carried out in Ethiopia. The main objective of this PhD dissertation is to develop, apply and test the usefulness of DCEs in informing environmental policy and decision-making in Ethiopia. Specific methodological issues are tested, in particular the role of income, literacy and communication in public understanding and learning in DCEs where survey participants are typically asked to answer hypothetical questions for more or less familiar goods and services in a series of choice tasks. The contexts in which the DCEs are applied in this PhD thesis are somewhat heterogeneous, ranging from urban water supply to solid waste collection and the marketing of organic beans in a rural market in Ethiopia. The same applies to some extent to the discrete choice models that are used to analyze them, although these models are all rooted in the same random utility modelling theory.

In the first case study in this dissertation, a DCE aiming to elicit public preferences for improvements in solid waste services is administered across different socio-economic zones in the city of Hawassa. The study provides important insight into the social benefits of public investment decisions to improve the quality of solid waste management services in large cities in Ethiopia. Only a limited number of stated preference studies have been conducted in developing countries so far focusing on solid waste collection, applying almost all the contingent valuation method, of which three were carried out in Addis Abeba, Mekelle and Bahir Dar in Ethiopia. In the case study in Hawassa, observed and unobserved preference heterogeneity are analyzed using mixed logit choice models. The results show that there exists substantial public WTP to increase collection frequency and separate recyclable waste. New is the focus on child labor in the waste management sector. Significant gender effects are furthermore found: women are more interested than men in increasing waste collection frequency and value the abolishment of child labor higher, just as higher income households.
As expected, respondents living in wealthier neighborhoods are more likely to pay higher service charges.

The second study in this PhD dissertation examines households’ WTP for improved urban drinking water supply services in the city of Hawassa, where the estimated economic values can be used in policy appraisals of improved drinking water supply investment decisions. Ethiopia is among the countries in Sub-Saharan Africa with the lowest water supply coverage. Improving existing drinking water supply services in developing countries depends crucially on available financial resources. Cost recovery rates of these services are typically low, while demand for more reliable services is high and rapidly growing. Most stated preference based demand studies in the developing world apply the contingent valuation method and focus on rural areas. Contrary to contingent valuation, DCEs allow for the estimation of the multiple characteristics or attributes of drinking water supply, i.e. supply reliability and water safety. The findings in this case study indicate that, despite significant income constraints, households are willing to pay up to 80 percent extra for improved levels of water supply over and above their current water bill, especially households living in the poorest parts of the city with the lowest service levels. Women value the improvement of water quality most, while a significant negative effect is found, as expected, for averting behavior and expenditures. New in this study from a methodological point of view is the preference learning detected in survey participants’ choice behavior when going through a sequence of hypothetical choice tasks.

The third case study in this PhD dissertation informs policymakers in Ethiopia about the most important terms and conditions to incentivize farmers to enter into a contractual agreement to invest in water harvesting on their land. The uptake of water-harvesting technology is expected to be more effective and last longer if farm households are involved in their design. In order to
test the influence of the way the specific contractual terms and conditions are communicated to farm households, many of whom are illiterate, a split sample approach is applied with and without visual aids for technical, institutional and economic contract characteristics. The findings show that both samples generate significantly different results, highlighting the importance of how information is conveyed to farm households. This pattern is confirmed when examining the self-reported importance attached to the various contract characteristics. Equality constrained latent class models show that the contract characteristics for which visual aids were developed are considered more attentively, emphasizing the importance of adequate communication tools in a developing country context where literacy rates are limited to increase water technology innovation uptake and reduce farm household vulnerability to droughts.

In the fourth and final case study in this PhD dissertation, one of the main criticisms underlying stated preference approaches like DCEs is tested, namely hypothetical bias. New in this study is the inter-respondent and intra-respondent comparisons of choice behavior, applying a hypothetical and real choice experiment in a local agricultural market in Ethiopia. The inter-respondent comparison commonly applied in the literature consists of a control group of buyers who are asked to hypothetically choose in a DCE between conventional and organic haricot beans and an experimental group of buyers who are financially endowed to actually purchase the same conventional and organic haricot beans using an identical experimental design. Hypothetical bias is tested by comparing (i) hypothetical and real choices, (ii) preference parameters of the estimated choice models related to the hypothetical and real choices, and (iii) hypothetical and real willingness to pay (WTP). In addition, factors explaining the observed hypothetical bias are examined. Controlling for potential choice sequencing and associated
preference learning, actual WTP appears to be higher than hypothetical WTP, but differences between hypothetical and real food purchases are only significant in the case of the inter-respondent comparison, not the intra-respondent comparison. This suggests that results from previous tests of hypothetical bias applying inter-respondent comparisons may have to be reconsidered.

In conclusion, the four case studies each contribute in their own way to some of the main methodological challenges encountered in stated preferences research in the context of a developing country where willingness to pay is severely constrained by ability to pay and literacy affects the way the public is able to participate in social research and surveys. The four studies at the same time fill important gaps to integrate non-market values in mainstream policy and decision-making. Policymaker understanding of these nonmarket values is considered paramount in sustainable economic transitions, accounting for the non-monetary social and environmental costs and benefits associated with investment decision-making.
1. Introduction

1.1 Background and problem statement

In the early days GDP-based global economy, built or human made capital was the main limiting factor while natural capital was relatively speaking abundantly available. It made sense in that context to think of the economy as primarily consisting of marketed goods and services and to think of the most important social objective as increasing the production of these goods and services for consumption (Costanza et al., 2014). The environmental Kuznet hypothesis, taken to its extreme, suggests that the poorest nations in the world do not regard the environment as anything worth preserving. However, as people get richer and GDP continues to rise, their demand for environmental quality increases, initially through the adoption of public health legislation and demand for clean water and air, and ultimately the conservation of nature more generally (Barbier, 2005). It is becoming increasingly clear that the current environmental crisis (e.g. global climate change, water scarcity, soil and air pollution, waste) is unequivocally caused by human economic activities that disregard the environment, and there is growing recognition that this crisis hinders progress towards achieving the sustainable development goals as articulated most recently by the United Nations in 2015. Moreover, it is highly likely that allowing environmental degradation to proceed along with economic growth might breach irreversible thresholds (Bennett and Birol, 2010). It is therefore necessary to integrate environment and development at policy, planning, and management levels and improve the knowledge and information base for decision-making (Ruffeis et al., 2010). In other words, the integration of the trade-offs between the economic, environmental and social impacts of policies, programs and projects into decision-making processes is expected to significantly contribute to efforts to mitigate further environmental degradation as a result of the global development challenges. This thereby preserves natural systems essential to life and
well-being and keeps the balance between development and conservation (Costanza et al., 2012). This integration into decision-making surrounding environmental policies, programs or projects is expected to benefit in particular developing countries that suffer most from environment-related challenges such as water scarcity, air pollution, soil degradation and associated ecosystem services and productivity losses. These challenges impede the eradication of poverty in these countries where there is a strong correlation between poverty and environmental degradation (Shifera et al., 2005). One way of including the environmental and social impacts of policies, programs and projects into decision-making processes is by applying environmental non-market valuation in their cost-benefit assessments.

According to Bennett and Birol (2010), the integration of environmental non-market values into decision-making processes has been one of the growth areas in the formulation of policy advice in developed countries over the past decades. This integration was triggered by a number of driving forces. First, the growth in policy maker demand for information about the non-market effects of their policies, especially on the environment and different groups in society. Second, the improvements in the validity and reliability on the supply side of valuation studies. This includes McFadden's (1974) development of the random utility model (RUM) that underpins most of the stated preference techniques nowadays for estimating non-market values (Johnston et al., 2017), and the increasing empirical evidence base of non-market values worldwide.

Non-market valuation has a just as important role to play in the developing world as in developed countries in estimating the economic value of natural resources and environmental change (Georgiou et al., 1997). The valuation methods that are being used are not substantially
different, just the context in which they are applied (Hanley and Barbier, 2009). The increasing scarcity of environmental assets and growing public and policymaker awareness of their importance facilitate in developing countries the translation of these changes in public preferences and environmental non-market values. The dependency of large numbers of especially the rural poor on fragile environments and critical environmental goods and services means that proper valuation of their benefits is paramount. Moreover, there is a relatively higher mortality associated with certain environmental risk factors in developing economies, such as severe droughts and water shortages, unsafe drinking water, urban air pollution, sanitation and hygiene. Valuation methods are increasingly employed to assess public WTP for reducing these risks as well as for improving environmental quality in general (Birol and Bennett, 2010). Most importantly, the application of non-market valuation techniques in developing countries offers not only remedies to avoid the pitfalls developed countries have encountered through their growth phase, i.e. continued emphasis on GDP as the only indicator of social progress, but also a potential 'short cut' to the development process where economic growth, social well-being and environmental improvement can be achieved simultaneously.

Several noteworthy studies employing discrete choice experiments have been carried out in developing countries in recent years, including in Ethiopia (see for example, Scarpa et al., 2003a, 2003b; Othman et al., 2004; Naido and Adamowicz, 2005; Bienabe and Hearne, 2006; Asrat et al., 2007; Tesfaye and Brouwer, 2012). However, a review of the application of these studies in developing countries reveals that homegrown expertise and institutional capacities are still limited, since they are often instigated by international donors and higher educated local students in developed countries (Bennett and Birol, 2010). DCEs in developing countries seem to have followed a one-size-fit-all approach as many of these studies are designed by
experts or consultants coming from the Western world (Khorshed, 2005). Informing environmental policy in developing countries using non-market valuation approaches such as DCEs that are able to integrate environmental concerns into mainstream policy and decision-making is still rather weak. As a result, developing countries do not benefit from avoiding the trap developed countries encountered on their growth path. In general, the improvements that DCEs bring to the analysis of environmental policy have not been fully exploited in developing counties. Moreover, the prospects for choice experiment results to have a positive impact on resource use efficiency in developing countries means that its application can make a real difference to the lives of the people who currently suffer deprivation (Birol and Bennett, 2010).

1.2 Main objective, research questions and novelty

The main objective of this PhD dissertation is to develop, apply and test the usefulness of DCEs in informing environmental policy and decision-making in Ethiopia. This is done in four different case studies pertaining to urban and rural water supply, urban waste management, and organic farming. Specific methodological issues are tested in a developing country context, in particular the role of income, education (literacy) and communication in public understanding and learning in DCEs where survey participants are typically asked to answer hypothetical questions for more or less familiar public goods and services in a series of choice tasks. The way choice alternatives and their attributes are presented to survey participants is furthermore expected to influence their attendance, which is in turn expected to be influenced by literacy. Finally, the issue of hypothetical bias, one of the main criticisms of stated preference methods such as DCEs, is addressed in one of the case studies by comparing real and hypothetical purchase behavior. The structure and design of the PhD dissertation is visualized in Figure 1.
This dissertation aims to answer the following specific research questions:

1. How stable are choices of DCE participants when subjected to a sequence of choice tasks (Chapters 2 and 3)?

2. What is the influence of visual aids on choice behavior and attribute attendance of literate and illiterate DCE participants (Chapter 4)?

3. To what extent does hypothetical bias in DCEs differ in intra- and inter-respondent comparisons of stated and revealed preferences (Chapter 5)?

The first research question is not specifically new in the DCE literature (see, for example, Brouwer et al. (2010) for an application in the developed world). The second and third research question, however, have not been answered in the DCE literature before, neither in the
developed or developing world. Hence, besides presenting some of the first applications of DCEs in Ethiopia in the field of urban and rural water supply, urban solid waste management and the role of child labor, and organic farming, especially new in this PhD study are the (i) tests of the influence of visual aids on attribute attendance in DCEs and (ii) tests of intra-respondent and inter-respondent measures of hypothetical bias.

1.3 Discrete choice experiments

DCEs were initially developed by Louviere and Hensher (1982) and Louviere and Woodworth (1983). They share the Random Utility Model (RUM) (Luce, 1959; McFadden, 1974) as a common theoretical framework with dichotomous-choice contingent valuation, as well as a common basis of empirical analysis in limited dependent variable econometrics (Pearce et al., 2006). DCEs are member of the stated preferences methods used for non-market valuation, and were originally developed in the marketing and transport economics literature (see, for example, Louviere and Hensher, 1982; Louviere and Woodworth, 1983; Louviere, 1998; Louviere, 1992), but have been increasingly also adopted in other fields over the last two decades, such as environment, health and agriculture. The theoretical grounding of this method lies in Lancaster’s (1966) theory of value and its econometrics basis in random utility theory (Thurstone, 1927; Manski, 1977). DCEs are based on the notion that any good, service, policy or program can be described in terms of its characteristics or attributes and the levels that these attributes take, with or without interventions in the current situation or status quo. Once attributes and their levels are identified, experimental design theory is used to generate different profiles of the good, service, policy or program in terms of their attribute levels (Bennett and Birol, 2010). In general, a DCE consists of the following elements:
(1) a set of fixed choice options or subsets of the total options that may have explicit names or labels, such as brands or types of habits, or may simply be generic, identified by uninformative labels such as option A and option B, or alternative 1 and alternative 2, etc.;

(2) a set of attributes that describe potential differences in the choice options, and typically are included in a research project because they are expected to play a significant role in the choice behavior of interest and underlying preferences;

(3) a set of levels or values assigned to each attribute of each choice option to represent the relevant range of variation in that attribute for the study’s specific research or policy objectives. Different attributes may have different levels depending on the research objectives and some attributes may be binary, while others are discrete or continuous;

(4) a sample of subjects that evaluates all or a subset of the choice sets in the DCE and chooses one of the possible options available in each choice set. The choices supplied by the sample of subjects are used to estimate various types of choice models (Bennett and Blamey, 2001).

One of the attributes included in a DCE is monetary, such as the price for a private good or a tax for a public good. The monetary attribute and the random utility framework on which the DCE is based allow for the estimation of welfare estimates, that is, WTP or WTA compensation for changes in the levels of the attributes of the goods, services, policy or programs under evaluation.
Hanley et al. (1998a) define the DCE as “a highly structured method of data generation”. According to Bennett and Birol (2010), the DCE can equip decision-makers with four types of information that can inform the development of more comprehensive cost-benefit analysis. These include:

1. which attributes are significant determinants of the values that stakeholders (for example, local or national public, consumers of a product or service, farmers, visitors to a recreational site) place on the goods, services, policies or programs;
2. the implied rankings of these attributes amongst the relevant stakeholders;
3. the value of changing one or more of the attributes at once; and
4. the total economic value of the good, service, policy or program to each stakeholder group and the overall population to which these stakeholders belong.

DCEs elicit preferences from the public, consumers or other stakeholders, providing information on the trade-offs they are willing to make between attributes of the alternatives they are asked to choose from. Once the behavioral model is developed, behavioral predictions can be made by estimating the trade-offs involved. The most important aspect of discrete choice modeling efforts in the analysis of choice behavior is that the task reflects the decision process of the respondent and that the attributes are constructed in a fashion that is consistent with the actual behavioral process being examined (Bennett and Blamey, 2001). Discrete choice models (DCMs) are based on the concept of utility maximization in economics where decision-makers are assumed to be rational, making choices which maximize the utilities they obtain from the alternatives. The alternatives are supposed to be mutually exclusive and collectively
exhaustive, while the rationality of the decision maker implies transitive and coherent preferences (Sorci et al., 2010).

A further application of DCMs is that it can be used to augment the well-known revealed preferences (RP) models to provide the flexibility of DCE data across a broader range than the current market, as well as the anchoring of stated choices on actual market behavior. In RP models, actual decisions made by consumers are observed and preferences are inferred from what was chosen and rejected (Adamowicz et al., 1994). The breakthrough is the recognition that preference data sources may differ primarily in the (co)variance content of the information captured by the random component of utility. If differences in variability can be identified and one data set can be rescaled relative to another to satisfy the covariance condition, then data sets can be pooled and combined and enrich the behavioral choice analysis (Louviere et al., 2000).

The popular enrichment strategy typically centers on combining revealed preference (RP) and stated preference (SP) data, although some studies combine multiple stated preference data sets. The appeal of combining RP and SP data is based on the premise that SP data are particularly good at improving the behavioral value of the parameters representing the relative importance of attributes in influencing choice, and hence increasing the usefulness of resulting marginal rates of substitution (measuring trade-offs) between pairs of attributes associated with an alternative. RP data, however, are more useful in predicting behavioral response in real markets in which new alternatives are introduced or existing alternatives are evaluated at different attribute levels (Louviere et al., 2000). In general, any preference elicitation method
that provides information about preference orderings for all or a subset of choice options should be consistent with RUT (Luce and Suppes, 1965).

1.4 Random utility model

Thurston (1927) originally developed the concept of utility in terms of psychological stimuli, leading to a binary probit model of whether respondents can differentiate between the levels of stimulus. Marschak (1960) interpreted the stimuli as utility and provided a derivation from utility maximization. Following Marschak, models that can be derived in this way are called random utility models (RUMs). Utility is defined as “the net benefit derived from taking some action” (Train, 2003). Utility is a latent construct, which is not directly observed by the modeler and is therefore treated as a random variable (Sorci et al., 2010). The derivation assures that the model is consistent with utility maximization, but it does not preclude the model from being consistent with other forms of behavior. The models can also be seen as simply describing the relation of explanatory variables to the outcome of a choice, without reference to exactly how the choice is made (Train, 2003).

In DCM, the most common approach is based on random utility maximization theory (McFadden, 1974), which assumes that the concept of individual choice behavior is intrinsically probabilistic. According to this theory, each individual has a utility function associated with each of the alternatives considered. This individual function can be divided into a systematic part, which considers the effect of the observable explanatory variables, and a random part that takes into account all the effects not included in the systematic part of the utility function. In other words, choices are modeled using a structural equation (equation (1.1)) representing individual preferences, where the explanatory variables are the alternative
attributes and individual decision-maker’s characteristics. The observed choice corresponds to
the alternative that maximizes the individual utility function, a process represented by a
measurement equation (equation (1.2)). Because the utility function has a random nature, the
output of the model corresponds to the choice probability of individual \( n \) choosing alternative
\( i \). The set of equations describing the standard discrete choice setting is given by:

\[
U_{ni} = X_{ni} \beta + \varepsilon_{ni} \quad (1.1)
\]

\[
y_{ni} = \begin{cases} 1 & \text{if } U_{ni} \geq U_{nj} \quad \forall j \in C_n, j \neq i \\ 0 & \text{otherwise} \end{cases} \quad (1.2)
\]

where \( U_{ni} \) corresponds to the utility of alternative \( i \) as perceived by individual \( n \); \( X_{ni} \) is a
row vector of attributes of alternative \( i \) and socioeconomic characteristics of individual \( n \); \( \beta \)
a column vector of unknown parameters; \( \varepsilon_{ni} \) an error term capturing different sources of
uncertainty, including unobserved attributes, unobserved individual characteristics,
measurement errors and instrumental variables (Sorci et al., 2010); \( y_{ni} \) indicates whether
alternative \( i \) is chosen by individual \( n \) or not; and \( C_n \) the individual set of available
alternatives (Hess and Daly, 2010). The random component does not suggest that individuals
make choices in some random fashion, rather it implies that important but unobserved
influences on choices exist that can be characterized by a distribution in the sampled
population. However, we do not know where any particular individual is located in that
distribution. Hence, we assign this information to that individual stochastically (Louviere et
al., 2000).
RUMs are derived as follows following Train (2009). A decision-maker, labeled \( n \), faces a choice among \( J \) alternatives. The decision-maker would obtain a certain level of utility (or profit) from each alternative. The utility that decision maker \( n \) obtains from alternative \( j \) is \( U_{nj} \), \( j=1, \ldots, J \). This utility is a latent construct perceived by the decision-maker, but not by the researcher, and is therefore treated as a random variable. The decision-maker chooses the alternative that provides the greatest utility. The behavioral model is therefore to choose alternative \( i \) over alternative \( j \) if and only if:

\[
U_{n_i} > U_{n_j}, \forall j \neq i
\]  

(1.3)

The researcher does not observe the decision maker’s utility, only the attributes of the alternatives faced by the decision-maker, labeled \( X_{nj}, \forall j \neq i \), and some characteristics of the decision maker, labeled \( S_n \), and the researcher can therefore specify a function that relates these observed factors to the decision maker’s utility. The function is denoted \( V_{nj} = V(X_{nj}, S_n) \).

Usually, \( V_{nj} \) depends on parameters that are unknown to the researcher and are therefore estimated statistically. Since there are aspects of utility that the researcher does not or cannot observe, utility \( U_{nj} \) is decomposed as \( U_{nj} = V_{nj} + \varepsilon_{nj} \) where \( \varepsilon_{nj} \) captures the factors that affect utility but are not included in \( V_{nj} \). This decomposition is fully general, since \( \varepsilon_{nj} \) is simply defined as the difference between true utility \( U_{nj} \) and the part of utility that the researcher captures through \( V_{nj} \). Given this definition, the characteristics of \( \varepsilon_{nj} \), such as its distribution, depend critically on the researcher’s specification of \( V_{nj} \). It is defined relative to a researcher’s representation of a choice situation. This distinction becomes relevant when evaluating the
appropriateness of various specific DCMs. The joint density of the random vector \( \varepsilon = (\varepsilon_{1n}, \ldots, \varepsilon_{jn}) \) is denoted \( f(\varepsilon) \). With this, the researcher can make probabilistic statements about the decision maker’s choice. The probability that decision maker \( n \) chooses alternative \( i \) in choice situation \( t \) is:

\[
P_{ni} = \text{Prob}(U_{ni} > U_{nj})
\]

\[
= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj})
\]

\[
= \text{Prob}(\varepsilon_{ni} - \varepsilon_{nj} < V_{ni} - V_{nj}), \forall j \neq i
\]  

(1.4)

This probability is a cumulative distribution, namely the probability that each random term \( \varepsilon_{nj} - \varepsilon_{nit} \) is below the observed quantity \( V_{nit} - V_{nj} \). Using the density \( f(\varepsilon) \), this cumulative probability can be rewritten as:

\[
P_{nit} = \text{Prob}(\varepsilon_{nj} - \varepsilon_{nit} < V_{nit} - V_{nj})
\]

\[
= \int I[\text{Prob}(\varepsilon_{nj} - \varepsilon_{nit} < V_{nit} - V_{nj})]f(\varepsilon_n)\,d\varepsilon_n, \forall j \neq i
\]  

(1.5)

where \( I(\cdot) \) is the indicator function, equaling 1 when the expression in parentheses is true and 0 otherwise. This is a multidimensional integral over the density of the unobserved portion of utility \( f(\varepsilon) \). Different DCMs can be obtained from different specifications of this density, that is, from different assumptions about the distribution of the unobserved portion of utility. The integral takes a closed form only for certain specifications of \( f(\varepsilon_n) \). The logit model in equation 1.6 has a closed-form expression for this integral. It is derived under the assumption that the unobserved portion of utility is distributed iid extreme value type I. The probit model is derived under the assumption that \( f(\varepsilon_n) \) is a multivariate normal, and the mixed logit model in equation 1.6 is based on the assumption that the unobserved portion of utility consists of a
part that follows any distribution specified by the researcher plus a part that is iid extreme value type I.

A mixed logit model is any model whose choice probabilities can be expressed in the form:

\[ P_{ni} = \int L_n(\beta) f(\beta) d\beta \]  

(1.6)

where \( L_{ni} \) is the logit probability evaluated at parameters \( \beta \):

\[ L_n(\beta) = \frac{e^{V_n(\beta)}}{\sum_{j=1}^{J} e^{V_j(\beta)}} \]  

(1.7)

\( V_n(\beta) \) is the observed portion of the utility which depends on the parameters \( \beta \). If utility is linear in \( \beta \), then \( V_n(\beta) = \beta'X_{ni} \). In this case, the ML probability takes the form:

\[ P_{ni} = \int \left( \frac{e^{\beta'X_{ni}}}{\sum_{j=1}^{J} e^{\beta'X_{nj}}} \right) f(\beta) d\beta \]  

(1.8)

The ML probability is a weighted average of the logit formula evaluated at different values of \( \beta \) with the weights given by the density function \( f(\beta) \). In the statistics literature, the weighted average of several functions is called a mixed function, and the density that provides the weights is called the mixing distribution. ML is a mixture of the logit function evaluated at different \( \beta \)’s with \( f(\beta) \) as the mixing distribution (Train, 2003). With probit and ML, the resulting integral does not have a closed form and is evaluated numerically through simulation (Train, 2009).
Like the ML model, the latent class (LC) model allows the analyst to harvest a rich variety of information about behavior from a panel or repeated measures data set (Crockett et al., 2010). The underlying theory of the LC model posits that individual behavior depends on observable attributes and on latent heterogeneity that varies with factors that are unobserved by the analyst. It is assumed that individuals are implicitly sorted into a set of finite classes, but which class contains any particular individual, whether known or not to that individual, is unknown to the analyst. Following Hensher and Green (2008), for a given class of respondent $C$, the probability of the discrete choice among $J$ alternatives by individual $n$ in choice task $t$ can be specified as:

$$P_{nit|c} = \frac{\exp(\beta'_c X_{nit})}{\sum_{i=1}^{J} \exp(\beta'_c X_{nit})} \quad (1.9)$$

where $P_{nit|c}$ is the probability of individual $n$ in class $C$ choosing alternative $i$ in choice task $t$, $\beta'_c$ is a set of estimated coefficients for class $C$; $X_{nit}$ is a set of attributes that describe the choice alternative. This standard logit choice model is specified with its own set of coefficients for the $C$ classes of respondents.

