UNIVERSITY OF COPENHAGEN

Københavns Universitet

Danish consumers' willingness to pay for certified wood products a contingent ranking study

Ladenburg, Jacob; Martinsen, Louise

Publication date:
2004

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Ladenburg, J., & Martinsen, L. (2004). *Danish consumers' willingness to pay for certified wood products: a contingent ranking study*. Kongelige Veterinære Landbohøjskole. Skrifter fra FOI: samfundsvidenskabelige serie http://www.foi.life.ku.dk/English/Publications/Reports/~/media/migration%20folder/upload/foi/docs/publikationer/rapporter/nummererede%20rapporter/unit%20of%20economics/2004/14.pdf.ashx

Download date: 06. Oct. 2026

SKRIFTER
FRA
Sektion for Økonomi

SAMFUNDSVIDENSKABELIGE SERIE 14
2004

DANISH CONSUMERS' WILLINGNESS TO PAY FOR CERTIFIED WOOD PRODUCTS

- A CONTINGENT RANKING STUDY

Jacob Ladenburg
Louise Martinsen

SOCIAL SCIENCE SERIES

Papers from Unit of Economics
Danish Research Institute of Food Economics
Royal Veterinary and Agricultural University, Denmark
Den Kongelige Veterinær- og Landbohøjskole

Preface

The present paper on Danish consumers' willingness to pay for certified wood products is a master thesis in Environmental and Natural Resource Economics at the Royal Veterinary and Agricultural University (KVL), Copenhagen, Denmark.

The primary purpose of the study is to assess Danish consumers' willingness to pay for certified wood products. The study is based on a survey of 900 respondents using the stated preference method Contingent Ranking. The study primarily addresses people with an interest in the application of environmental valuation methods and eco-labelling.

We would like to take this opportunity to thank the people who have been of great help during the process of preparing the report and carrying out the survey.

First of all, we would like to thank Scancom International, Tropical Forest Trust and Vemmetofte Kloster for the financial support, which have made this project possible.

For discussion and constructive comments during different phases of the project we would like to thank our supervisor Mr. Alex Dubgaard, Mr. Jørgen D. Jensen and Mr. Claus Ekstrøm.

Last, but not least, we would like to thank Mr. Michael. G. Arentsen and Ms. Maria Skotte for both inspiring discussion and moral support throughout the preparation of the project.

Copenhagen May 2004

Jacob Ladenburg Louise Martinsen

Abstract

The main objectives of the present study are to assess Danish consumers' Willingness to Pay (WTP) for certified wood products, and to explore the link between this WTP and the welfare economic value of the environmental benefits associated with the transition from Conventional Forest Management (CFM) to Sustainable Forest Management (SFM).

The assessment of consumers' WTP for certification is based on a mail-administered survey of 376 randomly sampled Danish consumers using the Contingent Ranking Method. Potential differences in WTP across product types are investigated by eliciting respondents' WTP for three different products: toilet paper, cutting boards and table tops. Furthermore, the effect of information on WTP for toilet paper and cutting boards is investigated by sub-sampling with two different levels of information about certification.

A thorough account is made of the different steps of the analysis leading to the derivation of the final models. These are all in accordance with economic theory and intuition, and have acceptable explanatory power. Despite this, the estimated WTP's are found to represent overestimates of actual WTP's. Thus, the average sample WTP's as a percentage of the price of the products range from 57–97%. The results reveal no definite relationship between the type or price of a product and WTP. Information was only found to be significant for cutting boards, where increasing information was found to increase WTP significantly.

The link between WTP for certified for wood products and welfare economic value of the benefits associated with SFM is primarily explored on a theoretical basis, but it is also discussed in the context of the empirical findings of the study. Certification may be perceived to represent the benefits associated with SFM. Despite this, it is found that estimates of WTP for certified products cannot readily be extended into an estimate of the welfare economic value of the transition from CFM to SFM.

The weak complementarity between wood products and the environmental benefits associated with certification impose limitations on individuals' options for expressing their preferences for certification and thereby SFM. Moreover, only that part of the benefits entailed by SFM, which individuals perceive as complementary to wood products will be reflected in their WTP. On a theoretical level, WTP's for certification can therefore only be expected to provide a lower bound estimate of the welfare economic value associated with SFM. However, obtaining such a lower bound estimate is quite complex, since it requires aggregation of WTP across total purchases of certified wood products.

Table of Contents

1 INTRODUCTION 1

1.1 OBJECTIVES 3
1.2 LIMITATIONS 4
1.3 OUTLINE OF REPORT 4

2 FOREST MANAGEMENT AND CERTIFICATION 6

2.1 THE TOTAL ECONOMIC VALUE OF FORESTS 6
2.1.1 USE VALUES 7
2.1.2 NON-USE VALUES 7
2.2 THE RELATIVE IMPORTANCE OF THE VALUE COMPONENTS 8
2.3 TEV AND PUBLIC GOODS 8
2.4 MAXIMISING FOREST REVENUE – SFM VS. CFM 10
2.5 INTERNALISING THE EXTERNALITIES OF CFM 12
2.6 CERTIFICATION 13
2.7 VALUE COMPONENTS AND TYPE OF GOODS AFFECTING WTP 14
2.8 SUMMARY 16

3 ECONOMIC THEORY OF THE CONSUMER 17

3.1 ECONOMIC THEORY OF THE CONSUMER 17
3.1.1 PREFERENCE RELATIONS 17
3.1.2 THE UTILITY FUNCTION 19
3.1.3 THE LANCASTER APPROACH TO ECONOMIC THEORY 22
3.2 WELFARE MEASURES 23
3.2.1 MARSHALLIAN CONSUMER SURPLUS (M-CS) 23
3.2.2 HICKSIAN WELFARE MEASURES 23
3.2.3 VARIATION MEASURE 23
3.2.4 SURPLUS MEASURE 24
3.3 CHOICE OF WELFARE MEASURE 24
3.3.1 VARIATION VS. SURPLUS MEASURE 25
3.3.2 COMPENSATING VS. EQUIVALENT 26
3.4 DERIVATION OF THE COMPENSATING SURPLUS MEASURE 26
3.4.1 CS FOR AN IMPOSED CHANGE 28
3.