The probability that a respondent belongs to a given class can be based on their observed choices, their observed characteristics and their reported attitudes. This is achieved using another logit model that shows the probability that individual $n$ belong to class $C$ as:

$$P_{nc} = \frac{\exp(\theta_c Z_n)}{\sum_{c=1}^{C} \exp(\theta_c Z_n)} \quad (1.10)$$
where \( P_{nc} \) is the probability that individual \( n \) belongs to class \( C \); \( \theta C \) is a set of \( C \) coefficient vectors, and \( Z_n \) a set of observable characteristics. The joint probability that a randomly chosen individual \( n \) chooses alternative \( j \) and is in class \( C \) is given by the multiplication of the probabilities in equations (1.9) and (1.10) as expressed in equation (1.11):

\[
P_{nj} = \sum_{c=1}^{C} P_{nitc} P_{nc}
\]  

(1.11)

Specifying utility in terms of random and deterministic influences makes the modeling ideally suited for econometric analysis of choices and provides the probability distributions for the estimates of welfare measures. However, the RUM, as a basis for DCMs and welfare measurement, departs from the standard neoclassical model in two ways (Nancy, 2007). First, it models an individual’s behavior on a choice occasion, that is, it models a single choice among a finite set of mutually exclusive alternatives. This contrasts with the neoclassical model that characterizes consumption decisions as a budget allocation process within a specific period of time, for example a year. Second, the random utility model incorporates a stochastic term reflecting the researcher’s lack of knowledge right from the start rather than adding it in an ad hoc way to the demand function after the entire constrained utility maximization process has been rationalized. Both departures from the neoclassical model give the discrete analysis an element of realism.

1.5 Outline of the PhD dissertation

This dissertation focuses on the potential role and use of DCEs in informing environmental policy and decision-making in a developing country context based on case studies from
Ethiopia. The nature of the contribution of this dissertation is both empirical and methodological (see Figure 1). Three of the four case studies in this dissertation focus on improvements in public and private services, i.e. urban solid waste management (Chapter 2), urban drinking water supply (Chapter 3) and rural irrigation water supply (Chapter 4). The fourth case study tests household demand for organic produce on a rural agricultural market (Chapter 5).

Chapter 2 is the first DCE study carried out in this PhD dissertation, focusing on public preferences for improvements in solid waste services in Hawassa, Ethiopia. Of particular interest in this study is the abolition of child labor in the waste management sector. Observed and unobserved preference heterogeneity are analyzed using mixed logit choice models, which is the basis for the subsequent estimation of the welfare implications of different solid waste service improvement policy scenarios.

Chapter 3 estimates the public benefits of improved urban drinking water supply in a DCE in Hawassa, Ethiopia. Ethiopia is among the Sub-Sahara African countries with the lowest water supply coverage. Besides estimating the welfare implications of different policy scenarios with regards to drinking water supply improvements based on the estimated choice model in this chapter, the stability of public preferences is tested when making repeated choices in the DCE.

Chapter 4 is more methodological in nature than the previous two chapters. Preference learning is addressed in the margin, and most attention goes out to the influence of visual aids in DCEs on attribute attendance of literate and illiterate survey participants. The policy context of chapter 4 is that it informs local and regional policy makers in Ethiopia about the most
important terms and conditions needed to incentivize farmers to enter into contractual agreements to invest in water harvesting technology on their land to secure food production.

Chapter 5 tests potential hypothetical bias in DCEs based on inter-respondent and intra-respondent comparisons of choice behavior, applying a hypothetical and real choice experiment in a local agricultural market in Ethiopia. This chapter too is primarily methodological in nature. Hypothetical bias is tested by comparing (i) hypothetical and real choices, (ii) preference parameters of the estimated choice models related to the hypothetical and real choices, and (iii) hypothetical and real willingness to pay (WTP). In addition, factors explaining the observed hypothetical bias are examined.

Finally, Chapter 6 concludes and discusses the limitations of this PhD dissertation and avenues for future research.
2 Public preferences for improved urban waste management

2.1 Introduction

Access to reliable solid waste services is an essential component to improve public health, a safe environment and sustainable development, particularly in low-income countries (Atlaf and Deshazo, 1996; Ahmed and Ali, 2004; Scheinberg et al., 2010). During the last decades, the sheer volume of the generated solid waste, particularly in large cities, has been increasing at an unprecedented rate because of rapid rates of urbanization and rising standards of living (NEERI, 1994; Beede and Bloom, 1995; CPCB, 2000; UN, 2000; Rathi, 2007). Lack of planning, poor or no separation of waste at the source and ill designed disposal systems, insufficient public and private funds and corruption in the public sector characterize many developing countries’ solid waste management (SWM) services (Adedibu, 1985; Diallo and Coulibaly, 1991; Gupta et al., 1998; Buenrostro et al., 2001). The negative externalities associated with increasing levels of unmanaged solid waste are furthermore exacerbated by the inadequate provision of other basic infrastructure and services such as water supply, sanitation facilities, sewage and transportation systems.

In Ethiopia, waste management is a problem that affects water quality and public health in urban and rural communities. Solid waste management practices are generally characterized by poor quality services in terms of collection, recycling, open dumping and other forms of

---

improper final disposal, scavenging at landfill sites by waste pickers, and improper and unlined landfill sites and corresponding environmental and public health issues (Baraki, 2003; Tolina, 2006; Walelegne, 2003). Currently an estimated 65 percent of all solid waste produced per day is collected and disposed by municipalities in dumpsites, 5 percent is recycled, 5 percent is composted and the remaining 25 percent of the solid waste is uncollected and dumped in unauthorized areas such as open fields, ditches, sewers, streets and other spaces in Ethiopian cities (Abera and Ali, 2015). Only 2 of the 15 regional cities in Ethiopia control and have a protocol for their solid waste disposal practices (Degnet and Maru, 2005).

Being one of the largest cities in Ethiopia, Hawassa has been expanding rapidly because of the influx of unemployed rural residents and improvements in infrastructure and booming construction works. However, solid waste management in Hawassa is far from adequately organized, and collected in traditional ways, where people providing the service in the city use their children as young as 10 years old to collect solid waste using donkey pulled carts without any safety regulations. Africa has the largest number of child labourers in the world: 59 million children between the age of 5 and 17 are involved in hazardous work (ILO, 2013). According to the CSA report of the Ethiopian Demographic and Health Survey (2011), 27 percent of the children between the age of 5 and 14 (31% of all male and 24% of all female children) is involved in paid and unpaid child labor, either in family businesses or working for non-family members. Ninety-five percent of the collected amount of solid waste in Hawassa is indiscriminately thrown away at dumping sites on the periphery of the city, and sometimes dumped illegally at empty lots scattered throughout the city. There is no provision of bins and waste collection service charges, collected directly door-to-door by the waste operators, are not
uniform, varying from 2 Birr to 15 Birr (10 to 85 US$ cents) per household per week depending on the size of the sack used to dispose the waste).

Most attempts to improve solid waste management in cities in Ethiopia have focused on the technical supply aspects of the service, mostly ignoring the demand side of the solid waste management services, including the ability to pay the service charge in order for the service to be provided in a financially viable way based on full cost recovery and without relying on child labor. A limited number of stated preference studies have been conducted in developing countries focusing on solid waste collection, applying the contingent valuation (CV) method (e.g. Altaf and Deshazo, 1996; Marchand, 1998; Rahji and Oloruntoba, 2008). In Ethiopia, similar CV studies were carried out, in Addis Abeba (Walelegne and Alebel, 2002) based on 494 household interviews yielding an average WTP per household per month between 4.6 and 16.7 Birr, in Mekelle (Gebreegziabeher et al., 2012) targeting 226 households resulting in an average WTP between 7.8 and 11.9 Birr per household per month, and Bahir Dar (Bishaw, 2011) sampling 164 households and producing an average WTP of 14.5 per household per month.

In this paper, we develop and apply a discrete choice experiment (DCE), another stated preference method, to assess public preferences in Hawassa for improved solid waste services, defining the service provision as a bundle of services as in, for example, Adetola and Benedicta (2010), including an increase in collection frequency and waste separation, and including the abolishment of child labor as a separate characteristic of the future solid urban waste

\footnote{2 Birr is Ethiopia’s national currency. At the time of the study, 1 Birr was equal to approximately US$ 0.057.
management sector. The main objective of the study is to assess public preferences for improved urban solid waste services, estimate public WTP, and identify what determines public WTP to inform future policy and decision-making related to the solid waste management sector. The study presented here differs from existing studies applying DCEs related to solid urban waste (e.g. Sakata, 2007; Karousakis and Birol, 2008) in that it includes the labor reform of the waste collection sector in a developing country context as an important additional choice characteristic.

The next sections describe the design of the DCE (Section 2), the underlying econometric choice model (Section 3), and the main results (Section 4). Section 5 concludes the paper.

2.2 The choice experiment

The choice attributes and levels were fixed based on expert interviews and a focus group discussion organized in Hawassa City Hall. The focus group consisted of 10 residential households, one expert in solid waste management, 5 door-to-door waste collectors, and 4 landfill operators. The group discussion aimed to identify the main issues related to solid waste management in Hawassa and translate these issues into concrete choice attributes. The meeting also served to clarify issues related to who provides the services, what methods are used to collect the waste, waste transport and disposal, the number of landfills around Hawassa, how payments are fixed and by whom. Domestic waste is collected, transported and disposed off by private households organized in micro-enterprises. Although education is compulsory in Ethiopia for children between 7 and 16 years, children as young as 10 years engage in waste collection, transportation and disposal activities in the city. Child labor was identified by the group participants as one of the main issues in the waste management sector. Other problems
related to irregular waste collection, the lack of waste separation (often carried out afterwards on landfill sites by children), the lack of fixed collection rates and the need to renegotiate rates (currently varying between 1 and 20 Birr per month) regularly on an individual household basis.

In the DCE, respondents were therefore presented with a series of possible solid waste service improvement scenarios, differing in (i) the level of weekly waste collection frequency, (ii) whether or not the waste is separated at the source and waste bins are provided so households can sort their waste into recyclables and non-recyclables, and (iii) whether the sector only employs adults instead of children in providing the solid waste services. These improvements come at a price, which consists in this case of an additional monthly service charge just like existing water and electricity bills, paid to municipality approved door-to-door service providers. Current waste collection frequency of approximately once per week was extended with 1 and 2 days extra per week, while the two other non-monetary choice attributes (waste separation into recyclable and non-recyclable waste and the abolishment of child labor) were defined as simple binary variables (i.e. ‘yes’ or ‘no’). Five price levels were identified for the increase in the household’s monthly waste service charge of 1, 5, 10, 15, and 20 Birr. This design appeared to work well in a follow-up pre-test based on 20 individual face-to-face interviews with a random selection of households in Hawassa.

Alternative policy scenarios were created by combining the four choice variables based on their different attribute levels in SPSS using a main effects fractional factorial orthogonal resolution III design, which assumes that interaction effects of the attributes are negligible (Kuhfeld, 2005). Main effects typically account for 70 to 90 per cent of the explained variance in these
kind of experimental designs (Dawes and Corrigan, 1974). Twenty-four alternatives were generated and converted into 12 pairwise choice tasks, which were randomly assigned to each respondent (e.g. Hanley et al., 2014). Hence, each respondent answered 12 choice cards. Although there is little consensus, and even less empirical data regarding the optimum level of choice tasks given the choice task complexity, participants in a DCE typically evaluate between one and sixteen choice sets, with the average being somewhere around eight choice sets per respondent (Louviere et al., 2003).

Each choice card shows two hypothetical choice alternatives describing a future improvement of current solid waste services along with the option to choose none of the two. Inclusion of this latter ‘status quo’ alternative is instrumental to be able to estimate welfare measures that are consistent with demand theory (Bateman et al., 2003) and not to force respondents to choose the proposed alternatives. Respondents were told that they would not have to pay anything extra if they choose the opt-out. In order to account for the fact that some of the survey participants are illiterate, interviewers read the information on each choice card out loud during the face-to-face interviews. An example of a choice card is presented in Table 2.1.

Table 2.1: An example of a choice card

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Option 2</th>
<th>Current situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection frequency per week</td>
<td>2 times</td>
<td>1 time</td>
<td>1 time</td>
</tr>
<tr>
<td>Waste separation using bins</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Service providers</td>
<td>Children</td>
<td>Adults only</td>
<td>Children</td>
</tr>
<tr>
<td>Increase in monthly service charge (Birr)</td>
<td>10</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>I prefer</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>
The DCE was implemented in February 2012 through 201 in-person interviews in three distinct city zones in Hawassa (Misrak Wukro, Manaharia and Tabor). The city zones were selected with the help of the Hawassa Urban Planning, Sanitation and Beautification Agency. No statistical information is available about each zone’s residential household characteristics, but the zones differ considerably in terms of their socio-economic status. Misrak Wukro is the poorest part of the city, while households living in Manaharia and Tabor are better off.

In each zone, first households paying for door-to-door solid waste collection were identified. From this, a random selection of households was chosen in each zone. In most cases, women were interviewed since they were at home during the day, and they are traditionally more responsible for handling solid waste in the household (Gebregziabeher et al., 2012). The response rate was 100 percent, which is not unusual for this kind of stated preference research in developing countries (Whittington, 1998). Trained local enumerators were used for the interviews. After data screening, all the 201 interviews could be used in the analysis, 68 from Misrak Wukro, 68 from Manaharia and 65 from Tabor.

2.3 The choice model

Public preferences for improved solid waste services are modeled in terms of McFadden’s (1974) Random Utility Model (RUM), allowing for a separation of utility \( U_{ijt} \) into a deterministic part \( V_{ijt} \) and a stochastic part \( \varepsilon_{ijt} \) as in (2.1):

\[
U_{ijt} = V_{ijt} + \varepsilon_{ijt}
\] (2.1)
The deterministic part of utility is typically specified to be linear in parameters including a constant term. DCEs fall in the class of attribute-based methods in which the deterministic part of utility for individual $i$ for good $j$ in choice task $t$ is described in (2.2) as a linear function of its attributes $X_{ijt}$, other explanatory variables $Z_{ijt}$ and a constant that is specific to an alternative $k_j, \forall j = 1$ alternatives (Train, 2003):

$$V_{ijt} = k_{j-1} + \beta X_{ijt} + \alpha Z_{ijt}$$  \hspace{1cm} (2.2)

The alternative-specific constant (ASC) captures the average effect on utility of all factors that are not included in the model. In this particular case, the ASC can be considered as a specific parameter of the utility function capturing a bias toward that alternative (Train, 2003). In each choice task the respondent is presented with a limited set of proposals $D_n$ to improve current solid waste services. The stochastic term is assumed to follow an IID extreme value distribution of type I.

In order to account for possible preference heterogeneity, the preference parameters are allowed to vary across respondents, applying different mixing distributions. Equation (2.3) describes the mixed logit (ML) probability of individual $i$ selecting alternative $j$ in choice task $t$ over other choice alternatives $k$. The utility coefficients $\beta$ vary across individuals, hence $\beta_i$, with density $\Delta(\beta_i \mid b)$. This density can be a function of any set of parameters and represents in this case the mean and covariance of $\beta$ in the sample population.
ML-models assume heterogeneity to be continuous over the interval spanned by the assumed distribution for the preference parameters (Scarpa et al., 2005). Treating preference parameters as random variables requires estimation through simulated maximum likelihood. Procedurally, the maximum likelihood algorithm searches for a solution by simulating draws from distributions with given means and standard deviations. Probabilities are calculated by integrating the joint simulated distribution. Applications of ML-models have shown that this model is superior to the standard multinomial logit model in terms of overall fit and accuracy of welfare estimates (e.g. Morey and Rossmann, 2003; Provencher and Bishop, 2004; Brouwer et al., 2010a).

Mixed logit models account for preference heterogeneity and repeated choices (Train, 2003). Even if unobserved heterogeneity is accounted for in a ML-model, the model may still fail to explain the sources of this heterogeneity (Hynes et al., 2008). To this end, interactions of respondent specific household characteristics with choice specific attributes can be included in the utility function to improve the model fit (Revelt and Train, 1998). Finally, the price attribute included in the DCE allows for the calculation of welfare estimates (e.g. Hensher et al., 2005). The welfare measures calculated in this study represent the monetary values arising from changes (improvements) in the bundle of solid waste services and are calculated as follows (e.g. Bateman et al., 2003):

\[
P_{ij} = \left( \sum_{j \in D} \exp\left( \frac{\beta_i X_{ij} + \alpha Z_{ij}}{\beta_i} \right) \right) \Lambda(\beta_i | b) d\beta_i \quad \forall j \in D_{ij}
\]  

(2.3)
where the utility associated with the proposed change in the waste service \( (V') \) is subtracted from the utility associated with the status quo level of service provision \( (V_0) \). If the model is linear in the monetary choice attribute, this indirect utility difference is then divided by the negative of the coefficient associated with the service charge \( \beta_{\text{service charge}} \).

### 2.4 Results

#### 2.4.1 Sample characteristics

Table 2.2 reports the main socio-demographic characteristics of the interviewed households in the three city zones. The three samples are homogeneous and hence comparable in terms of gender (81% of all respondents is female), age (respondents are 18 years or older, the oldest respondent is 75 years) and household size (on average each household consists of 4.9 persons), and also no significant differences exist between the sub-samples with regards to the average amount of waste they generate (on average 1.2 kg per household per day) and the average price they pay for their waste collection (prices varied between 1 and 25 Birr per household per week, with half of the sample paying 3 Birr per week and the sample average is 3.7 Birr per week).
Table 2.2: Main socio-demographic and solid waste service characteristics of the sample population across the 3 districts in Hawassa

<table>
<thead>
<tr>
<th>Variable</th>
<th>Misrak Wukro</th>
<th>Manaharia</th>
<th>Tabor</th>
<th>Kruskal-Wallis test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (% female respondents)</td>
<td>0.78 (0.05)</td>
<td>0.81 (0.05)</td>
<td>0.85 (0.05)</td>
<td>0.964 (p=0.618)</td>
</tr>
<tr>
<td>Average age (years)</td>
<td>28.1 (1.3)</td>
<td>30.3 (1.6)</td>
<td>27.7 (1.3)</td>
<td>0.554 (p=0.758)</td>
</tr>
<tr>
<td>Average household size (persons)</td>
<td>4.9 (0.3)</td>
<td>4.6 (0.3)</td>
<td>5.3 (0.2)</td>
<td>3.692 (p=0.158)</td>
</tr>
<tr>
<td>Share illiterate (%)</td>
<td>11.8</td>
<td>8.8</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Share who completed primary school only (%)</td>
<td>36.8</td>
<td>30.9</td>
<td>41.5</td>
<td>2.615 (p=0.271)</td>
</tr>
<tr>
<td>Share who completed secondary school (%)</td>
<td>32.4</td>
<td>27.9</td>
<td>30.8</td>
<td></td>
</tr>
<tr>
<td>Average household income (Birr/month)</td>
<td>846.4 (79.0)</td>
<td>1,489.7 (153.0)</td>
<td>4,165.4 (551.7)</td>
<td>81.854 (p=0.001)</td>
</tr>
<tr>
<td>Average amount of waste per day (kg)</td>
<td>1.21 (0.11)</td>
<td>1.12 (0.09)</td>
<td>1.23 (0.09)</td>
<td>1.519 (p=0.468)</td>
</tr>
<tr>
<td>Average service price (Birr/household/week)</td>
<td>4.2 (0.4)</td>
<td>3.4 (0.1)</td>
<td>3.4 (0.2)</td>
<td>3.746 (p=0.154)</td>
</tr>
<tr>
<td>Number of respondents</td>
<td>68</td>
<td>68</td>
<td>65</td>
<td></td>
</tr>
</tbody>
</table>

Note: ¹ Test of equality of distributions over all education levels across the three samples.
The main difference between the three sub-samples is, as expected, found in the income levels. Average household income is lowest in Misrak Wukro (US$ 48 per household per month) and highest in Tabor (US$ 237 per household per month). Divided by 30 days and the average household size, average income per person per day is less than the World Bank poverty line of US$1.25 in Misrak Wukro and Manaharia. Although the share of illiterate respondents is highest in Misrak Wukro (12%) and lowest in Tabor (1%), the distribution of respondents over the different education levels is statistically not significantly different across the three samples. Differences in the share of respondents who completed primary and especially secondary school are less pronounced.

2.4.2 Estimated choice models

A large majority of the respondents (98%) is interested in and willing to pay extra for improved solid waste services and always chose at least one of the hypothetical alternatives over the status quo. Across all 2412 choice occasions (201 respondents answering 12 choice sets), the opt-out was chosen in 2.3 percent of the cases. No protest voters were encountered, i.e. respondents who consistently chose the opt-out alternative in all choice occasions for reasons unrelated to their preferences or income constraints (e.g. Brouwer and Martin-Ortega, 2012). Choice behaviour was modelled accounting for the panel structure of the collected choice data. For efficiency reasons, the models were estimated using a Halton sequence of 500 replications in a quasi-Monte Carlo maximum likelihood simulation (Bhat, 2001) in NLOGIT version 5.0.

The ML choice model results without and with the inclusion of covariates are presented in Table 2.3. The estimated choice models are highly significant as can be seen from the outcome
of the $\chi^2$ test statistic testing differences in model fit when going from a model which includes the constant only and the fitted models presented in Table 2.3, and the relatively high pseudo
Table 2.3: Estimated mixed logit choice models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Attributes only</th>
<th>Including covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient estimate</td>
<td>St. error</td>
</tr>
<tr>
<td><strong>Mean coefficients choice attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>6.818*** 0.842</td>
<td></td>
</tr>
<tr>
<td>Increased collection frequency (times/week)</td>
<td>0.099** 0.046</td>
<td></td>
</tr>
<tr>
<td>Waste separation (1=yes)</td>
<td>0.239*** 0.088</td>
<td></td>
</tr>
<tr>
<td>Service provider (1=adults only)</td>
<td>0.655*** 0.085</td>
<td></td>
</tr>
<tr>
<td>Increase in service charge (Birr/month)</td>
<td>-0.068*** 0.007</td>
<td></td>
</tr>
<tr>
<td><strong>Standard deviation random parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>3.547*** 0.475</td>
<td></td>
</tr>
<tr>
<td>Increased collection frequency (times/week)</td>
<td>0.277*** 0.072</td>
<td></td>
</tr>
<tr>
<td>Waste separation (1=yes)</td>
<td>1.431*** 0.148</td>
<td></td>
</tr>
<tr>
<td>Service provider (1=adults only)</td>
<td>0.865*** 0.231</td>
<td></td>
</tr>
<tr>
<td>Increase in service charge (Birr/month)</td>
<td>0.061*** 0.008</td>
<td></td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (1=male respondent) x collection frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x service provider</td>
<td>-0.280** 0.111</td>
<td></td>
</tr>
<tr>
<td>Household size (number of persons) x waste separation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income (Birr/month) x service provider</td>
<td>0.473·10^-4 *</td>
<td></td>
</tr>
<tr>
<td>Resident in Manaharia (1=yes) x service charge</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model summary statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1706.764</td>
<td></td>
</tr>
<tr>
<td>$\chi^2 (p&lt;)$</td>
<td>1886.177 (0.001)</td>
<td></td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>Number of respondents</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>2412</td>
<td></td>
</tr>
</tbody>
</table>

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$. 
R² (Hoyos, 2010). All random coefficients are also statistically significant, indicating that the choice attributes are characterized by significant preference heterogeneity. The distribution of the random parameters is normal for the continuous choice attributes and uniform for the dummy variables (e.g. Hensher et al., 2005). The coefficient estimates for the choice attributes are all statistically significant and have the expected signs. The significant positive outcome of the ASC implies that respondents prefer some change instead of no change from the current situation. Respondents positively value an increase in collection frequency, waste separation and the abolishment of child labor, while a price increase for the waste services affects their utility negatively.

Applying a systematic search for the impact of respondent characteristics on both choice attributes and the ASC, a limited set of socio-demographic covariates is found to interact significantly with the choice attributes, but not with the ASC.

Significant gender effects are found for two choice attributes: women attach more value than men to the proposed increase in collection frequency and the abolishment of child labor. Bigger households are furthermore less interested in the separation of recyclable solid waste. One possible reason might be that this would entail more work for bigger households as they produce more waste. However, no significant relationship could be detected between household size and the amount of generated waste based either on correlation or the cross-tabulation χ² test statistic. Another possible explanation can be found in the fact that a significant association exists between household size and education level (the χ² test statistic based on cross-tabulation is 119.105, p<0.001). Respondents representing bigger households, that is, consisting of more than the average of 5 persons, are relatively speaking compared to
smaller households consisting of less than 5 persons more often illiterate, while respondents from smaller households have relatively more often completed primary and secondary school. Education was expected to play a direct role in explaining choice behavior for the waste service improvement scenarios, in particular for recyclable waste separation and the abolishment of child labor, but none of the education interaction terms turned out to be statistically significant. For example, Adetola and Benedicta (2010) found that education generally, i.e. interacting with the ASC but not in interaction with choice attributes such as collection frequency or waste separation, has a positive impact on choice behavior in favor of the improvement of solid waste services in Nigeria. In this study, we find a significant income effect on the abolishment of child labor: richer households attach a small, but statistically significant higher value to the abolishment of child labor in the waste management sector than poorer households.

Finally, a significant effect is found for the zone in which respondents live. Respondents living in the richer zone Manaharia are willing to pay a significantly higher increase in service charge than respondents living in the poorer zone Misrak Wukro (i.e. the baseline zone in the estimated choice model with covariates). No such significant effect is found for Tabor.

2.4.3 Preference stability across subsets of choice tasks

Preference stability was tested across sub-sets of choice tasks. In choice experiments, the utility function of each individual is assumed to be stable throughout the choice sequence (Brouwer et al., 2010b). Choice behaviour at the start of the experiment is expected to be consistent with choice behaviour at the end of the experiment. Specifically, we test whether the utility parameters $\beta$ and the confounded scale parameter $\lambda$ estimated in the choice model remain the same across all 12 choice tasks. To this end, a pairwise comparison is made between the 12
choice cards. Due to the limited number of observations, the 12 choice tasks are split into 3 groups of 4 tasks. In a first step, the estimated ML-model provides efficient estimates for $\lambda_1\beta_1$, $\lambda_2\beta_2$ and $\lambda_3\beta_3$ and a likelihood function for each set of four choice tasks. Because the estimated parameters are confounded with the scale parameters, we have to isolate possible differences in variance before comparing the preference parameters (Louviere et al, 2000). This means that the scale parameter of the first set of four choice tasks is normalized to $\lambda_1 = 1$. A pooled model is then estimated in a second step by including the first and second set of choice tasks, which has the effect of imposing equality on the preference parameters ($\beta_1 = \beta_2$). In this model, the relative scale parameter $\lambda_2 / \lambda_1$ is, however, not set equal to 1. A search procedure over a range of relative scale parameters is applied to estimate the combination of scale and pooled preference parameters that provides the best model fit. At each possible relative scale parameter, the data for the second set of choice tasks are re-scaled such that an ML-model can be estimated. After the best-fit model has been identified, a standard chi-square distributed Likelihood Ratio (LR) test using the Log Likelihoods (LLs) of the models from step 1 and the best-fit model of step 2 is used to test the difference between the preference parameters in the two choice sets under the null hypothesis that they are the same.

The third step tests for differences in scale across choice tasks. In principle, this step is conditional on the outcome of the mentioned LR test. It requires the estimation of an ML-model for the same pooled model as in step 2, but with equality imposed on both preference and scale parameters this time ($\beta_1 = \beta_2$ and $\lambda_1 = \lambda_2$). Again, a chi-square distributed LR test can be applied to compare the LL of the estimated model to the LL of the pooled model with varying scale parameters.
The results of this test procedure are presented in Table 2.4. Due to the confounding of preference and scale parameters, it is impossible to attribute the detected differences between the estimated groups of models either to differences in preference parameters, scale parameters or both (Louviere et al., 2000).

2.4.4 Economic welfare measures for solid waste service improvements
Based on the coefficient estimates in the first attributes only model, WTP values and their standard errors can be calculated, using the Krinsky-Robb (1986) method. Marginal WTP for one extra day per week of solid waste collection in the whole sample is 2.9 Birr per household per month (Table 2.5), while respondents are willing to pay 4.3 Birr extra when their waste is collected 2 days extra per week. This is equivalent to an increase in the sample’s average waste collection charge (see Table 2.5) of 20 and 30 percent, respectively. Marginal WTP is 3.5 Birr (with a standard error of 1.3) per household per month for separating and recycling solid waste and 9.6 Birr (with a standard error of 1.5) for the abolishment of child labor. Adding waste separation and the abolishment of child labor to the increase in collection frequency (the maximum benefit scenario in Table 2.4) results in an additional increase of public WTP for the whole sample of a factor 4 to 5 (from 2.9 to 16.0 Birr and from 4.3 to 17.4 Birr for an increase in solid waste collection frequency of 1 and 2 days, respectively). Public WTP is significantly higher in the richer city zones Manaharia and Tabor than in Misrak Wukro, but no significant difference exists between the former two (Manaharia and Tabor). In Misrak Wukro, public WTP for waste separation and the abolishment of child labor in the waste management sector is the same for one or two extra days of waste collection. This corresponds to a 56 percent increase over and above the current average service charge households pay. In Manaharia this
increase over and above the current service charge is between 174 and 187 percent and in Tabor between 152 and 174 percent.
Table 2.4 Test results of equality of preference ($\beta$) and scale ($\lambda$) parameters between sets of choice tasks

<table>
<thead>
<tr>
<th>Comparison of choice tasks</th>
<th>LogLik$_i$</th>
<th>LogLik$_j$</th>
<th>LogLik$_{i+j}$ ($\lambda_i \neq \lambda_j$)$^a$</th>
<th>LR-test$^b$ (9 d.o.f.)</th>
<th>Reject</th>
<th>Relative scale ($\lambda_j / \lambda_i$)</th>
<th>Relative variance ($\sigma^2_j / \sigma^2_i$)</th>
<th>LogLik$_{i+j}$ ($\lambda_i = \lambda_j$)$^d$</th>
<th>LR-test$^e$ (1 d.o.f.)</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4 and 5-8</td>
<td>-581.796</td>
<td>-591.974</td>
<td>-1197.258</td>
<td>46.977</td>
<td>Yes</td>
<td>0.51</td>
<td>6.324</td>
<td>-1201.747</td>
<td>8.977</td>
<td>Yes</td>
</tr>
<tr>
<td>1-4 and 9-12</td>
<td>-581.796</td>
<td>-580.867</td>
<td>-1187.660</td>
<td>49.993</td>
<td>Yes</td>
<td>0.41</td>
<td>9.785</td>
<td>-1195.805</td>
<td>16.291</td>
<td>Yes</td>
</tr>
<tr>
<td>5-8 and 9-12</td>
<td>-591.974</td>
<td>-580.867</td>
<td>-1206.716</td>
<td>67.749</td>
<td>Yes</td>
<td>0.62</td>
<td>4.279</td>
<td>-1208.413</td>
<td>3.394</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes:

$^a$ Pooled mixed logit model with scale normalization.
$^b$ LR=-2(LL$_{pooled}$ – (LL$_i$+LL$_j$)) with $\beta$-1 degrees of freedom (d.o.f.).
$^c$ Likelihood Ratio (LR) tests performed at 1% significance level.
$^d$ Pooled mixed logit model without scale normalization.
$^e$ LR=-2(LL$_{pooled}$-without - LL$_{pooled}$-with) with 1 degree of freedom (d.o.f.).

Table 2.5: Estimated mean WTP for improvements in solid waste services (Birr/household/month)

<table>
<thead>
<tr>
<th>Change in collection frequency</th>
<th>Baseline conditions:</th>
<th>Service improvement:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No waste separation, child labor</td>
<td>Waste separation, no child labor</td>
</tr>
<tr>
<td></td>
<td>Whole sample</td>
<td>Misrak Wukro</td>
</tr>
<tr>
<td>1 extra day per week</td>
<td>2.9</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td>(1.4)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>2 extra days per week</td>
<td>4.3</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(3.1)</td>
</tr>
</tbody>
</table>

Note: standard errors between brackets.

58
2.5 Discussion and conclusions

This study contributes to the limited stock of stated preference valuation studies focusing on the urban waste sector in a developing economy. The study provides valuable insights in household preferences for improvements in current levels of service provision, in particular public WTP over and above current service charges for more frequent collection times and recyclable waste separation through the provision of waste bins. The novelty of the study is found in the fact that we explicitly address the employment of children in waste collection and disposal. Although other studies exist in a developing country context applying DCEs to elicit public preferences for more reliable waste collection services and waste separation, none of these studies addresses the phenomenon of vast employment of child labor in the waste sector.

In this study, almost all of the sampled households are willing to pay extra to improve the existing urban solid waste management system. The WTP values found in this study are remarkably close to the WTP values found in previous CV studies carried out in Ethiopia aiming to improve solid waste collection in cities such as Addis Ababa and Mekelle. Public WTP in the whole Hawassa sample for 1 or 2 extra days of waste collection per week is on the lower end of the range of WTP values found in the existing literature (4.6-16.7 Birr/household/month), while public WTP in the whole sample for the maximum improvement scenario is on the higher end of these existing CV-based WTP values.

Importantly, households are willing to pay a significant premium over and above existing service charges to abolish child labor in the waste sector, especially households with female household heads and higher income levels. Although official statistics are not available in Ethiopia, many waste collectors use their children to go door-to-door to collect waste, exposing
them to all kinds of health risks and depriving them from education. Moreover, there is public interest in the separation of recyclable and non-recyclable waste when households are provided with waste bins, except from larger households, the heads of whom in this study appeared to be generally lower educated.

These results provide important information to municipality officials and the waste management sector as these attributes can be targeted to generate additional revenues and design appropriate strategies to improve existing solid waste management services. Households in Hawassa are willing to pay 20-30 percent extra over and above their current service charge if the current collection frequency would be increased from once per week to twice or three times per week. Significant differences exist, however, between city zones where households earn distinct levels of income. Depending on the city zone, public WTP increases by 50 to 190 percent over and above the current waste collection service charge if bins are provided for recyclable waste separation and child labor is abolished. These values can be used in future policy appraisals of investment decisions to improve solid waste services in Hawassa and other cities in Ethiopia facing similar waste management challenges. Besides affordability, education was expected to play an important role in explaining public preferences especially for the choice attributes waste separation and child labor abolishment. Education only seemed to play an indirect role through household size and only affected the choice attribute waste separation.

Finally, although the design of the DCE was simple and straightforward, consisting of 4 choice attributes only, of which half were binary variables (‘yes-no’, ‘children-adults’), the number of choice tasks was considerable. Respondents were asked to evaluate and make a decision in 12
choice tasks. This did not seem to pose any problems in this study, such as fatigue, not during the pre-test or the main survey. In order to assess how stable the estimated utility function is and to what extent preference learning or fatigue played a role, choice behaviour at the start of the experiment was compared with choice behaviour at the end of the experiment (Brouwer et al., 2010b). This was investigated by comparing preference and scale parameters over the choice sequence following the Swait and Louviere (1993) test procedure. The scale parameter is inversely related with the variance of the error term (Louviere et al., 2000). If the scale parameter increases, choice variance decreases, and one could conclude that people make more accurate decisions as they make repeated choices. Obtaining estimates for the scale parameters hence provides insight into possible preference refinement during a choice sequence (Holmes and Boyles, 2005). Examining the relative variance in Table 2.4 suggests that the choice model variance decreases by 32 percent when comparing the first and second group of choice tasks (first row) with the second and third group of choice tasks (third row). This would point to some degree of preference learning and refinement during the choice sequence, providing further confidence in the robustness of the empirical results.
3 Estimation of the public benefits of urban water supply improvements\textsuperscript{3}

3.1 Introduction

Safe domestic water supply is an essential component of primary health care and plays a vital role in poverty alleviation. Inadequate water supply and sanitation services impact upon the lives of billions of poor people in the developing world (World Bank, 2004). Two in every ten persons lack access to safe water supply, five have inadequate sanitation, and nine do not have their wastewater treated. Yet, these estimates are believed to underestimate the extent of the drinking water supply problem. In many countries where water supply systems have been installed, the quality of the services provided is poor. Many consumers who are connected face unreliable water supply and when available, it is often not safe to drink (World Bank, 2004).

Water supply services in Ethiopia are among the lowest in Africa, with an average consumption of only 15 liter per capita per day in urban areas, which is far below the World Health Organization (WHO) standard of 45 liters per person per day. Safe domestic water supply is estimated to be available to 36 percent of the population in rural areas and 80 percent in urban areas (EWSRFA, 2004). According to the WHO, Ethiopia had the lowest level of water supply coverage in Sub-Saharan Africa in 2000 (39% compared to an average of 56% in Sub-Saharan Africa) and the second-lowest level of sanitation coverage (EPER, 2004). Unprotected water

\textsuperscript{3} This chapter has been published as Tarfasa, S. and Brouwer, R. (2013). Estimation of the public benefits of urban water supply improvements in Ethiopia: a choice experiment. Applied Economics, 45(9): 1099-1108.
supply sources are one of the most important problems related to water supply quality. Consequently, a majority of the Ethiopians use unsafe and polluted water and are, as a result, exposed to a large variety of water-borne diseases. This is especially the case for the rapidly growing urban population.

Besides limited protection of water supply sources, financial constraints also play an important role in the current state of water supply in Ethiopia. Investment and operation costs of domestic water supply facilities are only partly covered by the consumers (15%). The central government contributes more or less the same as the consumers, but the majority of the costs (70%) are funded by financial sources from outside Ethiopia, primarily through international aid (Teshome, 2007). There exists a huge gap between the finance required to maintain and operate the existing water supply system and the revenues generated through the existing water tariff system and other sources to match the growing demand for more reliable and safe water supply. Hence, additional funding is needed.

The main objective of this study is to assess household willingness to pay extra for improved water supply services in an urban centre in Ethiopia. Hawassa is the 8th largest city of Ethiopia with a population of almost 160 thousand people (CSA, 2007) and located almost 300 km south of the capital Addis Ababa. Little to no empirical evidence is available in Ethiopia on how much urban households would be willing to pay extra to improve water supply services for use in water supply investment decision-making or tariff setting. The only published stated preference study by Kinfe and Berhanu (2007) asked 240 randomly selected households in Addis Ababa in a contingent valuation survey for their willingness to pay (WTP) for a bucket of extra water. They found positive WTP values ranging between 15 and 20 US dollar (USD)
cents per bucket (i.e. approximately 1.5-2.0 USD cents per liter) depending on the estimated statistical model. Probably best known in the area of contingent valuation of drinking water supply in developing countries is the work by Whittington (e.g. Whittington et al., 1990; Whittington, 1998). However, most of these studies were conducted in rural areas. Very few applied contingent valuation studies exist for urban areas in the developing world, exceptions being Soto Montes de Oca and Bateman (2006) and Vásquez et al. (2009) in Mexico.

In the study presented here, we apply a more advanced stated preference choice experiment (e.g. Blamey et al., 1999; Scarpa et al., 2007), where households are asked to choose between different policy scenarios of improved water supply services at different water price levels. In the design of the choice experiment, a distinction is made between improved supply reliability and water quality. The limited number of choice experiments conducted in this area in the developed world focused on WTP to avoid water restrictions, for instance due to droughts (Hensher et al., 2006). In the analysis of the results, special attention is paid to the stability of the respondents’ choice behavior as they go through the sequence of choice tasks and their socio-demographic characteristics.

The paper is organized as follows. Section 2 presents the theoretical choice model and Section 3 the design of the choice experiment and its implementation through the household survey. Section 4 presents the results while Section 5 concludes and discusses the study’s main policy implications.
3.2 The choice model

Preferences are modeled in terms of McFadden’s (1974) Random Utility Model (RUM), allowing for a separation of utility \( U_{ijt}^e \) into a deterministic part \( V_{ijt}^e \) and a stochastic part \( \varepsilon_{ijt}^e \). Choice experiments fall in the class of attribute-based methods in which the deterministic part of utility for individual \( i \) for good \( j \) in choice task \( t \) is described in (1) as a linear function of its attributes \( X_{ijt} \) and other explanatory variables \( Z_{ijt} \) (Train, 2003):

\[
U_{ijt} = V_{ijt} + \varepsilon_{ijt} = \beta X_{ijt} + \alpha Z_{ijt} + \varepsilon_{ijt} \quad \forall j \in D_{it} \quad (3.1)
\]

In each choice task the respondent is presented with a limited set of policy proposals \( D_{it} \), each proposing an improvement in domestic water supply and water quality. The stochastic term is assumed to follow an IID extreme value distribution of type I.

To account for preference heterogeneity, the preference parameters for the non-price attributes are allowed to vary across respondents, applying different mixing distributions. Equation (3.2) describes the mixed logit (ML) probability of individual \( i \) selecting alternative \( j \) in choice task \( t \) over other choice alternatives \( k \). The utility coefficients \( \beta \) vary according to individual (hence \( \beta_i \)) with density \( \Delta(\beta_i | b) \) for the non-price attributes. This density can be a function of any set of parameters and represents in this case the mean and covariance of \( \beta \) in the sample population.

\[
P_{ijt} = \frac{\exp \left( \sum_{j \in D} \exp \left( \beta_j X_{ijt} + \alpha Z_{ijt} \right) \right)}{\sum_{j \in D} \exp \left( \beta_j X_{ijt} + \alpha Z_{ijt} \right)} \Delta(\beta_i | b) d\beta_i \quad \forall j \in D_{it} \quad (3.2)
\]
ML-models assume heterogeneity to be continuous over the interval spanned by the assumed distribution for the preference parameters (Scarpa et al., 2005). Treating preference parameters as random variables requires estimation through simulated maximum likelihood. Procedurally, the maximum likelihood algorithm searches for a solution by simulating draws from distributions with given means and standard deviations. Probabilities are calculated by integrating the joint simulated distribution. Recent applications of ML-models have shown that this model is superior to the standard multinomial logit model in terms of overall fit and accuracy of welfare estimates (e.g., Breffle and Morey, 2000; Layton and Brown, 2000; Morey and Rossmann, 2003; Provencher and Bishop, 2004; Brouwer et al., 2010a).

Mixed logit models account for respondent differences (preference heterogeneity) and repeated choices (Train, 2003). Even if unobserved heterogeneity is accounted for in a ML-model, the model may fail to explain the sources of heterogeneity (Hynes et al., 2008). To this end, interactions of respondent specific household characteristics can be included with choice specific attributes in the utility function to improve the model fit (Revelt and Train, 1998). We test to what extent data from repeated individual choices can be combined into an aggregate choice model using the Swait and Louviere (1993) test procedure. As part of this procedure, equality of scale parameters is tested. Scale parameters, and as a result choice variance, may differ across repeated choice sequences, for instance due to preference learning (Brouwer et al., 2010b).

Error component (EC) models furthermore accommodate correlation between the utilities of alternatives (Brownstone and Train, 1999). Correlation between alternatives is accounted for by including an error component with zero mean in the utility function specification to allow
for heteroscedasticity between those alternatives that are likely to be correlated. Scarpa et al. (2005) recommend applying EC models when comparing less familiar (hypothetical) alternatives with better known (existing) ones (the opt-out in this case).

Finally, if a price attribute is included in the choice experiment, welfare estimates can be derived (e.g., Hensher et al., 2005). The welfare measure represents the monetary value arising from a change in the bundle of water supply services, also referred to as the compensating surplus (CS). In the study presented here, the economic welfare implications are estimated of different water supply improvement policy scenarios. We now turn to the design of the choice experiment.

3.3 The experiment

The choice experiment was designed in collaboration with the Hawassa Water Supply and Sewerage Office (HWSSO), who is responsible for maintaining and operating the water system in the city and collecting water fees from connected domestic households. Most consumers have a private connection (73% of all water consumption), followed by public fountains (20% of the total water consumption) and yard connections where one or more taps are installed in a household’s backyard with a water meter (7% of the total consumption). In this study only those customers were interviewed who have a private connection.

The HWSSO employs a progressive monthly water tariff system that is based on the amount of water consumed by the customer. For customers consuming less than 5 m³ per month the
tariff is 1 Birr, for those consuming between 5 and 10 m³ the tariff is 3.4 Birr. Consumption levels between 10 and 20 m³ have a tariff of 4.4 Birr and for amounts greater than 20 m³ the tariff is 5.5 Birr.

In the choice experiment, respondents were presented a series of possible water supply improvement scenarios, differing in the level of water supply reliability and water quality, at different increases in their water fee. Respondents were asked to choose their most preferred policy alternative. Based on expert interviews and focus group discussions, two relevant attributes for the water supply services were selected together with their levels. Domestic water supply was extended with 1, 2 or 3 days per week and water quality was qualified as either needing boiling for infants only or no boiling at all. In addition, five price levels were identified: an increase in the household’s monthly water bill of 3, 5, 10, 15 or 20 Birr.

Alternative policy scenarios are created by combining these three variables based on their different attribute levels. Because respondents cannot be shown all different choice options, the number of possible combinations was reduced to 24 choice sets of 12 choice tasks each based on an orthogonal fractional factorial design generated in the statistical software SPSS, enabling the estimation of main effects and two-way interactions. Each respondent was randomly shown one of these 24 choice sets of 12 choice cards. Each choice card shows two hypothetical choice alternatives describing a future policy scenario along with the option to choose none of the two. Inclusion of this latter ‘status quo’ alternative is instrumental to be

---

1Birr is Ethiopia’s national currency. At the time of the study, 1 Birr was equal to approximately 0.06 USD.
able to estimate welfare measures that are consistent with demand theory (Bateman et al., 2003). It was emphasized that respondents would not have to pay anything extra if they choose the opt-out. An example of a choice card is presented in Table 3.1.

Table 3.1: Example choice card

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Option 2</th>
<th>Current situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra days per week</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Boiling</td>
<td>No</td>
<td>For infants</td>
<td>For infants</td>
</tr>
<tr>
<td>Increase in monthly water bill</td>
<td>Birr 10</td>
<td>Birr 3</td>
<td>Birr 0</td>
</tr>
<tr>
<td>I prefer</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

The design of the choice experiment was first pretested and subsequently implemented in June 2010 through 170 in-person interviews in three zones in Hawassa city distinguished by the HWSSO: Misrak Wukro, Manaharia and Mahal Piasa. The response rate was 100 percent, which is not unusual for this kind of stated preference research in a developing country (Whittington, 1998). Although up-to-date statistical information about household characteristics is not available for the different zones, the zones differ considerably in terms of their socio-economic status. Misrak Wukro is generally considered the poorest part of the city, whereas households in Manaharia and Mahal Piasa are better off. In each zone, first households with a private compound were identified and secondly from this a random selection chosen. A third sample selection criterion was that a more or less equal share of men and women in the households was to be interviewed to test for possible gender effects in domestic water supply valuation. Trained local enumerators were used for the interviews. After data screening, 145 of
the 170 interviews could be used in the analysis (50 from Misrak Wukro, 52 from Manaharia and 43 from Mahal Piasa). The results are presented in the next section.

3.4 Results

3.4.1 Sample characteristics

Table 3.2 reports the socio-economic characteristics of the interviewed households. A majority of the respondents (63%) is female and respondents are, on average, 34 years old. Respondents were between 18 and 80 years of age. Average monthly household income is USD 145. Given the average household size, this equals a monthly per capita net income of USD 26. This is lower than the World Bank's international (PPP-adjusted) poverty line of USD 1.25 per day.

Table 3.2: Key demographic and water supply characteristics of the sample population

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean</th>
<th>St. dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share female respondents</td>
<td>0.63</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average age</td>
<td>33.9</td>
<td>6.0</td>
<td>18</td>
<td>80</td>
</tr>
<tr>
<td>Average household size</td>
<td>5.6</td>
<td>3.3</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Household income (Ethiopian Birr/month(^1))</td>
<td>2388</td>
<td>1665</td>
<td>50</td>
<td>8000</td>
</tr>
<tr>
<td>Average daily water consumption (liters)</td>
<td>99.2</td>
<td>87.3</td>
<td>10</td>
<td>800</td>
</tr>
<tr>
<td>Average number of days water supply</td>
<td>4.1</td>
<td>1.7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Average monthly water bill (Birr(^1))</td>
<td>36.2</td>
<td>31.2</td>
<td>2.0</td>
<td>210.0</td>
</tr>
<tr>
<td>Average monthly aversion costs (Birr(^1))</td>
<td>32.1</td>
<td>57.2</td>
<td>0</td>
<td>252</td>
</tr>
</tbody>
</table>

\(^{1}\) One Birr equaled approximately USD 0.06 at the time of the study.

Turning to the household’s domestic water supply characteristics, consumption varies between 10 and 800 liters per day. On average, a household has 4 days per week access to drinking
water supply and they pay USD 2.2 per month for their water bill. This is 1.5 percent of their monthly household income. Most households (58%) also spend every month a considerable amount of money on substitutes such as bottled water.

3.4.2 Aggregate choice model

All 145 respondents are interested in and willing to pay extra for improved water supply services and completed all 12 choice tasks. No protest voters were encountered in the data, i.e. respondents who consistently chose the opt-out alternative on all choice occasions. Across all 1740 choice occasions, the opt-out was chosen in 12 percent of the cases. As expected in unlabelled choice experiments, an equal distribution of choices is found between the two hypothetical alternatives. Choice behaviour was modelled using a combination of random parameter and error component models, accounting for the panel data structure of the choice model. For efficiency purposes, the models were estimated using a Halton sequence of 100 replications in a quasi-Monte Carlo maximum likelihood simulation (Bhat, 2001) in NLOGIT version 4.0.

Two models are presented in Table 3.3. The first model only includes the attributes, whereas the second model includes additional socio-demographic respondent characteristics. Several possible interactions between the attributes and socio-demographic respondent characteristics were tested for their statistical significance. The model that came out best from a statistical point of view after systematic testing of all possible interactions, i.e. including statistically significant variables only at the 10 percent level, is presented in Table 3.3.
Table 3.3: Estimated choice models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model including water supply characteristics</th>
<th>Model including household characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>ASC</td>
<td>2.611***</td>
<td>0.425</td>
</tr>
<tr>
<td>Water supply (1,2,3 extra days/week)</td>
<td>1.309***</td>
<td>0.134</td>
</tr>
<tr>
<td>Standard deviation (Normal distributed)</td>
<td>0.792***</td>
<td>0.218</td>
</tr>
<tr>
<td>Water quality (dummy: no boiling)</td>
<td>4.805***</td>
<td>0.588</td>
</tr>
<tr>
<td>Standard deviation (Uniform distributed)</td>
<td>7.900***</td>
<td>1.105</td>
</tr>
<tr>
<td>Water bill (3, 5, 10, 15, 20 Birr/month))</td>
<td>-0.251***</td>
<td>0.017</td>
</tr>
<tr>
<td>ASC x household income (Birr/month)</td>
<td>-</td>
<td>0.0004***</td>
</tr>
<tr>
<td>Water supply x living in Misrak Wukro</td>
<td>-</td>
<td>0.296**</td>
</tr>
<tr>
<td>Water quality x female respondents</td>
<td>-</td>
<td>0.822**</td>
</tr>
<tr>
<td>Water quality x aversion costs (Birr/month)</td>
<td>-</td>
<td>0.011***</td>
</tr>
<tr>
<td>Sigma error component</td>
<td>3.191***</td>
<td>0.523</td>
</tr>
</tbody>
</table>

Log likelihood function                      | -1260.349           | -1158.358      |
McFadden pseudo R-squared                    | 0.341               | 0.394          |
N                                             | 1740                | 1740           |

Significance levels: * 10% ** 5% *** 1%.

The models are highly significant. The outcome of the $\chi^2$ is 1302.473 with 7 degrees of freedom and 1506.455 with 11 degrees of freedom for the first and second model respectively. The estimated error component is significant in both models at the one percent level, indicating that respondents perceived the two hypothetical alternatives distinctly from the existing situation. The significant positive outcome of the alternative specific constant (ASC) in both models implies that respondents prefer a change instead of no change from the current situation.

The attribute parameters are highly significant and have the expected signs. Households value the availability of additional drinking water supply and improved quality. An increase in the water bill is, as expected, valued negatively, implying that the utility of the households decreases as the monthly water bill increases. Water supply is included as a continuous variable...
and water quality as a dummy variable with the value one if the water needs no boiling. The attributes are characterized by significant preference heterogeneity. The standard deviations of the water supply service attributes are also significant and sizeable, indicating that we captured unobserved heterogeneity in the random parameter specification. The standard deviation for water supply was best captured by a Normal distribution and the standard deviation for water quality by a Uniform distribution (e.g., Hensher et al., 2005). Based on these coefficient estimates marginal WTP and standard deviations were calculated using the Krinksy and Robb (1986) procedure. Marginal WTP for one extra day of domestic water supply is USD 0.3 per month, while marginal WTP for water that requires no boiling is USD 1.2 per month. This is an increase in the average current water bill of 14 and 53 percent respectively.

The second model shows that including additional explanatory variables significantly improves the model fit. The value of the log likelihood function is reduced further compared to the first model, while the pseudo R-squared increases to almost 40 percent, which is high for this type of models based on cross-section data. The size of the impact of all the attributes (and the ASC) is reduced through the inclusion of the theoretically expected variable household income (richer households are more likely to be willing to pay extra for improved water supply services than poorer households), but also variables such as the zone in which the respondent lives (respondents living in the poorest zone Misrak Wukro attach significantly more value to the improvement of water supply) and current spending on bottled water as an alternative to water supplied by the HWSSO (respondents who spend on average more money on bottled water value a water quality improvement significantly more than respondents who spend less money on bottled water). As expected, women who usually cook and take care of the children in the household value water quality improvements more than men to a level where it does not have
to be boiled anymore for infants. No significant effect could be detected for any of the other socio-demographic respondent characteristics (e.g., age, education or household composition).

3.4.3 Tests of preference stability

One of the main assumptions underlying stated preferences is that respondents know their preferences and that these preferences are stable and coherent (e.g. Brown et al., 2008). This implies that individuals consistently know their preference ordering for a set of goods or services and the rate at which they are willing to trade off good characteristics, such as price and quality. Hence, from a set of alternatives the individual is assumed to be capable of selecting the most preferred one based on its characteristics. Choice behavior at the start of the experiment is expected to be consistent with choice behavior at the end of the experiment (Brouwer et al., 2010b). However, lack of familiarity and experience with the good or service involved and the hypothetical choice setting may undermine these a priori assumptions (Shaikh et al., 2007).

We examine possible learning and preference refinement effects by comparing the scale parameter over the choice sequence using the Swait and Louviere (1993) procedure. The scale parameter is inversely related to the variance of the error term (Louviere et al., 2000). If the scale increases, variance decreases, that is, people are making a more accurate choice between the presented alternatives. Obtaining estimates for the scale parameter therefore provides insight into preference refinement during a choice sequence (Holmes and Boyle 2005). Through repetition respondents are expected to be capable of making more precise and consistent decisions, because they learn about the survey format, the associated hypothetical market and their own preferences (List, 2003).
We test whether the utility parameters $\beta$ and the confounded scale parameters $\mu$ in the estimated choice model presented in Table 3.3 remain the same across the 12 choice tasks. To this end, a pair-wise comparison is performed between the choice tasks. We split the 12 choice tasks in four sets of three tasks. In a first step, the estimated ML-model from Table 3.3 provides efficient estimates for $\mu^1\beta^1$, $\mu^2\beta^2$, $\mu^3\beta^3$ and $\mu^4\beta^4$ and a likelihood function for each set of three choice tasks. Then the scale parameter of, for example, the first set of three choice tasks is normalized to $\mu^1=1$ for identification purposes. In a second step, a pooled model including the first and second set of choice tasks is estimated, which has the effect of imposing equality on the preference parameters ($\beta^1=\beta^2$). In this model the relative scale parameter $\mu^2/\mu^1$ is, however, not set equal to 1. A search procedure over a range of relative scale parameters is applied to estimate the combination of scale and pooled preference parameters providing the best model fit. At each possible relative scale parameter the data for the second set of choice tasks are rescaled such that a ML-model can be estimated. After the best fit model has been identified, a standard chi-square (Likelihood Ratio) test using the log likelihoods of the models from step 1 and the best fit model of step 2 can be used to test the difference between the preference parameters in the two choice sets under the null hypothesis that they are the same.

The third step tests for differences in scale across choice tasks. In principle this step is conditional on accepting the mentioned chi-square test. It requires the estimation of a ML-model for the same pooled model as in step 2, but with equality imposed on both preference and scale parameters this time ($\beta^1=\beta^2$ and $\mu^1=\mu^2$). Again a chi-square test can be applied to compare the log likelihood of the estimated model to the log likelihood of the pooled model with varying scale parameters. The results of this test procedure are presented in Table 3.4.
Table 3.4: Test results equality of preference ($\beta$) and scale ($\mu$) parameters between choice tasks

<table>
<thead>
<tr>
<th>Comparison choice tasks (t)</th>
<th>LL$_{ti}$</th>
<th>LL$_{tj}$</th>
<th>LL$_{ti+j}$ ($\mu_i \neq \mu_j$)</th>
<th>LR-test (10 d.f.)</th>
<th>Reject</th>
<th>Relative scale ($\mu_i/\mu_j$)</th>
<th>Relative var. ($\sigma^2_i/\sigma^2_j$)</th>
<th>LL$_{ti+j}$ ($\mu_i=\mu_j$)</th>
<th>LR-test (1 d.f.)</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1-3 &amp; 4-6</td>
<td>-270.001</td>
<td>-313.504</td>
<td>-605.149</td>
<td>43.29</td>
<td>Yes</td>
<td>1.40</td>
<td>0.51</td>
<td>-605.484</td>
<td>0.67</td>
<td>No</td>
</tr>
<tr>
<td>2 1-3 &amp; 7-9</td>
<td>-270.001</td>
<td>-329.888</td>
<td>-623.483</td>
<td>47.19</td>
<td>Yes</td>
<td>0.41</td>
<td>5.95</td>
<td>-628.437</td>
<td>9.91</td>
<td>Yes</td>
</tr>
<tr>
<td>3 1-3 &amp; 10-12</td>
<td>-270.001</td>
<td>-239.144</td>
<td>-564.342</td>
<td>110.40</td>
<td>Yes</td>
<td>4.01</td>
<td>0.06</td>
<td>-569.474</td>
<td>10.26</td>
<td>Yes</td>
</tr>
<tr>
<td>4 4-6 &amp; 7-9</td>
<td>-313.504</td>
<td>-329.888</td>
<td>-652.259</td>
<td>17.73</td>
<td>Yes</td>
<td>0.42</td>
<td>5.67</td>
<td>-660.236</td>
<td>15.96</td>
<td>Yes</td>
</tr>
<tr>
<td>5 4-6 &amp; 10-12</td>
<td>-313.504</td>
<td>-239.144</td>
<td>-586.327</td>
<td>67.36</td>
<td>Yes</td>
<td>1.83</td>
<td>0.30</td>
<td>-590.800</td>
<td>8.95</td>
<td>Yes</td>
</tr>
<tr>
<td>6 7-9 &amp; 10-12</td>
<td>-329.888</td>
<td>-239.144</td>
<td>-610.236</td>
<td>82.41</td>
<td>Yes</td>
<td>2.65</td>
<td>0.14</td>
<td>-619.113</td>
<td>17.75</td>
<td>Yes</td>
</tr>
</tbody>
</table>

LL: log likelihood; Likelihood Ratio (LR) tests performed at 10% significance level.

$^a$ pooled mixed logit model with scale normalization; $^b$ pooled mixed logit model without scale normalization.
When comparing the estimated choice models between the four sets of choice tasks, both the null hypothesis of equality of preference parameters and scale parameters are rejected at the 10 percent significance level, except the scale parameter between the first and second choice set (last column in Table 3.4). Hence, choice behavior at the start of the choice experiment is different compared to choice behavior at the end of the experiment. Remarkable is especially the increase in variance in the third set of choice tasks (7-9) when examining the relative variance in rows 2 and 4 of Table 3.4. The variance between the first two sets of choice tasks is not significantly different, the variance then makes a significant jump in the third set and decreases significantly again in the last set of choice tasks compared to the first set.

Comparing the estimated choice models across the four choice sets, the model based on the third choice set is the only model where one of the attributes (extra days of water supply) is not statistically significant at the 10 percent level. The attributes are statistically significant in all other choice sets, including the increase in the water bill. Variation across the choice models is primarily found in the interaction terms between the attribute water supply and aversion cost and the attribute water quality and female respondents. These results suggest that overall the attributes remain fairly stable, but not the influence of the socio-demographic characteristics on choice behavior.

3.4.4 Economic welfare measures for water supply improvement policy scenarios

The welfare implications of different water supply improvement policy scenarios were calculated based on the estimated model in Table 3.3 including household characteristics. Currently, consumers receive, on average, drinking water four days per week. In most cases, they have to boil the water before they can drink it. Table 3.5 presents the estimated mean WTP
values associated with improvements in urban water supply services for the average respondent. Standard errors are presented in brackets and are based on the Krinsky-Robb (1986) procedure. WTP for the water supply improvement policy scenarios are presented both in local currency (Birr) and US dollars.

Table 3.5: Estimated WTP for water supply improvements per household per month

<table>
<thead>
<tr>
<th>Policy scenario</th>
<th>Without water quality improvement¹</th>
<th>With water quality improvement²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Birr</td>
<td>USD</td>
</tr>
<tr>
<td>1 extra day per week water</td>
<td>9.8</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>2 extra days per week water</td>
<td>13.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>(2.7)</td>
<td>(0.2)</td>
</tr>
<tr>
<td>3 extra days per week water</td>
<td>17.9</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(1.8)</td>
<td>(0.1)</td>
</tr>
</tbody>
</table>

¹ Water needs boiling for infants. ² Water needs no boiling.

In case only water supply improves, not water quality, urban households are willing to pay, on average, between 25 and 50 percent extra over and above their current monthly water bill of 36.2 Birr (USD 2.2). If water quality improves at the same time, the increase in the water bill households would be willing to pay varies between 60 and 80 percent depending on the extra days of domestic water supply.

If the sample is representative, these values can be aggregated across the population from which it was drawn in order to calculate a total economic value for the policy scenarios. This total economic value can subsequently be compared with the necessary investment and maintenance
costs for the improvement of current water supply services in the city of Hawassa. Based on the factors that turned out to be statistically significant in the second choice model presented in Table 3.2, women are expected to benefit most from improvement scenarios which include water quality. Given the fact that we also found a significant spatial effect for one of the zones, investments in water supply improvement will be most beneficial to households living in the poorest zone Misrak Wukro.

3.5 Conclusions

Reliable water supply services are of paramount importance to the expanding urban populations in the developing world. Improving existing water supply services depends crucially on available financial resources. Cost recovery rates of water supply services provided by public utilities in developing countries are typically low, while demand for more reliable services is high and rapidly growing. This study examined public preferences for improved water services in an urban area in a country with the lowest water supply coverage in Sub-Saharan Africa with the aim to estimate the non-market value of specific water supply service improvements. This economic value can be used in policy and project appraisals of improved supply investment decisions.

This study adds to the limited stock of applied water supply reliability valuation studies. A number of contingent valuation studies exist focusing on willingness to pay for improved water supply services in developing countries, in particular in rural areas and few in urban zones. Almost no studies exist that apply choice experiments. Those that do focus on WTP to avoid domestic water restrictions or improve environmental river quality in developed countries. The
choice experiment in this study estimated compensating surplus welfare measures by asking a random sample of households in Hawassa, Ethiopia for their willingness to pay for improved water supply services, both in terms of quantity and quality. Currently, households have, on average, 4 days per week access to drinking water, and consume almost 100 liters per day for which they pay USD 2.2 per month. Based on the average household size, daily per capita consumption is only 18 liters. This is very low compared to the estimated average per capita consumption levels in the rest of the developing world (UNESCO, 2003). Almost 60 percent of the interviewed households pay every month an equal amount of money for alternative drinking water sources such as bottled water. All households boil their water before they drink it. Water quality is not constant and often unsafe to drink.

Despite significant income constraints, all households appeared to be willing to pay substantially extra for improved levels of water supply, especially those households living in the poorest part of the city with the lowest service levels and who pay more for alternative bottled water. Although estimated preference parameters vary throughout the choice sequence in the choice experiment, choice variability decreases significantly towards the end of the choice experiment. This suggests the presence of learning effects whilst going through the choice tasks. Mean WTP for more reliable water supply varies between 25 and 50 percent over and above their current household water bill. If water quality is improved at the same time, this results in an almost twice as high additional WTP depending on the extra days of water supply. Women who take care of infants in the household value the improvement of water quality to a level where boiling for infants is not necessary anymore most.
Aggregating the estimated individual household WTP values across the total number of households in Hawassa (12,500) under the assumption that the survey sample is representative yields a rough indicator of the total benefits of future investment plans in improved water supply services. The value of these benefits vary from USD 90,000 per year if water supply services are improved by one day only to USD 165,000 per year if water supply is improved by three days to seven days per week. If water quality is improved in the latter case at the same time, the value of the total benefits equals USD 270,000. Discounted over a 25-year time period at a discount rate of 10 percent this amounts to a present value of approximately USD 2.5 million. This discounted value can be compared to the capital costs of any future investment decision in improved water supply services in the city.
4 Informing water harvesting technology contract design using choice experiments

4.1 Introduction

Water harvesting is generally considered a suitable strategy for adapting to water shortages caused by climate change (Lasage, 2015). Typically rainwater harvesting has proven to be an affordable and sustainable solution in places which are suitable to collect rainwater, have dispersed populations and where the costs of developing surface or groundwater infrastructure are high (UNEP, 2005). Rainwater harvesting is defined as a method of inducing, collecting, storing and conserving local surface water runoff for agriculture (Lasage, 2015). However, attempts to spread and intensify this practice have had limited success (Bouma et al., 2015). In Ethiopia, where this study is conducted as part of the European Commission funded project Water Harvesting Technologies Revisited (WHaTeR) to explore the potential for innovation, improvement and upscaling, there has been growing interest in agricultural water management through rainwater harvesting to enhance agricultural productivity and alleviate poverty. Here too though the technology has not been adopted to the extent anticipated (e.g. Hagos et al., 2007; Sisay, 2013), mainly because the program was top-down, technocratic and non-participatory (e.g. Moges et al., 2011; Searnet, 2013). Even in cases where adoption of the

---

5 This chapter has been published as Tarfasa, S., Brouwer, R., Sheremet, O. and Bouma, J. (2017). Informing water harvesting technology contract design using choice experiments. Water Resources Research, 53(10): 8211–8225.
technology was participatory, farmers were not so much consulted as persuaded to accept water harvesting technology on their land (Bewket, 2007), ultimately resulting in their abandonment. Rainwater harvesting is still a relatively new technology to most farm households in Ethiopia and little is known about it at institutional and community level, hence resulting in a low adoption of the technology (e.g. Feyisa, 2014). In order to realize the full benefits of adopting rainwater harvesting technologies to enhance agricultural productivity and reduce vulnerability to water scarcity, it is important to understand the institutional-economic factors that constrain farm households in their decision-making to invest in this technology.

In this case study, a choice experiment is developed with the main aim to test what triggers farm households to invest in and maintain water harvesting ponds on their land. To our knowledge there exists no such study yet, which estimates farm households’ demand for improved water harvesting technology in Ethiopia. In the choice experiment, farmers living in a drought prone area in Ethiopia are offered a menu of contracts, following the approach developed in Tesfaye and Brouwer (2012). The contractual agreements include different combinations of institutional, economic and technological terms and conditions, and are introduced to a random sample of farm households in the study area to assess how to best encourage them to invest in improved water storage and hence reduce their vulnerability to water scarcity.

Analyzing farmers’ choices for alternative contract designs reveals the importance they attach to the various terms and conditions (Tesfaye and Brouwer, 2012). In view of the relatively high illiteracy rate in Ethiopia (according to UNICEF (2014), over the period 2008-2012 on average 37% of the male and 53% of the female population could not read or write), pictograms were
used in the choice experiment with as little text as possible to describe the levels of the contract characteristics. In a control group, the same choice experiment was also administered without the visual aids to test their impact on choice behavior. The first hypothesis tested here is that the way the information was conveyed to farm households in the choice experiment (with and without visual aids) significantly affects choice behavior, in particular among the illiterate population. In addition, the extent to which farmers pay attention to the different contract characteristics, with and without visualizations, was assessed using both self-reported and inferred attribute (non-) attendance approaches (e.g. Hensher and Green, 2011; Campbell et al., 2011). This leads to the second alternative hypothesis tested here that respondents do not pay equal attention to all contract terms and conditions, i.e. there exists attribute non-attendance.

The remainder of this chapter is structured as follows. The next section first describes the methodological design of the study, including the underlying econometric model to estimate farm household demand for the presented contractual agreements to invest in improved water harvesting technology on their land. This is followed by a description of the case study area and data collection procedure in Section 3. Section 4 presents the results, while conclusions are drawn in the final Section 5.

4.2 Methodological design

4.2.1 Choice experiment design

Choice experiments (CEs) are part of the family of stated preference methods (e.g. Hanley et al., 2011). Preferences for existing or new products, technologies or policy programs are elicited using a social survey format, such as in-person interviews (e.g. Bennett and Blamey,
Although CEs were originally developed in marketing research and applied to a wide variety of consumer products in developed countries, their application in developing countries has seen an exponential growth (e.g. Bennett and Birol, 2010). CEs typically ask respondents in a series of choice tasks to choose between two or more alternative policies, which are described based on their relevant characteristics (e.g. Hanley et al., 2003). This allows estimation of public demand for these policies, and preferences for their specific characteristics or attributes. Usually the benefits associated with the policies are described and the costs involved, for example the price individual respondents would have to pay in order to secure the described policy benefits. The inclusion of a monetary price attribute in the CE allows estimation of public willingness to pay (WTP) for the policy. In this study, the alternatives between which farmers in Ethiopia are asked to choose are contracts.

The contract design here aims at investigating farmer’s preferences for different contractual arrangements, which they conclude either with regional governments or non-governmental organizations who provide them with the relevant technical and institutional-economic support to invest in private household ponds. The contract provides farmers opportunities and certainties to reduce their climate vulnerability. The contract specifies the terms and conditions of the provision of credit to finance the construction of “best practice” water ponds. So, the good on offer is water security through the implementation of improved water harvesting technology, which is financed by a proposed micro-credit scheme in the case study area. In exchange for the water security benefits that farmers obtain from adopting improved rainwater harvesting technology, they periodically repay the loan they receive from the preferred credit provider plus interest. Farmers are expected to invest the credit received from the credit provider in the construction of state-of-the-art household water ponds according to the contract.
specifications. Contracts are offered for the duration of 5, 10 and 15 years. The descriptions of the contractual arrangements in the CE are presented in Table 4.1.

Table 4.1: Description of the contractual arrangements in the choice experiment

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract provider</td>
<td>Regional government or non-governmental organization</td>
</tr>
<tr>
<td>Contract duration</td>
<td>5, 10 or 15 years</td>
</tr>
<tr>
<td>Pond capacity</td>
<td>8,000, 27,000, 64,000 or 125,000 liter</td>
</tr>
<tr>
<td>Pond Lining</td>
<td>Plastic or cement</td>
</tr>
<tr>
<td>Pond cover</td>
<td>Yes, no</td>
</tr>
<tr>
<td>Labor input</td>
<td>10, 20 or 30 person days</td>
</tr>
<tr>
<td>Monthly repayment loan</td>
<td>100, 200, 300, 400, 500 Birr per household</td>
</tr>
</tbody>
</table>

The contract characteristics were selected based on an extensive literature review, expert interviews and focus group discussions. They include (i) the contract provider, (ii) contract length, (iii) water pond size, (iv) lining of the water pond, (v) water pond coverage (i.e. whether or not the pond is covered), and (vi) farmer labor input to construct the pond. The repayment rates are set based on group discussions and pretest results conducted before the main survey, and are independent of the contract length, both in terms of the applied statistical design and measurement of behavioral intent. Contract length is primarily included to measure a farm household’s longer-term commitment to maintaining the water pond. The group discussions showed that using land certificates as collateral, farmers can get loans ranging from 2,000 to
5,000 Birr (USD 105-265) at an interest rate of 10 percent for 12 months from the regional government. The monthly payments for such a loan over 12 months vary between 300 and 700 Birr. Based on the focus group discussions, a repayment period of 12 months was considered a constraint given average income levels in the study area. As a result, many if not most farmers are unable to participate in the loan service provided by the local government. Contract duration was therefore extended to range between 5 and 15 years to give farmers enough time to repay the loan, and at the same time assess their longer-term commitment to their investment and maintain the water ponds. The combinations of monthly payments and contract duration in the design reflect the actual situation and imply the provision of loans in the range of 1,000 to 5,000 Birr at a market interest rate of 10 percent.

This is hence also the range applied in the CE, i.e. five price levels were identified in terms of a household’s monthly loan repayment, varying between 100 and 500 Birr (USD5-25). Besides the regional government, international NGOs are considered the most important alternative contract providers in the design of the CE since they have been providing financial and material support to local farmers to build community ponds in the study area. The inclusion of labor input in the contract design is based on the fact that farmers currently already contribute labor days for digging the community ponds. Labor input with levels fixed at 10, 20 and 30 person days per pond is therefore another characteristic of the contract considered in this study as an in-kind payment. Based on field visits to the study area, the community ponds constructed for

6  Birr is Ethiopia’s national currency. One Birr equaled approximately USD 0.0494 at the time of the study.
the farmers in the past differ in the lining and sealant used compared to available alternative technologies, justifying their inclusion as characteristics of the contract.

Contractual arrangements are created by combining the above listed characteristics based on their levels. Because respondents cannot be shown all the possible combinations as choice options, a D-efficient design was generated in Ngene (Scarpa and Rose, 2008), consisting of 9 blocks of 10 choice tasks each. Assuming an underlying multinomial logit (MNL) model, the D-error reduced in the design generating process from an initial value of 0.041 at the start of the simulation to 0.029 for the selected design. Bliemer and Rose (2010) show that D-efficient MNL-based designs also perform well for mixed logit models. For this reason, the design was derived assuming a MNL specification. Each respondent was randomly shown one of these 9 blocks of 10 choice cards. Following previous experiences applying CEs in rural Ethiopia and in order to limit the cognitive burden on farmers (Tesfaye and Brouwer, 2012), each choice card shows two hypothetical contracts along with the option to choose none of the two. Inclusion of this latter ‘status quo’ alternative is instrumental to estimate welfare measures that are consistent with demand theory (Bateman et al., 2002) and not to force respondents to accept the proposed contracts if they are not interested in either of them. During the pretest of the CE, special attention was paid to whether the design included combinations of attributes that were considered unrealistic or unfeasible by respondents.

In order to test choice consistency and potential preference learning (Brouwer et al., 2010), the first choice card was shown again at the end of the choice sequence as the final choice task without telling respondents. In addition, half of the respondents were presented the choice cards without pictograms and the other half with pictograms to test the effect of these visual aids in
the CE. An example of a choice card with pictograms is presented in Figure 4.1. The choice card without the pictograms is exactly the same except for the pictograms.

Figure 4.1: Example choice card with pictograms

The introduction of the CE was identical in both versions. Before the start of the CE, all respondents were shown the same instruction choice card, with visualizations in the sample with visualizations and without visualizations in the sample without visualizations, to explain the objective of the CE and the choice tasks. Enumerators were trained to introduce the CE in the same way to all respondents and guide each respondent in the same way through the example choice card to ensure all respondents clearly understood the choices they were
expected to make in the subsequent CE. After going through the instruction choice card, i.e. before the start of the CE, each respondent was asked if (s)he understood the choice task. Enumerators only proceeded with the CE once the respondent indicated that (s)he fully understood the choice task and (s)he was ready for the CE. They were then shown the sequence of choice tasks in the CE, where the enumerator read out loud the information on each choice card to illiterate respondents, and literate respondents were given the option to read the information on the choice cards self or have the enumerator read it out loud to them.

4.2.2 Econometric choice model

CEs are based on Lancaster’s (1966) attribute-based utility theory. The related econometric basis is found in McFadden’s (1974) Random Utility model (RUM), which can be approximated by an appropriately specified random parameters logit model (McFadden and Train, 2000). The random parameters logit (RPL) model overcomes the main limitations of standard multinomial logit models by accounting for random taste variation, unrestricted substitution patterns and unobserved correlation between alternatives (Train, 2003). Random parameters logit models can represent heterogeneity across individuals in both observed and unobserved preferences, but require computationally intensive procedures to estimate probabilities (Newman, 2003).

The RUM can be described in the context of this specific study as follows. Suppose a farmer faces a choice among \( J \) contractual agreements related to different water harvesting technologies and contract specifications.
The utility $U_{ijt}$ associated with each contract alternative $j$, as evaluated by each farmer $i$ in choice situation $t$ can be specified as in equation (4.1):

$$U_{ijt} = \beta_i' X_{ijt} + \epsilon_{ijt} \quad \forall j \in J$$

(4.1)

where $X_{ijt}$ is the vector of explanatory variables that are observed by the analyst, including the terms and conditions (attributes) of the presented contract alternatives, the socio-economic characteristics of the individual farmer and the descriptors of the decision context and choice task in choice situation $t$ (Hensher et al., 2013). $\beta_i$ and $\epsilon_{ijt}$ are not observed and treated as stochastic influences (Train, 2003). $\epsilon_{ijt}$ is the vector of error terms with an iid extreme value type I distribution. Equation (4.2) describes the probability of individual $i$ selecting alternative $j$ compared to other choice alternatives $k$ given that cost associated with each alternative is $Y_{ijt}$:

$$P_{ij} = \left\{ \frac{\exp[\beta_i' X_{ijt} + \beta_{yi} Y_{ijt}]}{\sum_{k \neq j} \exp[\beta_i' X_{ikt} + \beta_{yi} Y_{ikt}]} \right\} f(\beta_i) d\beta_i \quad \forall j \neq k$$

(4.2)

In order to account for unobserved preference heterogeneity across individual farmers, the coefficients associated with the contract characteristics are specified as random terms, allowing the utility coefficients $\beta_i$ to vary across individuals, applying different mixing distributions with density $f(\beta_i/\theta)$. $\theta$ refers here to the density parameters such as the mean and covariance of $\beta$ in the sample population. Following Hensher et al. (2005), a uniform distribution is used
for the choice attributes’ random dummy variables, the random terms of the other choice attributes, including the alternative specific constant (ASC), follow a normal distribution as this appeared to produce the best statistical fit. The price coefficient is included as a random term too because we are interested in how individual respondents attend to and weigh each individual choice attribute, and not necessarily in the estimation of theoretically correct welfare measures. For efficiency purposes and ensure model stability, the models are estimated using a Halton sequence of 500 replications in a quasi-Monte Carlo maximum likelihood simulation (Bhat, 2001). Even if unobserved heterogeneity is accounted for in a RPL-model, the model may fail to explain the sources of heterogeneity (Hynes et al., 2008). To this end, interactions of respondent specific household characteristics can be included with either choice specific attributes in the utility function or the alternative specific constant to improve the model fit.

Although the RPL is the standard model nowadays to model choice behavior in CEs, the model may not properly estimate attribute coefficients if there are some respondents who do not attend to all attributes. The RPL model will assign zero attribute coefficients to these respondents, which will result in lower average sample coefficients. The inferred attribute non-attendance (ANA) model picks up those attributes that are most frequently overlooked or down-weighted by respondents in their analysis of the presented choice situations. Given the emphasis in this study on how attributes are communicated to respondents, attribute (non)attendance is of particular interest here.

The estimation of an ANA model is done within the Equality Constrained Latent Class (ECLC) framework (Hensher and Greene, 2010). This procedure allows us to derive the implied degree of attribute processing by respondents, relying only on the information incorporated in the
respondents’ choices. It has been argued that the inferred ANA provides less biased estimates of ANA than respondents’ self-reported estimates of the amount of attention they pay to each choice attribute (e.g. Campbell and Lorimer, 2009; Scarpa et al., 2009; Hensher and Greene, 2010; Campbell et al., 2011).

The ECLC approach typically requires the model to include all possible $2^K$ combinations of ignored attributes (where $K$ is the total number of attributes) and, correspondingly, to define an equivalent number of classes in which the coefficients of ignored attributes are set to zero, while other coefficients are constrained to be equal across each class (for a detailed description, see for example Hensher and Greene, 2010 or Campbell et al., 2011). This is modeled using a discrete distribution of the model’s parameters over a finite set of classes. Based on their individual characteristics, respondents are implicitly placed in an unobserved or ‘latent’ class $c$, each with distinct utility parameters (hence $\beta_c$). Within each class, individual decisions over consecutive choice cards are assumed to be independent, while conditional on being in a specific class, choice probabilities are generated using a multinomial logit model specification:

$$\Pr(y_{it} = j \mid \text{class} = c) = P_{ij|c} = \frac{\exp(\beta'_c x_{ijt})}{\sum_{q=1}^J \exp(\beta'_c x_{ijqt})}. \quad (4.3)$$

Probabilities for unobserved class membership may be assumed constant: $\Pr(\text{class}=c) = Q_{ic}$. Correspondingly, calculating the expected value of the class-specific probabilities over classes,
we obtain the unconditional probability of the sequence of choices:

\[
Pr(y_j = j) = P_{ij} = \sum_{c=1}^{C} Q_{ic} \prod_{t=1}^{T} P_{ijtc} = \sum_{c=1}^{C} Q_{ic} \prod_{t=1}^{T} \frac{\exp(\beta_c'x_{ijt})}{\sum_{q=1}^{Q} \exp(\beta_c'x_{ijq})}.
\]

(4.4)

Using Bayes theorem, it is possible to obtain the posterior estimate of individual class membership probabilities:

\[
Pr(\text{class} = c | \text{choices, data}) = Q_{ic}^* = \frac{Q_{ic} \prod_{t=1}^{T} P_{ijtc}}{\sum_{m=1}^{M} \prod_{t=1}^{T} P_{ijtm}},
\]

(4.5)

and given the conditional estimates of the class probabilities and class-specific parameters, we can further calculate an individual-specific posterior estimate of the parameters:

\[
\hat{\beta}_i = \sum_{c=1}^{C} Q_{ic}^* \hat{\beta}_c.
\]

(4.6)

Individual preference heterogeneity manifests itself in this case in the way how respondents analyze the attributes in each choice situation, in particular how much attention they pay to each of the attributes. To model inferred ANA, the ECLC model is estimated with several latent classes representing different heuristic rules for attribute processing. A range of classes is created where the first class is associated with respondents who consider all attributes, then several more classes are associated with respondents who ignore one of the attributes, and the last class with those who ignore all attributes. In accordance with the underlying assumption
of non-attendance, the coefficients for ignored attributes in the respective class have the value zero. Conversely, the values of the non-zero coefficients are constrained to be equal across all classes, so that the coefficients capture only the attribute-processing strategies and not preference heterogeneity across respondents. The estimated individual latent class membership probabilities provide an estimate of the weights that each individual assigns to each attribute, or in other words, the probability that an individual attends to a particular attribute. Summed class probabilities across all classes where a particular coefficient is zero (including the class where all attributes are ignored) provide an estimate of the share of respondents who ignore the given attribute.

Due to the limited data size and the large number of attributes in our CE design (yielding 121 possible combinations of non-attended attributes), we aimed to estimate the most parsimonious and logical model, and therefore limit the model to the latent classes where none, all, or only one attribute is ignored. A similar approach is applied by Scarpa et al. (2012). Including all 121 possible combinations is technically not feasible to estimate. This avoids the risk of mining for the best model specification over a limited sample size or finding a model that is intractable for interpretation. This approach gives us nine classes in total: class 1 is defined for all attributes attended, class 2 to 8 where every time one of the choice attributes is ignored and class 9 assumes that none of the attributes received any attention. Due to the constraint on the sum of latent class probabilities, the probability for class 9 is calculated after the model estimation. How individual attribute class probabilities would change as a result of including additional combinations of attribute classes is not tested further here. Using this approach, the main goal is to see how class membership, and hence the role individual attributes play in the choice process, differs across the two samples with and without the visual aids.
4.3 Case study area and data collection

The study is carried out in the administrative district Halaba, located in the Southern Nations and Peoples Regional state, 310 km south of Addis Ababa, the capital city of Ethiopia. The area of 640 square kilometers is located in a relatively flat, semi-arid region, crossed by the river Bilate. The average elevation is 1,800 meters above sea level (m.a.s.l.), ranging from 1,550 to 2,150 m.a.s.l. In combination with the rainfall intensity, the gentle slopes generate sufficient overland flow to be stored in ponds, lending itself as a suitable technique to reduce water scarcity that limits agricultural production in the area.

The total population in the district was estimated in 2007 at just over 232 thousand people, with almost all inhabitants residing in rural areas. The rural villages are accessible in the dry season by means of a dry weather road (Deneke and Abebe, 2008). There are two rainy seasons: one from March to May with an average rainfall of 240 mm (standard deviation of 120 mm) and one from July to September with an average rainfall of 320 mm (standard deviation of 135 mm) (Amha, 2006). The latter period is considered the main rainy season. The average annual temperature ranges from 17.6 to 22.5 degrees Celsius, and the average evaporation rate is about 1,750 mm per year (Shewangizaw and Michael, 2010). The soils in the study area are fertile. Besides teff, the main food grain in Ethiopia, other food crops grown in the study area are maize, wheat, barley, beans and chili pepper (Amha, 2006). In the absence of rivers in large parts of the area, communities largely depend for their drinking water supply either on community ponds or increasingly on groundwater from public boreholes. However, the groundwater table is deep (97-360 meters), the soil is sandy, hence accelerating seepage, and dependency on deep water boreholes entails a high risk of dental fluorosis (Kocanda, et al., 2013). Farmers use private ponds for supplementary irrigation of higher value crops and
livestock drinking. In addition, water harvesting is used occasionally for domestic consumption, such as washing clothes, but not as drinking water.

In this study, 150 farm households without private water ponds living in 12 different villages in the case study area were surveyed and interviewed in person following a stratified random sampling procedure. The households were proportionally divided over the 12 villages and one half answered the textual/numerical choice cards, while the other half answered the choice cards with visualizations (pictograms). Blocks of choice sets were assigned randomly to respondents in the 12 villages. Between 12 and 13 interviews were carried out per village, and each block was answered by 16 or 17 respondents. The head of each of these farm households was interviewed by a trained enumerator based on a thoroughly pre-tested questionnaire. Each interview lasted between 30 and 45 minutes.

4.4 Results

4.4.1 Sample characteristics

The split samples’ household and farm characteristics are presented in Table 4.2. No official statistics are available to assess the representativeness of the samples. Despite the random sampling procedure, a few significant differences exist between the two sub-samples, most importantly related to monthly household income. Differences were tested using the non-parametric Mann-Whitney test statistic. Test results reported here are available from the authors. The sample who received the version of the CE with the visualizations has, on average, an income level that is 33 percent lower than that of the sample who received the version without the visualizations. This is partly related to respondents’ education level. However, the standard deviation of the mean income statistic for the sample that received the visual aids is
very high with a variation coefficient of 84 percent compared to 49 percent for the sample without the visual aids, indicating that the mean value is not very accurate and there exists a substantial spread in the self-reported income levels in this sub-sample. Moreover, closer inspection of the self-reported income values reveals that almost a quarter of the respondents in the sample with visualizations refused to disclose their income level, whereas all respondents except one in the other sample stated their income. Also the share of respondents who cannot read or write is substantially (22%) higher in the sample with the visualizations. These differences will therefore have to be controlled for when comparing the choice modeling results in the two split samples.

The other differences between the two sub-samples are limited. The sample with attribute visualizations includes, on average, slightly older respondents from smaller households with slightly less primary education, but more completed secondary school education. The share of respondents having access to credit or receiving remittances from family members abroad is more or less the same across the two samples.
Table 4.2: Sample household and farm characteristics across the two sub-samples

<table>
<thead>
<tr>
<th>Household and farm characteristics</th>
<th>Without attribute visualizations</th>
<th>With attribute visualizations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General socio-demographic household characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average age (years)</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>Share of male respondents (%)</td>
<td>85</td>
<td>79</td>
</tr>
<tr>
<td>Average number of household members (persons)</td>
<td>6.7</td>
<td>5.9</td>
</tr>
<tr>
<td>Share of respondents who cannot read or write (%)</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>Share of respondents with primary school education (%)</td>
<td>57</td>
<td>41</td>
</tr>
<tr>
<td>Share of respondents with secondary or higher education (%)</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Average monthly household income (Birr) (st. dev.)</td>
<td>898 (441)</td>
<td>599 (506)</td>
</tr>
<tr>
<td>Share of households with off-farm income activities (%)</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Share of respondents receiving remittance (%)</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Share of households having (access to) credit (%)</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td><strong>Farm characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of households with a land use certificate (%)</td>
<td>95</td>
<td>97</td>
</tr>
<tr>
<td>Average size of land owned by a household (ha)</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Average number of crops grown per household</td>
<td>1.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Average number of livestock (TLU)</td>
<td>2.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Share facing water shortages (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>once every 2 years</td>
<td>68</td>
<td>55</td>
</tr>
<tr>
<td>once every 5 years</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>Share with dead livestock due to droughts (%)</td>
<td>48</td>
<td>43</td>
</tr>
<tr>
<td>Average annual drought damage (Birr/household) (st. error)</td>
<td>2,893 (190)</td>
<td>3,812 (426)</td>
</tr>
<tr>
<td>Share expecting droughts to increase in the next 10 years (%)</td>
<td>69</td>
<td>49</td>
</tr>
<tr>
<td>Number of observations</td>
<td>75</td>
<td>75</td>
</tr>
</tbody>
</table>

Turning to the farm characteristics, differences between the two samples are very small and in most cases negligible. Similar sample shares have a land use certificate, necessary to get access to credit, and have on average the same amount of land per household, growing on average the same number of crops. Livestock ownership is included as another measure of household wealth and is measured in Tropical Livestock Units (TLU). The sub-sample with attribute visualizations has slightly more livestock (3.4 TLU) than the sub-sample without (2.9 TLU).
Finally, self-reported water shortage in the dry season over the past 2 years is somewhat higher in the sample without visualizations than in the sample with visualizations. The same applies to the share of farm households who faced dead livestock as a result of droughts. However, the average annual damage costs are higher in the sample with visualizations. The difference is not significant though at the 5 percent level. A majority of 69 percent of the respondents in the sample without visualizations and 49 percent in the sample with visualizations expect rainfall shortages to increase in the next 10 years.

4.4.2 Choice modeling results
A majority of the farmers are interested and willing to enter into a contractual agreement to invest in improved household water ponds to reduce their vulnerability to water scarcity and thereby improve their livelihoods. Sixteen of the 150 respondents (11%) were not interested in the contract and consistently chose the opt-out, either because they could not afford to pay or their land was considered too small to build a new water pond. As expected, an equal distribution of choices is found between the two hypothetical contract alternatives. The first contract alternative is chosen in 45 percent of the choice occasions and the second in 44 percent of the cases, confirming there is no selection bias present in the hypothetical contract alternatives due to design errors. As expected, the higher the payment level, the lower the share of farmers choosing these more expensive alternatives. Note that average monthly income in the sample with visualizations is close to the highest bid level in the CE (Table 4.2). As a result, alternatives with the highest bid level were chosen in only 4.9 percent of the choice occasions in the sample with visualizations and in 3.8 percent of the choice occasions in the sample without the visualizations.
As an aside, comparing the alternative chosen in the first choice task with the one in the 11th identical last choice task, the consistency of the choices is remarkably high. Eighty-eight percent of the respondents chose the same contract alternative in the identical first and last choice task. This is higher than the consistency rates found in previous studies (e.g. Brouwer et al., 2010). No significant influencing factors could be detected when regressing consistent and non-consistent choices on a number of socio-demographic and socio-economic respondent characteristics (e.g. age, gender, education, farm characteristics) and design characteristics (levels of the choice attributes), possibly due to the limited number of non-consistent choices.

Two models are estimated, splitting the sample households as before into two (those who received visual aids and those who did not) as we wanted to see what effect the visual aids have on choice behavior (our first hypothesis). The contract provider, pond lining and pond cover are included in the models as dummies, whereas contract length, labor input, capacity of the pond and payment are included in the model as continuous variables. In addition to the characteristics of the proposed contract, also the most relevant covariates are included in Table 4.3, namely respondents’ household income and literacy rate, based on the differences observed in Table 4.2. The former is interacted with the price attribute and the latter with the ASC to assess the overall impact of literacy on choice behavior. Household income and hence ability to pay is expected to influence a respondent’s WTP a specific contract price, while there is no expectation beforehand to assume that a respondent’s literacy rate would affect one or more specific contract attributes. No significant effects could be detected for other influencing factors, in particular farmers’ current risk exposure and damage costs and perception of future droughts.
Table 4.3: Estimated random parameters logit choice models for the two sub-samples including covariates

<table>
<thead>
<tr>
<th>Variables</th>
<th>Without attribute visualizations</th>
<th></th>
<th></th>
<th></th>
<th>With attribute visualizations</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient estimate</td>
<td>St. error</td>
<td>St. dev.</td>
<td>St. dev.</td>
<td>Coefficient estimate</td>
<td>St. error</td>
<td>St. dev.</td>
<td>St. dev.</td>
</tr>
<tr>
<td>ASC</td>
<td>3.548***</td>
<td>0.751</td>
<td>3.211***</td>
<td>0.562</td>
<td>5.521***</td>
<td>0.928</td>
<td>2.834***</td>
<td>0.548</td>
</tr>
<tr>
<td><strong>Choice attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contract provider = regional government</td>
<td>0.417***</td>
<td>0.146</td>
<td>0.689</td>
<td>0.489</td>
<td>0.059</td>
<td>0.194</td>
<td>1.837***</td>
<td>0.337</td>
</tr>
<tr>
<td>Contract duration (years)</td>
<td>-0.035**</td>
<td>0.017</td>
<td>0.040</td>
<td>0.029</td>
<td>0.034*</td>
<td>0.017</td>
<td>0.046</td>
<td>0.034</td>
</tr>
<tr>
<td>Pond capacity (m$^3$)</td>
<td>0.013***</td>
<td>0.003</td>
<td>0.019***</td>
<td>0.004</td>
<td>0.003**</td>
<td>0.002</td>
<td>0.007**</td>
<td>0.003</td>
</tr>
<tr>
<td>Pond lining = cement</td>
<td>0.139</td>
<td>0.163</td>
<td>1.419***</td>
<td>0.368</td>
<td>-0.406**</td>
<td>0.169</td>
<td>1.502***</td>
<td>0.333</td>
</tr>
<tr>
<td>Top of pond = covered</td>
<td>0.788***</td>
<td>0.187</td>
<td>1.720***</td>
<td>0.311</td>
<td>1.182***</td>
<td>0.211</td>
<td>1.949***</td>
<td>0.360</td>
</tr>
<tr>
<td>Non-monetary payment (labour days)</td>
<td>0.008</td>
<td>0.008</td>
<td>0.023</td>
<td>0.015</td>
<td>-0.006</td>
<td>0.009</td>
<td>0.038***</td>
<td>0.014</td>
</tr>
<tr>
<td>Monthly payment (Birr)</td>
<td>-0.003***</td>
<td>0.001</td>
<td>0.004***</td>
<td>0.001</td>
<td>-0.004***</td>
<td>0.001</td>
<td>0.005***</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC x 1 if respondent is illiterate</td>
<td>-1.419**</td>
<td>0.688</td>
<td></td>
<td></td>
<td>-0.514</td>
<td>0.868</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payment x 1 if income is higher than average</td>
<td>0.002</td>
<td>0.001</td>
<td></td>
<td></td>
<td>-0.002</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model summary statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-616.168</td>
<td></td>
<td></td>
<td></td>
<td>-562.596</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.320</td>
<td></td>
<td></td>
<td></td>
<td>0.379</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi square statistic (18 d.o.f.)</td>
<td>580.375***</td>
<td></td>
<td></td>
<td></td>
<td>687.518***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>825</td>
<td></td>
<td></td>
<td></td>
<td>825</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance levels: *p<0.10; **p<0.05; ***p<0.01
The RPL choice models are estimated in NLOGIT version 5.0, accounting for the panel data structure of the model and unobserved preference heterogeneity. Both models are highly significant as can be seen from the chi-square test statistic. The pseudo R-square is high for this type of cross-section analysis and slightly higher for the sample with the attribute visualizations than for the model without the attribute visualizations. All 11 choices are included in the model estimation procedure, yielding 825 observations per sub-sample, i.e. 75 respondents making 11 choices in each sub-sample. The reason for including the 11 instead of 10 choices is that a comparison of the RPL models based on 10 and 11 choices shows that the latter explain slightly more of the choice variation than the latter. This suggests that for some of the respondents who answered differently in the 11th choice task (12% of the sample population) their choice in the repeated choice task seems to better reflect their preferences, further reducing the variance in choice behavior.

As expected based on the relatively low number of opt-out choices, the ASC’s relating to the hypothetical contract alternatives are positive and highly significant in both sub-samples. However, they are also surrounded by substantial preference heterogeneity as can be seen from the significant standard deviation of the random parameters. The contractually requested in kind contribution (household labor) is not statistically significant in either model and hence

7 The Akaike Information Criterion (AIC) adjusted for the number of observations is reduced from 1.571 to 1.537 for the RPL models with 10 and 11 choice tasks, respectively in the sample without attribute visualizations, and from 1.442 to 1.408 in the sample with attribute visualizations. Similarly, McFadden’s pseudo R² increases by 4 percent from 0.307 to 0.320 in the sample without attribute visualizations, and from 0.366 to 0.379 in the sample with attribute visualizations.
plays no role in contract selection, also not when interacted for example with covariates such as household size as a proxy for the availability of household labor. The finding that this contract condition is neglected by the farmers in their decision-making process suggests that it is not considered important or that the opportunity costs of labor may be low in the study area. This latter explanation could not be verified as this was not further investigated during the survey. The pretest results showed that this variable might not be significant at the 10 percent level, but it was kept in as a contract condition in view of the fact that it is common practice to provide own labor in building community and private ponds. Moreover, the coefficient estimate was negative based on the pretest results, even if farmers in both sub-samples do not seem to care much about this contract condition. The monetary payment (credit payback with interest) is highly significant and as expected negative: the higher the payback sum, the less likely a farmer will choose a contract alternative. Although the size of the coefficient’s standard deviation is considerable, suggesting that some respondents attach a positive utility to the payment, the normal distribution is highly significant and produced a better statistical fit than assuming that the coefficient is constant across respondents as a fixed effect or assuming a bounded triangular distribution as in Kragt (2013). Contrary to expectations, no significant income or wealth effect on WTP could be detected in the estimated choice models.

In the sub-sample without the attribute visualizations the specific lining of the water pond does not play a significant role, while the contract provider is not considered important in the sub-sample with the attribute visualizations. Remarkably, the standard deviations of the corresponding random parameters in both samples are statistically significant suggesting that there exists variation in the valuation of these attributes across farmers. In the sub-sample without the visual aids, farmers prefer that the contracts are provided by the regional...
government instead of international NGO’s (the baseline category), possibly due to the fact that the regional government also provides other inputs already such as fertilizer and seeds (e.g. Raman and Dessie, 2013), and in the sub-sample with the visual aids plastic lining (the baseline category) is preferred to cement. This latter finding can be explained by the fact that the plastic lining retains water longer than cement, which often cracks and results in water leakage during the dry season. Note that this specific contract characteristic is surrounded by significant preference heterogeneity though.

Farmers in both samples prefer, all else being the same, bigger water ponds that are covered to avoid evaporation, health risks associated with attracting malaria mosquitoes and children from drowning. Whereas the positive impact of the specified size of the pond on choice behavior is bigger in the sample without visualizations, pond cover is valued higher in the sample with visual aids. Both contract characteristics are surrounded by significant preference heterogeneity. Contract duration is also significant in both samples, but short-term contracts are preferred in the sample without visual aids and long-term contracts in the sample with visual aids.

Finally, literacy rate has a significant effect on choice behavior in the sample without visualizations and not in the sample with visualizations. Illiterate farmers who have not been shown the pictograms are significantly less likely to choose one of the contract alternatives and more likely to stay with the status quo. The fact that the same effect is not statistically significant in the sample with the visual aids might suggest that the role of illiteracy has been offset when using visual aids. However, this result has to be interpreted with the necessary care.
in view of the fact that the literacy rate in the sub-sample without the visualizations is somewhat lower.

We also estimated the choice models for the two sub-samples of literate and illiterate farmers and included a dummy for the version of the CE (with or without visualizations). These choice model results, presented in the Annex to this chapter, confirm that the application of visual aids only significantly affects the sub-sample of illiterate farmers and not the sub-sample of literate farmers. Illiterate survey respondents are more likely to choose one of the hypothetical contract alternatives if they were shown the version with visual aids.

4.4.3 The influence of visual aids on choices for the contractual agreements

Comparing choice consistency across the two split samples, those who received the version of the CE including the attribute visualizations were slightly more consistent in their choices (89.5%) than those who received the version without visualizations (86.5%). The Swait and Louviere (1993) test procedure was carried out to compare the results from the split samples with and without the visualizations to formally test the first hypothesis. In the two step approach, the first hypothesis of equal preference parameters is convincingly rejected at the 1 percent significance level (Likelihood Ratio test statistic is 46.952 with 16 degrees of freedom), implying that the two estimated choice models are not the same, and the use of visual devices significantly affects choice behavior. Based on the outcome of this first hypothesis test, we know that the two models are not the same, but we are unable to identify whether this is due to differences in preference or scale parameters (Louviere et al., 2000). Hence, although somewhat redundant, we also report the outcome of the second hypothesis of equal scale parameters (Likelihood Ratio test statistic is 2.222 with only 1 degree of freedom). This second
hypothesis cannot be rejected at the 10 percent level, and suggests that there exist no statistically significant differences in the error variance of the estimated choice models.

### 4.4.4 Stated and inferred attribute (non-)attendance

The upper part of Table 4.4 shows the number of attributes that guided farmers’ choices in the CE (answers are missing for 5 respondents). Although respondents were allowed to select multiple attributes, 57 percent of the sample without visualizations stated one attribute only. This was less than a quarter in the sample with visualizations. Hence, more attributes were attended to in the sample receiving the visual aids. The bottom part of Table 4.4 shows how many respondents indicated that a specific choice attribute played a role and thus was attended to during the choice-making process. Because respondents were allowed to select as many attributes as they wanted, the numbers in the bottom part do not add up to the total number of respondents in each sub-sample. It is important to point out that not mentioning an attribute does not necessarily mean that the attribute was ignored. Alemu et al. (2012) find, for example, that respondents may not ignore attributes completely even if they claim so in the survey. The relatively low shares of attribute attendance presented here therefore have to be interpreted with the necessary care. Only one respondent indicated that none of the attributes guided his choices, suggesting he ignored all attributes.

Looking at the numbers, we can conclude that there is not a single attribute that was considered by all respondents. In other words, our second alternative hypothesis that respondents did not pay attention to all attributes is confirmed. Moreover, there is considerable variation in the degree of attention paid to different attributes: while about half of the respondents primarily considered pond cover and pond lining during their choices, only 5 respondents paid attention
to the requested household labor. In line with earlier findings in the literature (e.g. Campbell et al., 2008; Scarpa et al., 2009; Carlsson et al., 2010; Kragt, 2013), the monetary payment attribute was not attended by all respondents either. Only between 25 and 31 percent of the respondents stated that they accounted for this attribute in their choices. Still, pond cover, pond capacity and pond lining are among the more attended attributes compared to the somewhat lower levels of attendance to the institutional contract characteristics such as the contract provider and contract duration. Comparing stated ANA across sub-samples, we furthermore observe that the degree of non-attendance is significantly lower for pond capacity, lining and contract duration, that is, for three out of the seven attributes that received visual support in the split sample experiment.

Table 4.4: Stated Attribute Attendance Across the two Sub-Samples

<table>
<thead>
<tr>
<th>Number of attributes attended</th>
<th>Without attribute visualizations</th>
<th>With attribute visualizations</th>
<th>Equality of proportions p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of respondents</td>
<td>Share (%)</td>
<td>Number of respondents</td>
</tr>
<tr>
<td>None</td>
<td>1</td>
<td>1.4</td>
<td>0</td>
</tr>
<tr>
<td>1 attribute</td>
<td>42</td>
<td>57.5</td>
<td>17</td>
</tr>
<tr>
<td>2 attributes</td>
<td>17</td>
<td>23.3</td>
<td>19</td>
</tr>
<tr>
<td>3 attributes</td>
<td>4</td>
<td>5.5</td>
<td>26</td>
</tr>
<tr>
<td>4 attributes or more</td>
<td>9</td>
<td>12.3</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>100.0</td>
<td>72</td>
</tr>
</tbody>
</table>

Attended choice attribute

<table>
<thead>
<tr>
<th></th>
<th>Without attribute visualizations</th>
<th>With attribute visualizations</th>
<th>Equality of proportions p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of respondents</td>
<td>Share (%)</td>
<td>Number of respondents</td>
</tr>
<tr>
<td>Contract provider</td>
<td>11</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Contract duration</td>
<td>12</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Pond capacity</td>
<td>15</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>Pond lining</td>
<td>21</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Pond cover</td>
<td>41</td>
<td>55</td>
<td>36</td>
</tr>
<tr>
<td>Labor input</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Monthly payment</td>
<td>23</td>
<td>31</td>
<td>19</td>
</tr>
</tbody>
</table>

1 Multiple attribute selection was possible, hence the reason that the shares do not add up to 100%.
We also estimated the RPL choice models presented in section 4.2 for the two sub-samples accounting for self-reported ANA, by restricting the attributes respondents did not list to attend to zero in the estimation procedure. However, the latter estimated choice models were less robust and explained less of the observed choice behavior than the conventional RPL models without stated ANA. Similar results were found, for example, by Carlsson et al. (2010), Scarpa et al. (2012), and Alemu et al. (2013). These models are presented in the Annex to this chapter.

The ECLC-model to assess inferred ANA is estimated using the software Biogeme (Bierlaire, 2003). The usual search across the range of starting values was performed and the resultant models were compared on the grounds of their convergence and final likelihood. The statistically best-fit model estimation results are presented in Table 4.5. As before, the estimation results are reported separately for the two sub-samples. Note that for some class probabilities the estimation procedure reports very small numbers with very high standard errors, which is due to the optimization of the multi-parameter likelihood function with several close-to-zero coefficients.
Table 4.5: Estimated inferred attribute non-attendance ECLC models across the two sub-samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>Without attribute visualizations</th>
<th>With attribute visualizations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient estimate</td>
<td>St. error</td>
</tr>
<tr>
<td>ASC</td>
<td>-2.340***</td>
<td>0.269</td>
</tr>
<tr>
<td>Contract provider = regional government</td>
<td>4.240**</td>
<td>1.620</td>
</tr>
<tr>
<td>Contract duration (years)</td>
<td>0.002</td>
<td>0.015</td>
</tr>
<tr>
<td>Pond capacity (m$^3$)</td>
<td>0.003**</td>
<td>0.001</td>
</tr>
<tr>
<td>Pond lining = cement</td>
<td>-0.198</td>
<td>0.124</td>
</tr>
<tr>
<td>Pond cover = covered</td>
<td>0.972***</td>
<td>0.132</td>
</tr>
<tr>
<td>Labor input (labour days)</td>
<td>0.001</td>
<td>0.007</td>
</tr>
<tr>
<td>Monthly payment (Birr)</td>
<td>-0.004***</td>
<td>0.001</td>
</tr>
<tr>
<td>PrClass1: all attributes attended</td>
<td>0.035</td>
<td>0.026</td>
</tr>
<tr>
<td>PrClass2: contract provider ignored</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PrClass3: contract duration ignored</td>
<td>0.450***</td>
<td>0.081</td>
</tr>
<tr>
<td>PrClass4: pond capacity ignored</td>
<td>0.057*</td>
<td>0.030</td>
</tr>
<tr>
<td>PrClass5: pond lining ignored</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PrClass6: pond cover ignored</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PrClass7: labor input ignored</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>PrClass8: monthly payment ignored</td>
<td>0.019</td>
<td>0.027</td>
</tr>
<tr>
<td>PrClass9: all attributes ignored</td>
<td>0.440***</td>
<td>0.094</td>
</tr>
<tr>
<td>Initial Log-likelihood</td>
<td>-829.309</td>
<td></td>
</tr>
<tr>
<td>Final Log-likelihood</td>
<td>-626.397</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio test</td>
<td>405.8</td>
<td>$p&lt;0.001$</td>
</tr>
<tr>
<td>Number of observations</td>
<td>825</td>
<td></td>
</tr>
</tbody>
</table>

Significance levels: *$p<0.10$; **$p<0.05$; ***$p<0.01$
The first important observation from Table 4.5 is that in both sub-samples the share of respondents who attended to all attributes (class 1 probability) is very low and not statistically significant, which means that all respondents ignored certain attributes during their decision-making. At the same time, the probability of not attending to all attributes is quite high in both sub-samples, suggesting that our second alternative hypothesis cannot be rejected. This is somewhat contrary to the finding reported before in Table 4.4 that only one respondent stated not to attend to any of the attributes. A possible explanation may be that since the ECLC model is specifically used to address ANA, the coefficients are estimated only for those respondents who attended to at least some of the attributes. The high probability may also be related to the fact that class membership is based on an ex-post calculation procedure as described in Section 4.2.2. This would imply that other class probabilities, which are currently zero or non-significant, have been underestimated. This might to some extent be the case judging by the results presented in Table 4.4. Alternatively, it may be that some respondents truly ignored some attributes and had low weights for others, and the optimization routine used to estimate the ECLC model picked this up as an indication that they ignored all attributes. This would be in line with some of the small probability values and the non-significant class probabilities in other cases where respondents suggested that they ignored an attribute.

The second important observation is that almost all non-attendance shares, including the share of respondents ignoring all attributes, are considerably lower for the sub-sample of respondents who were shown the choice cards with pictograms. That is, the attributes for which visual aids were developed are considered more attentively by the respondents, as can also be observed for the stated attribute attendance in Table 4.4. This confirms our findings in the previous section 4.3 and our first alternative hypothesis that the way the information in the CE is
conveyed to respondents significantly affects their choice behavior. Third, in the absence of visualizations, farmers pay more attention to pond cover and the monthly payment. Fourth, in both sub-samples, contract duration is the most ignored contract characteristic. Finally, given the differences in the inferred degrees of ANA across the sub-samples, it is not surprising to observe that the estimated attribute coefficients are also different.

4.5 Conclusions

Introducing and implementing improved water harvesting technology on farm land in drought prone areas, as in this case study, is expected to be more effective and last longer if farm households are involved in the design of the water pond and better use is made of the available technical knowledge and expertise. The main objective of this study was to inform policymakers in Ethiopia about the most important terms and conditions to incentivize farmers to enter into a contractual agreement to invest in and maintain water ponds on their land and reduce their vulnerability to droughts. The novelty of the study is that we use a CE to identify the importance farmers attach to both the technical and institutional-economic contract characteristics to improve water management in dryland agriculture and reduce farm household vulnerability to increasing periods of droughts. Water harvesting has been identified by the FAO (2014) as a critical element in strengthening land users’ capacity to adapt to climate change, requiring substantial more action in Sub-Saharan Africa than currently is the case to improve water management in both rainfed and irrigated agricultural systems and reduce exposure to dry spells. The practical relevance of the results presented here is that they help regional policymakers identify how water harvesting can be rolled out as a farm household climate risk adaptation and mitigation strategy, accounting for the literacy rate of local farmers
in this developing country context, and support effective uptake of contractual agreements through adequate communication.

Sixty percent of the sampled farmers in this study indicated to face severe droughts and water shortages. Previous attempts to build heavily subsidized community water ponds in the area largely failed. Hence this attempt to assess if framing water security management in the study area as a private investment decision, facilitated through a contractual agreement, is of interest to the farmers involved. The results in this study clearly show that demand for the offered contractual agreements is high and farmers are willing to pay market interest levels to obtain the financial support (micro credit) for such an investment decision.

Making sure farmers understand the specific contractual terms and conditions they commit to was a challenge, hence the reason for the use of visual aids in a split sample approach. This was expected to also accommodate the fact that a considerable share is illiterate. Both samples indeed generated significantly different results, highlighting the importance of how information is conveyed to survey participants in a developing country context. Although choice consistency was very high in both sub-samples, there are indications that whilst controlling for differences in income and education levels the sample who received the version of the CE with the visualizations made more stable and less random choices.

The observed variation in the significance levels across the different contract characteristics indicate that some are considered more important than others. This pattern is confirmed when examining the self-reported importance attached to the various contract characteristics, with the water harvesting technology such as pond lining, cover and capacity being valued highest.
Here too we detect important differences between the two sub-samples and observe that the contract characteristics for which visual aids were developed are considered more attentively. In conclusion, the use of visual aids in CEs has a significant effect on choice behavior and ultimately on the economic values derived from the estimated choice models in this study carried out in a remote developing country context. The empirical evidence of such effects in the existing literature is very limited and deserves more attention in future stated preferences research.
Annex I: Estimated random parameters logit choice models for the samples with literate and illiterate survey participants

<table>
<thead>
<tr>
<th>Variables</th>
<th>Illiterate survey participants</th>
<th>Literate survey participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter estimate</td>
<td>St. error</td>
</tr>
<tr>
<td>ASC</td>
<td>3.687</td>
<td>2.484</td>
</tr>
<tr>
<td><strong>Choice attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contract provider = regional government</td>
<td>0.436**</td>
<td>0.184</td>
</tr>
<tr>
<td>Contract duration (years)</td>
<td>-0.006</td>
<td>0.016</td>
</tr>
<tr>
<td>Pond capacity (m3)</td>
<td>0.007***</td>
<td>0.002</td>
</tr>
<tr>
<td>Pond lining = cement</td>
<td>-0.167</td>
<td>0.163</td>
</tr>
<tr>
<td>Top of pond = covered</td>
<td>0.592***</td>
<td>0.163</td>
</tr>
<tr>
<td>Non-monetary payment (labour days)</td>
<td>0.004</td>
<td>0.008</td>
</tr>
<tr>
<td>Monthly payment (Birr)</td>
<td>-0.002***</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC x 1 if version includes pictograms</td>
<td>1.756**</td>
<td>0.776</td>
</tr>
<tr>
<td><strong>Model summary statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-517.306</td>
<td></td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.310</td>
<td></td>
</tr>
<tr>
<td>Chi square statistic (17 d.o.f.)</td>
<td>463.895***</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>682</td>
<td></td>
</tr>
</tbody>
</table>

Significance levels: *p<0.10; **p<0.05; ***p<0.01
Annex II: Estimated random parameters logit choice models accounting for stated attribute non-attendance

<table>
<thead>
<tr>
<th>Variables</th>
<th>Without attribute visualizations</th>
<th>With attribute visualizations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient estimate</td>
<td>St. error</td>
</tr>
<tr>
<td>ASC</td>
<td>3.433***</td>
<td>0.859</td>
</tr>
<tr>
<td><strong>Choice attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contract provider = regional government</td>
<td>0.192</td>
<td>0.192</td>
</tr>
<tr>
<td>Contract duration (years)</td>
<td>-0.018</td>
<td>0.017</td>
</tr>
<tr>
<td>Pond capacity (m³)</td>
<td>0.013***</td>
<td>0.003</td>
</tr>
<tr>
<td>Pond lining = cement</td>
<td>0.057</td>
<td>0.207</td>
</tr>
<tr>
<td>Top of pond = covered</td>
<td>0.670***</td>
<td>0.172</td>
</tr>
<tr>
<td>Non-monetary payment (labour days)</td>
<td>-0.017</td>
<td>0.012</td>
</tr>
<tr>
<td>Monthly payment (Birr)</td>
<td>-0.003***</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC x 1 if respondent is illiterate</td>
<td>-1.221</td>
<td>0.981</td>
</tr>
<tr>
<td>Payment x 1 if income is higher than average</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Model summary statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-637.935</td>
<td></td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.296</td>
<td></td>
</tr>
<tr>
<td>Chi square statistic (18 d.o.f.)</td>
<td>536.840***</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>825</td>
<td></td>
</tr>
</tbody>
</table>

Significance levels: *p<0.10; **p<0.05; ***p<0.01
5 Testing hypothetical bias in an agricultural field experiment

5.1 Introduction

Non-market valuation has developed into two main branches of research using revealed preference (RP) and stated preference (SP) methods. The former infers values for products or services by studying actual behavior of consumers in a closely related or surrogate market for the environmental resource in question, while the latter is able to measure economic values when there is no behavioral trail on which to rely, in particular for products or services which are either not directly related to any usage and traded in existing markets, or which are new and not yet marketed (Louviere et al., 2003). Both businesses and governments need reliable methods and models to estimate consumer demand and willingness to pay (WTP) associated with investment decisions in new product or policies.

SP methods elicit individuals’ preferences and WTP in a hypothetical setting. This has been one of their major criticisms, i.e. that what people say they would pay is often not the same as what they actually pay in a real market setting (e.g. Carson et al., 1996). People may not know what they would do if a hypothetical situation were real, or they may not be willing to reveal what they would do and behave strategically instead. This discrepancy between people’s behavioral intent as expressed in a SP study and their actual behavior is called hypothetical

---

8 This chapter has been accepted for presentation at the 6th World Congress of Environmental and Resource Economists (WCERE) in Gothenburg, Sweden, June 25 to 29, 2018.
bias, and has been subject to external validity testing of various SP methods, so far most importantly using contingent valuation approaches (e.g. Harrison and Rutström, 2008).

The number of studies investigating hypothetical bias by comparing SP and RP data is limited. The empirical evidence from these studies assessing hypothetical bias is furthermore mixed across a wide field of application. Most meta-analyses of the existing environmental valuation literature find an overstatement of hypothetical WTP over real payments. This includes List and Gallet (2001) who find that individuals overstate their preferences on average by about a factor 3 in a hypothetical setting based on 29 contingent valuation studies and Murphy et al. (2005) who find a lower ratio of hypothetical to actual values of 1.35 based on another 28 contingent valuation studies. Loomis (2011) reviews two meta-analyses and concludes that hypothetical WTP exceeds the actual value by a factor of two to three. Individual studies seem less conclusive. For instance, Carson et al. (1996) concluded that, in general, contingent valuation estimates were very similar and somewhat smaller than RP estimates, while Carlsson and Martinsson (2001) find no significant difference in WTP between a hypothetical and actual choice experiment. In a hypothetical and real referendum, Vossler and Kerkvliet (2003) find that hypothetical survey responses match actual voting outcomes and WTP estimates.

The comparison of SP and RP data has seen a small surge more recently in transportation studies, applying choice experiments focusing on travel time. Here, SP elicited through choice experiments seem to generally understate real values (e.g. Brownstone and Small, 2005). For example, Hensher (2010) finds that real values tend to be larger than the values derived from hypothetical markets, i.e. people are willing to pay more to save time than predicted in hypothetical markets. In their literature review of more recently applied choice experiments,
Beck et al. (2016) conclude that there is no clear a priori expectation in which direction hypothetical bias will manifest itself, if at all.

A number of possible ways have been proposed and tested to reduce or eliminate hypothetical bias and increase the incentive compatibility of SP, especially in contingent valuation studies (e.g. Carson and Groves, 2007; Vossler and Evans, 2009; Loomis, 2014). These recommendations have been tested in the field of choice experiments (e.g. Hensher, 2010; Fifer et al., 2014; Beck et al., 2016). A reduction in hypothetical bias is expected particularly in the context of private goods (List and Gallet, 2001). Numerous studies exist in the food safety domain, investigating public WTP for food labelling (e.g. McCluskey and Loureiro, 2003), in particular focusing on locally and organically grown food products (e.g. Scarpa and Del Giudice, 2004; Carpio and Isengildina-Massa, 2009), or genetically modified food products (e.g. Loureiro and Bugbee, 2005; Kimenjua and De Groote, 2008). Several of these studies apply alternative preference elicitation procedures to eliminate hypothetical bias such as experimental auctions (e.g. Soler et al., 2002). However, hardly any studies exist where hypothetical and real food purchases are compared directly. Exceptions are Volinskiy et al. (2011), Chowdhury et al. (2011) and Moser et al. (2014), who all find that estimated hypothetical WTP values are higher than the real ones.

In the experiment by Volinskiy et al. (2011) some participants were hypothetically and others actually given a 1 liter bottle of canola oil and the opportunity to acquire a different type of canola oil in a computer-based choice experiment. They find that there is hypothetical bias, but there also exist a group of participants who have the same choice behavior regardless of whether choices are actual or hypothetical. The other two studies also report hypothetical bias
when comparing real and hypothetical food purchases, i.e. orange sweet potato in Uganda (Chowdhury et al., 2011) and apples in Italy (Moser et al., 2014). Both studies furthermore employ cheap talk scripts to test their impact on hypothetical bias and find that it reduces the bias, but not entirely. While participants in the choice experiment in Chowdhury et al. (2011) were given a participation fee of about 30 US dollar cents that they were able to use during the choice experiment, Moser et al. (2014) asked participants in a supermarket to use their own money to make the payment.

The three studies above all examine hypothetical bias based on inter-respondent comparisons of choice data. Inter-respondent comparisons are considered a weaker convergent validity test in the parallel test-retest valuation literature to assess the temporal stability of SP data (e.g. Schaafsma et al., 2014; Brouwer et al., 2017) because of the lack of full control over relevant socio-demographic, financial and purchase experience characteristics of respondents in split samples. In this study, we therefore test hypothetical bias through both inter-respondent and intra-respondent comparisons of choice data related to food purchases in an agricultural market in Ethiopia. A random sample of market visitors is first asked in a hypothetical discrete choice experiment to choose between conventional and organic red haricot beans given different price levels, and then endowed with a lump sum of money to actually purchase the same conventional and organic red haricot beans using exactly the same experimental design (allowing for intra-respondent comparison). A control group of market visitors is asked to make exactly the same choices in a hypothetical choice experiment (allowing for inter-respondent comparison). Differences in choices, preference parameters and WTP are tested. In addition, factors explaining hypothetical bias are examined.
The remainder of this chapter is organized as follows. The next section first describe the design of the choice experiment, the econometric model and the data collection procedure. This is followed in Section 3 by a presentation of the main results, including sample characteristics and a comparison of choice behavior, choice models and WTP values based on the hypothetical and real purchases. Finally, Section 4 concludes.

5.2 Experimental design and data collection

5.2.1 The field experiment

Consumer preferences and WTP for conventional and organically grown red haricot beans, a local household staple food, are elicited in this study. The products look the same, but organically grown beans are generally free of pesticides and artificial manure. They may therefore be more damaged due to insects and diseases or essential nutrients may be in shorter supply, affecting their growth. Beans are labelled in the choice experiment as either conventionally or organically grown to see if agricultural practices matter in consumers’ decisions and guide agricultural policy regarding this crop. The beans’ characteristics are presented in Table 5.1. These characteristics were identified in consultation with experts (crop breeders and researchers), and informed by a literature review and a focus group discussion with a random selection of beans buyers and sellers before the main household survey.

The second attribute is whether the haricot beans originate from the local region (i.e. whether they are endemic) or are newly introduced by a government research center. Higher yield and disease resistant varieties are developed in government research centers, and actively introduced in rural agricultural development programs throughout Ethiopia (Bechere, 2007).
These new varieties are already sold on existing rural markets, including the market where this study was conducted.

Table 5.1: Experimental design

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Definition</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural practice</td>
<td>Whether the haricot beans are grown conventionally or organically</td>
<td>- Conventional</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Organic</td>
</tr>
<tr>
<td>Bean variety</td>
<td>Whether the haricot beans are endemic or newly introduced by a government research center</td>
<td>- Endemic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Introduced</td>
</tr>
<tr>
<td>Cropping pattern</td>
<td>Whether or not haricot beans are cultivated jointly with other crops</td>
<td>- Monoculture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Mixed</td>
</tr>
<tr>
<td>Price per kg</td>
<td>Price paid for a kilogram of haricot beans in the local market</td>
<td>Birr 5, 10, 15, 20</td>
</tr>
</tbody>
</table>

In view of the fact that buyers on rural agricultural markets in Ethiopia are often farm households themselves (e.g. Aragie and McDonald, 2014), their general level of understanding of agricultural practices is high. This also applies to the third attribute, i.e. whether the red haricot beans have been cultivated jointly with other crops, such as maize in a mixed farming system, or as part of a cash crop monoculture. The last attribute is the price of the haricot beans, which varies between 5 and 20 Birr per kilogram, based on existing market price levels. In August 2012 when the field experiment was conducted, 1 Birr equaled on average US$ 0.0558. Hence 70 Birr is equal to US$ 3.91.
Alternative product specifications were created by combining these four attributes based on their attribute levels. Based on a full factorial design, 16 choice sets showing 32 product specifications were generated to enable the estimation of main effects. Showing consumers all 16 choice sets was considered too much of a cognitive burden, possibly introducing biases in choice behavior due to fatigue. Therefore, the 16 choice tasks were randomly grouped into four blocks of 4 choice tasks. Consumers were randomly assigned two of these blocks and hence answered 8 choice tasks in total.

Each choice set shows two product specifications along with the option to choose none of the two. Inclusion of this latter ‘status quo’ alternative is instrumental in not forcing consumers to buy one of the two bean varieties, and is expected to help minimize hypothetical bias (Hensher, 2010). An example of a choice set is shown in Figure 5.1. In order to make the experiment look as real as possible and further minimize hypothetical bias, consumers were also shown the 1 kilogram bags of haricot beans, which were on display on a table in the market place, similar to Yue and Tong (2009). The latter applied a hypothetical choice experiment in the US focusing on locally and organically grown tomatoes and showed participants real tomatoes instead of using for example pictures. The information on the choice cards was read out loud to respondents in view of the fact that a considerable number of respondents cannot read or write.

In the choice experiment, the experimental group of consumers was first randomly assigned one block of 4 choice tasks and asked to choose their most preferred bean variety hypothetically. This experimental group of consumers was subsequently given 70 Birr for their participation in the second part of the choice experiment, in which they were told they could use the money endowment for the actual purchase of different varieties of haricot beans, or not
use it and keep the money (opt-out). They were asked to make the same choices again, without telling them, showing them in a second sequence of choice tasks 4 times two pairs of 1 kilogram bags of haricot beans with the same characteristics as in the previous 4 choice tasks, described on the same choice cards. This time they were asked to actually purchase each time one of the 1 kilogram bags they were shown. This allowed us to compare their hypothetical choices directly with their real choices in the intra-respondent comparison (see Figure 5.2).

Figure 5.1: Example choice card

<table>
<thead>
<tr>
<th>Variety</th>
<th>Conventional</th>
<th>Organic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Local</td>
</tr>
<tr>
<td>Cropping</td>
<td>Mono</td>
<td>Mixed</td>
</tr>
<tr>
<td>Birr/kg</td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

Which haricot beans do you prefer? ○ ○ ○ ○

In the end, each consumer who participated in the choice experiment went home with 4 kilograms of haricot beans, which they purchased based on the lump sum endowment. Any remaining money they were allowed to keep. Based on prior estimation of a local household’s average monthly beans consumption (40-50 kg/household), this is approximately equal to a household’s need for 2-3 days.
A second control group of consumers was presented two randomly assigned blocks of 4 choice tasks each and asked to hypothetically choose 8 times their most preferred beans variety at the given prices. The results from this control group will be used in the inter-respondent comparison (Figure 5.2).

**Figure 5.2: Experiment and control group comparisons**

![Diagram showing experimental and control group comparisons]

### 5.2.2 Econometric model

As for example in Carlsson and Martinsson (2001), hypothetical bias is tested by comparing (i) hypothetical and real choices, (ii) preference parameters of the estimated choice models related to the hypothetical and real choices, and (iii) hypothetical and real WTP. Choice experiments fall in the class of attribute-based methods in which the deterministic part of an individual $i$'s utility attached to product $j$ in choice task $t$ is described in (1) as a linear function...
of its characteristics (attributes) $X_{ijt}$, other explanatory variables $Z_{ijt}$ and a constant that is specific to an alternative $k_j$, $\forall j \neq 1$ alternatives (e.g. McFadden, 1974):

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} = k_{j-1} + \beta X_{ijt} + \alpha Z_{ijt} + \varepsilon_{ijt} \quad \forall j \in D_{it}$$

(1)

where the strength of preferences related to the specific product characteristics is captured by the coefficient vector $\beta$ and $\alpha$ is the vector of coefficients associated with $Z$, and $\varepsilon_{ijt}$ is the stochastic part of the utility function. In this study, a consumer $i$ is presented in each choice task $t$ a limited set of alternative bean products $D_{it}$, each consisting of different characteristics, either in a hypothetical or real market setting. Although the alternative specific constant (ASC) $k_{j-1}$ typically captures the effect of factors that are not included in the model on utility, here it captures the marginal utility of the label attached to haricot beans, i.e. whether the beans are produced in a conventional or organic way. The stochastic term $\varepsilon_{ijt}$ is assumed to follow an IID extreme value distribution of type I (Train, 2003).

In order to account for preference or taste heterogeneity across individual consumers, a mixed logit choice model is estimated, more specifically a random parameters logit (RPL) model. The RPL model overcomes the main limitations of standard multinomial logit models by accounting for random taste variation, unrestricted substitution patterns and unobserved correlation between alternatives (Train, 2003). RPL models can represent heterogeneity across individuals in both observed and unobserved preferences, but require computationally intensive procedures to estimate probabilities (Newman, 2003). RPL models were shown to
provide a better statistical fit in Volinskiy et al. (2011) and Moser et al. (2014), and are therefore also used here.

In the RPL model, preference parameters are allowed to vary across individuals, applying different mixing distributions. Equation (2) describes the mixed logit (ML) probability of individual $i$ selecting alternative $j$ in choice task $t$ over other choice alternatives $k$. The utility coefficients $\beta$ vary across individuals, hence $\beta_i$, with density $\Delta(\beta_i | \theta)$. This density can be a function of any set of parameters and represents in this case the mean and covariance of $\beta$ in the sample population.

$$P_{ijt} = \int \left[ \frac{\exp[(\beta_i X_{ijt} + \alpha Z_{ijt})]}{\sum_{j \in D_{it}} \exp[(\beta_i X_{ijk} + \alpha Z_{ijk})]} \right] \Delta(\beta_i | \theta) d\beta_i \quad \forall j \in D_{it}$$

Treating preference parameters as random variables requires estimation through simulated maximum likelihood. Procedurally, the maximum likelihood algorithm searches for a solution by simulating draws from distributions with given means and standard deviations. Probabilities are calculated by integrating the joint simulated distribution. For efficiency purposes and ensure model stability, the models are estimated using a Halton sequence of 500 replications in a quasi-Monte Carlo maximum likelihood simulation (Bhat, 2001). Even if unobserved heterogeneity is accounted for in a ML-model, the model may still fail to explain the sources of this heterogeneity (Hynes et al., 2008). To this end and in order to better capture taste heterogeneity (e.g. Scarpa and Del Giudice, 2004), interactions of respondent specific
household characteristics with choice specific attributes will be included in the utility function to further improve the model fit.

5.2.3 Data collection

The design of the choice experiment was first pretested and subsequently implemented in August 2012 through 211 in-person interviews (110 in the experimental group and 101 in the control group) in the agricultural market in the adjacent villages Tula and Dorebafana in the South Nations Nationalities and Peoples Regional state of Ethiopia, during market days (Tuesday and Friday). The response rate was 100 percent, which is not unusual for this kind of social survey research in a developing country context (Whittington, 1998). Besides the choice experiment, a questionnaire was used to collect consumer data on haricot beans consumption and consumers’ socio-economic and demographic background characteristics. Market visitors were recruited randomly on a next to pass basis in the market. Conducting the experiment in the market place was expected to create a sense of being in a real market and help minimize hypothetical bias.

The mentioned villages primarily produce and consume haricot beans. No official statistics are available about the number of residents in the two villages or the number of traders in the market. Based on our own counts of the number of people, between 500 and 1,000 sellers and buyers visit the market on the two days per week. Market access is generally good in the region since most counties in the state are connected to the main asphalt road between Addis Ababa, the capital city of Ethiopia, and Moyale, a market town on the Ethiopia-Kenya border. The area is located at 7°3’11”N latitude and 38°29’43”E longitude, approximately 250 km south of Addis Ababa. The main crops produced in the area are maize, enset (also called false banana), khat,
haricot beans, fruits and vegetables. Haricot beans are cultivated in the area with a mean annual
temperature between 17.5\(^{0}\) C and 27.5\(^{0}\) C, and an elevation between 500 and 2,200 meters
above sea level (JICA, 2012). Although monoculture also takes place, haricot beans are
typically part of a mixed subsistence farming system with maize and other perennial crops.
Farmers sell surplus (fresh or dried) haricot beans on the local market in the open air. Some
also sell their products to local traders who subsequently sell the beans in the bigger regional
market in the nearest city of Hawassa (Woysa, 2014).

5.3 Results

5.3.1 Descriptive summary statistics

Table 5.2 reports some of the main background characteristics of the two samples. No major
differences are encountered between the two samples in terms of their socio-demographic
characteristics. Most participants were female, aged on average 31 years, coming from
households with on average 6 household members and a monthly income level of 1,010 Birr.
Half of the sample population earns no more than 700 Birr per month. This is less than US$ 40
per household per month, and hence below the global poverty threshold of US$ 1.25 per person
per day. Almost 40 percent of the sample is unable to read or write. This is close to what has
been reported, for example, by UNICEF (2014) over the period 2008-2012 for the country as
a whole, that on average 37 percent of the male and 53 percent of the female population cannot
read or write. Some minor differences exist between the two samples in terms of the shares
who finished primary and secondary school. Most households are involved in subsistence
farming. The median tropical livestock unit equals 2.1, while the mean in both samples is
somewhat higher, namely 2.5. Most farm households own chickens and 1 or 2 cows, followed
by a few goats or sheep.
Table 5.2: Summary statistics consumer characteristics

<table>
<thead>
<tr>
<th>Consumer characteristic</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St. dev.</td>
</tr>
<tr>
<td>Share of female participants (%)</td>
<td>65.0</td>
<td></td>
</tr>
<tr>
<td>Average age (years)</td>
<td>33.0</td>
<td>8.7</td>
</tr>
<tr>
<td>Average household size (persons)</td>
<td>7.0</td>
<td>3.4</td>
</tr>
<tr>
<td>Net household income (Birr/month)</td>
<td>1,051.6</td>
<td>1,088.9</td>
</tr>
<tr>
<td>Tropical livestock units</td>
<td>2.6</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Education level

| Share who cannot read and write (%)         | 35.8   |          | 38.6   |          |
| Share who attended primary school (%)       | 57.8   |          | 46.5   |          |
| Share who attended secondary school (%)     | 5.5    |          | 14.9   |          |

Haricot beans consumption

| Average monthly beans consumption (kg)       | 34.2   | 15.4     | 56.8   | 12.2     |
| Share consuming local beans only (%)        | 14.5   |          | 50.5   |          |
| Share consuming imported beans only (%)     | 38.2   |          | 28.7   |          |
| Share consuming both types of beans (%)     | 47.3   |          | 20.8   |          |
| Range lowest prices paid in last 12 months (Birr/kg) | 1-8   |            | 1-12   |          |
| Range highest prices paid in last 12 months (Birr/kg) | 8-15 |          | 14-17  |          |
| Average midpoint price paid in last 12 months (Birr/kg) | 9.9  | 1.6      | 12.3   | 0.5      |
| Sample size (participants)                  | 110    |          | 101    |          |

More prominent are the differences observed in the summary statistics related to haricot beans consumption in the two samples, especially average bean consumption and the types of beans consumed. Half of the control group eats local beans only, while this share is 15 percent in the experimental group. Compared to the control group, more than twice as many participants in
the experimental group eat both local and imported beans. Although the self-reported lowest and highest price ranges overlap, there is a significant difference between the midpoint price of this range paid by the experimental and control group over the last 12 months. These differences will hence have to be accounted for in the inter-respondent comparison.

5.3.2 Comparing hypothetical and real purchases

Participants in the experimental sample first chose their preferred bean variety hypothetically in the choice experiment. They were subsequently endowed with a lump sum of money, enabling them to actually buy their preferred beans in the market place. Comparing the identical 4 hypothetical and 4 real choices, the latter deviate from their initial hypothetical choices in 15.9 percent of the choice occasions. Hence, the degree of hypothetical bias based on the intra-respondent comparison seems limited.

In a majority of the cases (67%) where consumers chose another alternative in the real experiment, this new alternative was more expensive than the one they chose originally in the hypothetical choice experiment. On average, these consumers paid 9.3 Birr per kg more compared to what they said they would pay initially during the hypothetical choice experiment. This suggests that the hypothetical bias examined in this study, defined as the difference between stated and revealed preferences, is negative, i.e. stated preferences underestimate what consumers are willing to pay in reality. This finding will be confirmed later in section 5.3.3 when comparing the mean WTP values derived from the estimated choice models.

On average, consumers spent 40.2 Birr on their preferred beans variety in the control group (last 4 choice tasks). This was slightly higher in the experimental group. On average consumers
in the experimental group spent 42.6 Birr in the first 4 choice tasks (hypothetically), while they spent on average 44.3 Birr during the real experiment after they were given a lump sum of money to actually purchase the beans. Only the difference in aggregate expenditures between the last 4 choices in the experimental and control group (inter-respondent comparison) is statistically significant, not the difference between the first and last 4 choices in the experimental group (intra-respondent comparison). The standardized Mann-Whitney test statistic is 1.002 ($p=0.317$) when comparing aggregate expenditures between the first and last 4 choice tasks in the experimental group (intra-respondent comparison) and 2.456 ($p=0.014$) across the last 4 choices in the experimental and control group (inter-respondent comparison). The distribution of the aggregate expenditures for the preferred (chosen) bean varieties across the 4 choice tasks in the three sub-samples is presented in Figure 5.3. The presented distributions are statistically the same: the non-parametric Kolmogorov-Smirnov test is unable to reject the null hypothesis of equal distributions at the 10 percent significance level. The Kolmogorov-Smirnov test statistic is 1.215 ($p=0.104$) when comparing aggregate expenditures across the last 4 choices between the experimental and control group (inter-respondent comparison) and 0.472 ($p=0.979$) when comparing aggregate expenditures between the first and last 4 choice tasks in the experimental group (intra-respondent comparison).
Consumers spent a minimum of 25 and a maximum of 70 Birr during the 4 choices in the experimental and control group. Most participants in the hypothetical choice experiments spent 25 Birr (23% of the participants in the experimental group and 28% of the participants in the control group). In the real experiment, most participants spent either 25 or 50 Birr (18 and 19% respectively). The maximum amount spent in the hypothetical choice experiment in the experimental group was 65 Birr. After endowing these participants with 70 Birr, they spent slightly (and insignificantly) more on their most preferred beans. From 45 Birr onwards, they spent either the same or slightly more than what they spent during the first 4 hypothetical choice tasks (see Figure 5.3). However, as the test results discussed before show, the endowment did not significantly influence their purchase behavior.
5.3.3. What explains hypothetical bias?

Whether or not participants’ choices were consistent in the hypothetical and real experiment was regressed in Table 5.3 on their sociodemographic characteristics, the characteristics of their beans consumption, and the design characteristics of the field experiments in a random effects probit model, accounting for the panel structure of the data (i.e. 4 choices per respondent). Hypothetical bias is measured in Table 5.3 as an inconsistency between the hypothetical and real choices. Correlation (heteroscedasticity) between choice tasks is captured by the error term $\sigma_u$ and the correlation coefficient $\rho$ in Table 5.3.

The following variables were found to exert a significant effect on the observed hypothetical bias: age (older participants are more likely to change their choice in the real experiment), literacy rate (illiterate consumers appear to be less inclined to change their choices in the real experiment), which type of beans consumers currently buy (consumers who only buy locally produced beans are less likely to change their choice), and the choice task sequence (choices are more likely to be changed in the second and third choice task of the real experiment compared to the same choices in the hypothetical experiment whilst accounting for possible correlation between choices in the model’s panel data structure). No significant effect of any of the other experimental design characteristics on choice consistency between the hypothetical and real choice experiment can be detected, such as the levels of the different attributes. Also no significant effect can be found for consumers’ gender, income or wealth (measured through the tropical livestock units they own), or the average amount of beans participants eat every month or the price they currently pay.
Table 5.3. Factors explaining hypothetical bias

<table>
<thead>
<tr>
<th></th>
<th>Coeff. estimate</th>
<th>St. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.734***</td>
<td>0.935</td>
</tr>
</tbody>
</table>

**Sociodemographic characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Coeff. estimate</th>
<th>St. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (1=female)</td>
<td>0.297</td>
<td>0.269</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.042***</td>
<td>0.016</td>
</tr>
<tr>
<td>Literacy (1=illiterate)</td>
<td>-1.010***</td>
<td>0.323</td>
</tr>
<tr>
<td>Income (100 Birr/month)</td>
<td>-0.013</td>
<td>0.015</td>
</tr>
<tr>
<td>Tropical livestock units</td>
<td>-0.027</td>
<td>0.048</td>
</tr>
</tbody>
</table>

**Beans characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Coeff. estimate</th>
<th>St. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beans consumption (kg/month)</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>Price paid for beans (Birr/kg)</td>
<td>-0.039</td>
<td>0.086</td>
</tr>
<tr>
<td>Beans variety (1=buys local beans only)</td>
<td>-1.044**</td>
<td>0.467</td>
</tr>
</tbody>
</table>

**Design characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Coeff. estimate</th>
<th>St. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice task 6</td>
<td>0.504*</td>
<td>0.271</td>
</tr>
<tr>
<td>Choice task 7</td>
<td>0.704**</td>
<td>0.275</td>
</tr>
<tr>
<td>Choice task 8</td>
<td>0.293</td>
<td>0.280</td>
</tr>
<tr>
<td>Label (1=organic)</td>
<td>0.245</td>
<td>0.194</td>
</tr>
<tr>
<td>Variety (1=new)</td>
<td>0.108</td>
<td>0.193</td>
</tr>
<tr>
<td>Cultivation pattern (1=mixed)</td>
<td>0.153</td>
<td>0.204</td>
</tr>
<tr>
<td>Price level (Birr/kg)</td>
<td>0.009</td>
<td>0.018</td>
</tr>
</tbody>
</table>

Log-likelihood: -164.785

Wald chi-squared: 24.81, 0.053

$\sigma_u$: 0.723, 0.179

$P$: 0.343, 0.112

Number of observations: 436

Sample size (participants): 109

Note: * $p<0.10$; ** $p<0.05$; *** $p<0.01$
5.3.4 Estimated choice models and equality of preference parameters

Four different choice models were estimated based on the first and last four choices in the experimental and control group (Table 5.4). The estimated mixed logit models are all highly significant. The estimated models for the control group have a slightly higher explanatory power when examining McFadden’s pseudo R-square. The R-square values are high for all 4 samples (e.g. Hoyos, 2010). All the choice attributes are, as expected, statistically significant, and price has, as expected, a negative sign. No prior expectations exist with regards to the direction of influence of the other choice attributes. Consumers prefer conventional beans to organic ones in the experimental group, while no such difference can be detected in the control group where the labels do not differ significantly from each other and have no significant influence on choice behaviour. However, significant preference heterogeneity surrounds the estimated coefficients for the labels and price, as can be seen in the lower part of Table 5.4, especially in the experimental group, implying that consumers differ in the utility they assign to these two characteristics of the beans. No significant preference heterogeneity can be found for the other choice attributes. Newly introduced bean varieties are preferred to local beans and consumers prefer mixed cultivation patterns to monoculture.
Table 5.4: Estimated mixed logit choice models

<table>
<thead>
<tr>
<th>Choice attributes</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First 4 choice tasks</td>
<td>Last 4 choice tasks</td>
</tr>
<tr>
<td></td>
<td>(hypothetical)</td>
<td>(real)</td>
</tr>
<tr>
<td>Label (1=conventional)</td>
<td>0.824*** 0.208</td>
<td>0.680*** 0.189</td>
</tr>
<tr>
<td>Variety (1=newly introduced)</td>
<td>0.927*** 0.144</td>
<td>0.890*** 0.150</td>
</tr>
<tr>
<td>Cultivation pattern (1=mixed)</td>
<td>1.426*** 0.167</td>
<td>1.117*** 0.151</td>
</tr>
<tr>
<td>Price (Birr/kg)</td>
<td>-0.118*** 0.022</td>
<td>-0.062*** 0.020</td>
</tr>
<tr>
<td>Co-variates (interaction terms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety x quantity (1&gt; average)</td>
<td>-0.358 0.494</td>
<td>-0.241 0.461</td>
</tr>
<tr>
<td>Variety x variety (1=local beans)</td>
<td>-0.095 0.341</td>
<td>-0.158 0.327</td>
</tr>
<tr>
<td>Price x income (1&gt; average)</td>
<td>0.072** 0.036</td>
<td>0.052 0.036</td>
</tr>
<tr>
<td>Distribution choice parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Label (1=conventional)</td>
<td>1.448*** 0.284</td>
<td>1.292*** 0.279</td>
</tr>
<tr>
<td>Variety (1=newly introduced)</td>
<td>0.015 0.367</td>
<td>0.051 0.485</td>
</tr>
<tr>
<td>Cultivation pattern (1=mixed)</td>
<td>0.069 0.640</td>
<td>0.281 0.818</td>
</tr>
<tr>
<td>Price (Birr/kg)</td>
<td>0.093*** 0.025</td>
<td>0.104*** 0.028</td>
</tr>
<tr>
<td>Model summary statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-280.935 -308.805</td>
<td>-249.896 -246.295</td>
</tr>
<tr>
<td>McFadden R²</td>
<td>0.419 0.361</td>
<td>0.437 0.445</td>
</tr>
<tr>
<td>Number of observations</td>
<td>440 440</td>
<td>404 404</td>
</tr>
</tbody>
</table>

Note: * p<0.10; ** p<0.05; *** p<0.01
Like the unobserved preference heterogeneity, also the influence of observed preference heterogeneity is limited. Control was included for several potentially influencing socio-demographic (e.g. age, gender, household size, education level) and other factors (e.g. wealth, livestock), but none played a consistent significant role in all four models. Given the differences between the experimental and control group observed in Table 5.2, control was included for the amount of beans households consume and the variety of beans. However, these covariates played no role in the estimated choice models. Only the interaction between the price of beans and a household’s income appeared to have a significant effect in the first part of the choice sequence in the experimental group, where consumers earning more than the sample average appear to be less sensitive to the proposed price levels.

Given the very limited impact of other explanatory factors, the mixed logit models containing the choice attributes only were compared across the four sub-samples applying the Swait and Louviere (1993) test procedure. The results are presented in the upper half of Table 5.5. The Likelihood Ratio test results show that there exist no significant differences between the two intra-respondent choice models, i.e. based on the choice data related to the first and last 4 choice tasks in the experimental group, not in terms of preference parameters or scale parameters. The latter implies that the error variance, which is inversely related to the scale parameter (Louviere, 2003), and hence the amount of noise surrounding choice behavior, for instance related to random choices, is the same between the two estimated models. On the contrary, the inter-respondent comparison yields significant differences, both when comparing the first 4 choice tasks between the experimental and control group and the last 4 choice tasks. The former suggests that the two groups were behaving differently from the start of the choice experiment.
Table 5.5: Test results equality of preference ($\beta$) and scale ($\lambda$) parameters and willingness to pay (WTP) between purchase choices in the experimental (E) and control (C) group

<table>
<thead>
<tr>
<th>Comparison of groups</th>
<th>LL(1)</th>
<th>LL(2)</th>
<th>LL(pooled)</th>
<th>LR-test (8 d.f.)</th>
<th>Reject?</th>
<th>LL(pooled)</th>
<th>LR-test (1 d.f.)</th>
<th>Reject?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) E - first 4 hypothetical choices</td>
<td>-283.267</td>
<td>-251.084</td>
<td>-552.788</td>
<td>36.590</td>
<td>Yes</td>
<td>-552.646</td>
<td>0.285</td>
<td>No</td>
</tr>
<tr>
<td>(2) C - first 4 hypothetical choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) E - first 4 hypothetical choices</td>
<td>-283.267</td>
<td>-310.101</td>
<td>-595.737</td>
<td>2.630</td>
<td>No</td>
<td>-594.683</td>
<td>2.108</td>
<td>No</td>
</tr>
<tr>
<td>(2) E - last 4 real choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) E - last 4 real choices</td>
<td>-310.101</td>
<td>-246.815</td>
<td>-567.749</td>
<td>17.367</td>
<td>Yes</td>
<td>-565.600</td>
<td>4.298</td>
<td>Yes</td>
</tr>
<tr>
<td>(2) C - last 4 hypothetical choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison of groups</th>
<th>WTP product specification (1)</th>
<th>WTP product specification (2)</th>
<th>WTP product specification (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) E - first 4 hypothetical choices</td>
<td>$p&lt;0.001$</td>
<td>$p&lt;0.001$</td>
<td>$p=0.283$</td>
</tr>
<tr>
<td>(2) C - first 4 hypothetical choices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) E - first 4 hypothetical choices</td>
<td>$p=0.082$</td>
<td>$p=0.106$</td>
<td>$p=0.105$</td>
</tr>
<tr>
<td>(2) E - last 4 real choices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) E - last 4 real choices</td>
<td>$p=0.002$</td>
<td>$p=0.001$</td>
<td>$p=0.076$</td>
</tr>
<tr>
<td>(2) C - last 4 hypothetical choices</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
1 Swait and Louviere (1993) test results.
2 Poe et al. (2005) test results.
LL: Log Likelihood; LR: Likelihood Ratio; (1): sample 1; (2): sample 2; d.f.: degrees of freedom.
5.3.5 Estimated willingness to pay based on stated and revealed preference data

Mean WTP values are estimated based on the choice models presented in Table 5.4 for three different product specifications (Table 5.6). In the estimation procedure the price coefficient is fixed in each subsample under the assumption of a constant utility of money income. Standard errors for the mean WTP values are calculated using the Krinsky and Robb (1986) bootstrapping procedure based on 10,000 draws. From both the intra-respondent and inter-respondent comparison in Table 5.6, a consistent negative hypothetical bias can be observed. In other words, mean WTP based on the stated choice data is always higher than mean WTP based on the revealed choice data. Across the three product specifications, the size of the negative hypothetical bias is fairly similar, varying between 28 and 34 percent (31% on average), when comparing the first 4 hypothetical choices with the last 4 real choices in the experimental group (intra-respondent comparison). The hypothetical bias varies more widely between 30 and 70 percent when comparing the last 4 hypothetical choices in the control group with the last 4 real choices in the experimental group in the inter-respondent comparison, and is 52 percent on average.
<table>
<thead>
<tr>
<th>Product specification</th>
<th><strong>Experimental group</strong></th>
<th></th>
<th><strong>Control group</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First 4 hypothetical choice tasks</td>
<td>Last 4 real choice tasks</td>
<td>First 4 hypothetical choice tasks</td>
<td>Last 4 hypothetical choice tasks</td>
</tr>
<tr>
<td>(1) Conventional</td>
<td>16.59 (2.69)</td>
<td>24.33 (7.43)</td>
<td>8.06 (1.15)</td>
<td>10.52 (2.28)</td>
</tr>
<tr>
<td>Mixed</td>
<td>6.55 – 15.18</td>
<td>5.57 – 27.21</td>
<td>0.69 – 4.98</td>
<td>1.12 – 8.85</td>
</tr>
<tr>
<td>(2) Conventional</td>
<td>10.87 (2.20)</td>
<td>16.39 (5.52)</td>
<td>2.84 (1.09)</td>
<td>4.99 (1.97)</td>
</tr>
<tr>
<td>Imported</td>
<td>6.55 – 15.18</td>
<td>5.57 – 27.21</td>
<td>0.69 – 4.98</td>
<td>1.12 – 8.85</td>
</tr>
<tr>
<td>Monoculture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) Organic</td>
<td>5.72 (0.85)</td>
<td>7.93 (2.34)</td>
<td>5.23 (0.52)</td>
<td>5.53 (0.90)</td>
</tr>
<tr>
<td>Local</td>
<td>4.05 – 7.38</td>
<td>3.36 – 12.51</td>
<td>4.22 – 6.24</td>
<td>3.77 – 7.30</td>
</tr>
<tr>
<td>Mixed</td>
<td>6.55 – 15.18</td>
<td>5.57 – 27.21</td>
<td>0.69 – 4.98</td>
<td>1.12 – 8.85</td>
</tr>
</tbody>
</table>

Table 5.6: Consumer willingness to pay (WTP) for different varieties of haricot beans (Birr/kg)
The differences between the WTP values are tested for their statistical significance using Poe et al.’s (2005) combinatorial test. The test results are presented in the bottom half of Table 5.5. The same test sequence is followed as for the Swait and Louviere test in the top half of Table 5.5. The differences between the estimated WTP values for the last 4 real choices and the first 4 hypothetical choices in the experimental group are not statistically significant at the 5 percent level (intra-respondent comparison). Hence, following the hypothetical choice experiment, when given the opportunity to actually purchase the same bags of beans, consumers’ WTP appears statistically speaking to be the same as their hypothetical WTP. The reverse result is found in the inter-respondent comparison. Here, the hypothetical bias is statistically significant at the 5 percent level for 2 out of the 3 product specifications.

Finally, 81 participants in the experimental group were identified who actually paid for the first product in Table 6 from their endowment of 70 Birr, 60 participants for the second product and 50 participants for the third product. The average amount of money these participants spent on the first product was 10.6 Birr per kg, while the average price participants paid for the second product was 11.1 Birr and 10.9 Birr for the third product. These values are not significantly different from each other, and perhaps more importantly, also not from the values of the same hypothetically chosen products in the experimental group based on the intra-respondent comparison. The Mann-Whitney test results for the intra-respondent comparison are 0.317 ($p=0.751$) for the first product specification, 0.145 ($p=0.885$) for the second product, and -0.090 ($p=0.928$) for the third product specification. The comparison between samples is not presented here because of the low number of observations in the control group for the specific bean products. In the hypothetical experiment, the specified beans were selected no more than 28 times by the control group. These values are considerably smaller than the estimated WTP.
values in Table 5.6 for the first 2 products (56% and 32% for the first and second product, respectively). Only in the case of the third product specification is the actual amount of money spent on a kilogram of beans higher (37%) than the estimated WTP value based on the model presented in Table 5.4 for the last 4 choice tasks in the experimental group.

5.4 Conclusions and discussion

The novelty of this study, aiming to assess the degree of hypothetical bias in a field experiment, is methodologically speaking found in the combination of the inter-respondent and intra-respondent comparison of hypothetical and real purchase behavior of visitors to an agricultural market in Ethiopia. To our knowledge, such a comparison of the two approaches has not been carried out yet in the existing literature. Existing studies use a ‘split sample’ approach. The approach followed here allows verification of the observed bias often found in the literature, in particular in the few discrete choice experiment studies focusing on food purchases.

This is also the first study to find evidence of a negative hypothetical bias in this particular field of application when comparing the collected hypothetical and real choice data, both in the inter-respondent and intra-respondent comparison. Most studies in environmental economics find a positive bias, where actual purchase behavior results in lower values than those associated with a hypothetical behavioral intent. Only the transportation literature reports a similar negative bias when valuing commuting time. Remarkably, the hypothetical bias is only statistically significant in the inter-respondent comparison, not in the intra-respondent comparison. This hence questions the general validity and reliability of the inter-respondent comparisons carried out so far in the literature to assess hypothetical bias.
Explanations for the observed hypothetical bias were examined by regressing differences in choice behavior in the intra-respondent comparison on socio-demographic and experimental design characteristics. Although the number of participants in the field experiment is similar to those used in the related literature reported in this paper (e.g. Chowdhury et al., 2011 or Moser et al., 2014), the number of significant explanatory factors is limited. The likelihood of hypothetical bias is significantly higher in the second and third choice task of the real experiment compared to the same choices in the hypothetical experiment, suggesting some degree of procedural bias influencing hypothetical bias. Additionally, age, literacy and current beans purchase behavior drive differences between the hypothetical and real choice data.

Possible explanations for the negative hypothetical bias found in this study are mainly speculative, as underlying drivers are not further investigated here. Possible procedural bias due to the endowment, as suggested by Moser et al. (2014), was dismissed as overall participant spending was only slightly more in the real experiment than in the hypothetical experiment. Also comparing actual purchase behavior and the real prices paid for different bean products with the prices of the same hypothetically chosen beans shows no significant difference in the intra-respondent comparison. The actual prices paid in the experiment are close to the midpoint estimate of the range of prices consumers indicated to have paid over the past 12 months for haricot beans.

Finally, noteworthy is that the real prices participants paid for the beans were considerably lower than the estimated WTP values based on the estimated choice models in the experimental group. This seems to have inflated the negative hypothetical bias. Using the actual prices paid in the real experiment instead of the estimated WTP values in the intra-respondent comparison
results in a different picture, where the negative hypothetical bias for the first product specification is reversed into a positive one (10.6 Birr paid in the real experiment compared to 16.6 Birr in the hypothetical experiment), there is hardly any difference anymore between hypothetical WTP and real expenditures for the second product specification (11.1 Birr paid in the real experiment compared to 10.9 Birr in the hypothetical experiment), and the negative hypothetical bias has more than doubled for the third product specification (10.9 Birr paid in the real experiment compared to 5.7 Birr in the hypothetical experiment). These latter results suggest a mixed outcome depending on the beans’ product specification. More research in this area seems warranted, to assess the robustness of the results presented here where existing split sample approaches were put to the test through intra-respondent comparisons.
6 Conclusions and recommendations

This PhD dissertation aimed to reduce information scarcity related to the broader non-market impacts of environmental policy in developing countries and their integration in mainstream policy and decision-making by using DCEs. The use and usefulness of DCEs to inform environmental policy and decision-making was tested focusing on key methodological issues related to the role of income, literacy and communication in public understanding and learning in DCEs where survey participants are typically asked to answer hypothetical questions for more or less familiar goods and services in a series of choice tasks. Section 6.1 summarizes the main methodological contribution of this dissertation and the advancement of DCEs from a developing country perspective, and presents recommendations for further research. Policy recommendations based on the case studies are covered in section 6.2.

6.1 Methodological advancements and future research

The case studies presented in this dissertation aimed to contribute in one way or the other to the methodological advancement of the application of DCEs in developing countries, where a considerable share of the population is poor and illiterate. DCEs consist of a series of hypothetical choice tasks that may impose a significant cognitive burden on survey participants in developed and developing economies. Given public unfamiliarity with DCE surveys and a relatively large illiteracy rate, thorough pretesting was crucial to ensure that a majority of the survey participants were able to understand the choice tasks in the DCE. All cases studies made use of extensive pre-testing procedures and focus group discussions with the target groups involved to support the framing and formulation of the choice tasks. Local enumerators who speak the local language were trained and formed an integral part of the pre-tests and
questionnaire design. DCE studies in developing countries like Ethiopia require careful consideration of the information load included in the survey and DCE. In order to ensure that the obtained results are valid and reliable on the one hand and representative on the other hand, i.e. including representative shares of the population who can and cannot read and write, the cognitive burden of the DCE to both literate and illiterate survey participants, and how this influences the obtained welfare estimates, has to be tested.

From the point of view of the analysis of the collected choice data, various state-of-the-art statistical methods were employed to investigate preference heterogeneity and possible correlation between alternatives (heteroscedasticity). All chapters (2-5) employed random parameters or mixed logit models, which account for unobserved preference heterogeneity. Observed heterogeneity across individual respondents was picked up where possible, through systematic search procedures and by including interaction terms of respondents’ socio-demographic characteristics and the choice attributes. In chapter 3, error components were furthermore added to the random parameters model to account for correlation between the hypothetical choice alternatives, while in chapter 4 also equality constrained latent class models were estimated to assess the impact of attribute attendance.

In order to answer the first research question of preference stability and possible preference learning during the sequence of choice tasks, the first two case studies explicitly examined the extent to which preference parameters and choice variance remained constant or changed during the DCE. To this end, the utility and scale parameters in the estimated choice models were estimated at the level of single or aggregated choice tasks following the Swait and Louviere (1993) test procedure to disentangle utility parameters from confounded scale
parameters. Contrary to one of the main assumptions underlying stated preferences research that respondents know their preferences and that these preferences are stable and coherent, the test results show that choice behavior at the start of the experiment is different compared to choice behavior at the end of the experiment, consistent with similar studies carried out in the developed world (e.g. Brouwer et al., 2010b). This suggests that important learning processes may occur when participants go through a sequence of choice tasks, which may make them especially sensitive to value cues implicitly or explicitly provided in these choice tasks, for example through price setting or the use of pictograms. It also raises the important question for future research to what extent aggregate utility functions can be used to reliably inform policy and decision-making.

Attribute attendance and the influence of visual aids was tested to address the second research question in this PhD dissertation. The way information is communicated in DCEs was expected to significantly influence stated choices, especially in samples with high illiteracy rates. This was therefore tested using a split sample approach where farmers were asked to choose their most preferred option from a menu of contractual agreements that include different combinations of institutional, economic and technological terms and conditions related to different water harvesting technologies. Making sure that farmers understood the specific contractual terms and conditions they committed to was a challenge, hence the reason why visual aids were developed and applied in one version of the DCE. Both samples indeed generated significantly different results, highlighting the importance of how information is conveyed to survey participants in a developing country context. Moreover, there are indications that the sample who received the version of the DCE with the visualizations made more stable and less random choices.
The variation in the significance levels across the different contract characteristics implies that some are considered more important than others. Although no formal tests were carried out to assess which of the estimated attribute (non-)attendance choice models (stated or inferred) best fitted the data, this pattern was confirmed when examining the self-reported importance attached to the various contract characteristics of water harvesting technology, and technical aspects such as pond lining, cover and capacity were valued highest. Interestingly, irrespective of the estimated choice model, the input of household labor never had a statistically significant effect on choice behavior. Since subsistence based in-kind trade instead of monetary transactions are not uncommon in a developing country context (e.g. Bennett and Birol, 2010), household labor was included in the offered contracts. Also because it is common practice to provide own labor in building community and private ponds in Ethiopia. It was shown that this contract condition plays no role in contract selection, also not when interacted with covariates such as, for example, household size as a proxy for the availability of household labor. This finding suggests that the in kind contribution is either not considered important or that the opportunity costs of labor are low in the study area. This latter explanation was not further verified during the survey, but deserves more in-depth investigation. The differences detected between literate and illiterate farmers indicate that future research has to carefully account for differences in educational background in the design of DCEs. The empirical evidence base of the effects of visual aids in the existing DCE literature is furthermore very limited and deserves more attention in future stated preferences research.

Hypothetical bias was tested in the third and final research question. A limited number of DCEs exist which use inter-personal comparisons. The novelty in the study in this PhD thesis is that hypothetical bias was tested using both the conventional split-sample inter-personal approach
applied in the literature, and an intra-personal approach, where the same sample of respondents was endowed to actually purchase the organic beans. Contrary to expectations, a negative bias is found. However, this hypothetical bias is only statistically significant in the inter-respondent comparison, not in the intra-respondent comparison. Most existing studies in environmental economics find a positive bias, where actual purchase behavior results in lower values than those associated with a hypothetical behavioral intent. Only the transportation literature reports a similar negative bias when valuing commuting time. Comparing actual purchase behavior and the real prices paid for different bean products with the prices of the same hypothetically chosen beans shows no significant difference in the intra-respondent comparison. This questions the validity of the inter-respondent comparisons carried out in the literature so far to assess hypothetical bias. Possible explanations for the negative hypothetical bias found in this study are hard to give. Possible procedural bias due to the endowment is dismissed since participant spending was only slightly more in the real experiment than in the hypothetical experiment. The real prices participants paid for the beans were considerably lower than the estimated WTP values based on the estimated choice models in the intra-respondent comparison. This may have inflated the negative hypothetical bias.

Using the actual prices paid in the experiment instead of the estimated WTP values in the intra-respondent comparison furthermore results in a somewhat different picture. Depending on the specific product specification, the negative hypothetical bias is either reversed into a positive bias as in the existing environmental economics literature, or results in a non-significant difference between hypothetical WTP and real expenditures for two of the three products investigated in this study. More intra-respondent comparisons are needed to assess the
robustness of the results presented here, putting existing split sample approaches further to the test.

### 6.2 Policy recommendations

A key outcome of the discrete choice experiments developed and applied in this PhD study is the generation of benefit estimates to be used in cost-benefit analysis of public and private investment decisions in urban waste collection and drinking water supply in the city Hawassa, water harvesting ponds on private land and organic beans sold on a rural agricultural market in the Southern Nations, Nationalities and Peoples’ Region in Ethiopia, of which Hawassa is the capital city.

Besides the generation of these benefit estimates, the studies investigating public WTP for urban waste collection and drinking water supply inform policy and decision-makers at the same time about the expected financial cost recovery of these investment decisions. Cost recovery rates of public utilities such as drinking water supply and wastewater treatment are typically low in developing countries, while demand for more reliable services is high and rapidly growing. Solid waste collection is organized in Hawassa to a large extent by private entrepreneurs, and hence cost recovery is less of an issue in this sector. The service would simply not be provided if no profit would be made.

In the case of investment decisions in private water ponds on agricultural land, the payment mechanism was the provision of a private loan, which had to be paid back by the farmers based on market interest rates over a period of time varying between 5 and 15 years. The current
repayment period of 12 months is a constraint given average income levels in the study area, and many if not most farmers are therefore unable to participate in the loan service provided by the local government. The practical relevance of the results from this survey is that they especially help regional policymakers identify how water harvesting can be rolled out as a farm household climate risk adaptation and mitigation strategy. Sixty percent of the sampled farmers in this survey indicated to face severe droughts and water shortages. Previous attempts to build heavily subsidized community water ponds in the area largely failed. In the DCE, we tried to assess if framing water security management in the study area as a private investment decision, facilitated through a contractual agreement, would be of interest to the farmers involved. The results in this case study clearly show that demand for the offered contractual agreements is high and farmers are willing to pay market interest levels for the provided micro credit to obtain the necessary financial support for this investment decision. The results from the DCE furthermore showed that effective uptake of contractual agreements has to be supported through adequate communication. It is very important that farmers understand the specific contractual terms and conditions they commit to. The characteristics of the water harvesting technology such as pond lining, cover and capacity were valued highly, and therefore have to be adequately addressed in future regional climate risk adaptation and mitigation strategies.

The results from the haricot beans study primarily guide rural agricultural policy to see if agricultural practices matter in consumers’ decisions. Consumers seem to prefer conventional beans to organic ones. Although the two products look the same, organically grown beans are generally free of pesticides and artificial manure, and may therefore be more damaged due to insects and diseases. Newly introduced bean varieties by government research centers in rural agricultural development programs throughout Ethiopia appear to be preferred to local beans,
suggesting that these programs have been successful in changing both producers’ and consumers’ preferences for beans varieties.

Public WTP appeared to be especially high for urban solid waste collection and safe drinking water supply. Waste collection is irregular while collection rates are renegotiated regularly on an individual household basis. Moreover, waste collection and separation are often carried out by children as young as 10 years old. Households in Hawassa are willing to pay 20-30 percent extra over and above their current service charge if the current collection frequency would be increased from once per week to twice or three times per week. Public WTP is significantly higher in richer neighborhoods. Importantly, households are willing to pay a significant premium over and above existing service charges to abolish child labor in the waste sector, especially households with female household heads and higher income levels. There is furthermore public interest in the separation of recyclable and non-recyclable waste when households are provided with waste bins, except from larger households, the heads of whom in this study appeared to be generally lower educated. These results provide important information to municipality officials and the waste management sector as these attributes can be targeted to generate additional revenues and design appropriate strategies to improve existing solid waste management services.

All households in our sample in Hawassa boil their water before they drink it. Water quality is not constant and often unsafe to drink. Despite significant income constraints, all households appeared to be willing to pay substantially extra for improved levels of water supply, especially those households living in the poorest part of the city with the lowest service levels and who pay more for alternative bottled water. Almost 60 percent of the interviewed households pay
every month the same amount of money for their tap water as for alternative drinking water sources such as bottled water. Mean WTP for more reliable water supply varies between 25 and 50 percent over and above the current household water bill. If water quality is improved at the same time, this results in an almost twice as high additional WTP depending on the extra days of water supply. Aggregating the estimated individual household WTP values across the total number of households in Hawassa under the assumption that the survey sample is representative yields a rough indicator of the total benefits of future investment plans in improved water supply services.

Although not a primary focus of the PhD thesis, important gender effects were detected in the latter two surveys related to urban waste collection and drinking water supply. This may not come as a surprise since the availability of safe drinking water supply and household waste collection affects in most cases women who play a central role in the households. These effects were not present in the rural studies focusing on water harvesting and organic beans. In the water harvesting survey this may be partly due to the fact that most of the participants were male household heads. In the beans market survey, most of the participants were, however, female household members. Women were expected to benefit most from improvement scenarios which include the abolishment of child labor and improve water quality. They take care of infants in the household and therefore value the improvement of water quality to a level where boiling for infants is not necessary anymore most, and they were more interested in increasing the collection frequency and valued the abolishment of child labor more than men.

Finally, DCEs are applied across many fields in developed countries and could play a similar role in developing countries. To this end, this dissertation explored the applicability of DCEs
in a series of case studies in Ethiopia. The fact that the case studies in this dissertation are all from Ethiopia and within specific fields of environmental and agricultural economics limits the generalization of the case study findings to other sub-Saharan countries and fields of applications. More studies in other African countries to assess the representativeness of the outcomes presented here are needed, as well as a broadening of the scope of applications in order to be able to generalize the present case study findings to other areas and developing countries.
References


Central Pollution Control Board (CPCB) (2000). Management of municipal solid waste, Ministry of Environment and Forests, Delhi, India.


Transportation Research Part B: Methodology, 44: 735-752.

Hensher, D.A. and Green, W.H. (2003). A latent class model for discrete choice analysis:

Cambridge University Press, Cambridge, UK.

respondents ignoring specific attributes. Transportation, 32: 203-222.

Hensher, D.A. and Greene, W.H. (2010). Non-attendance and dual processing of common-
413-426.

Cambridge University Press, Cambridge, USA.

to avoid drought restrictions. The Economic Record, 82(256): 56–66.

Proceedings from the inaugural international choice modeling conference, 410-427.


difference of empirical distributions: Application to internal and external scope tests in

private solid waste management services in Ibadan, Nigeria. Waste Management and
Research, 27(3): 961-965.

Region: Prospects and constraints. Proceedings of the 9th International Conference on the
Ethiopian Economy, Ethiopian Economic Association (EEA), Addis Ababa, Ethiopia.

Rathi, S. (2007). Optimization model for integrated municipal solid waste management in


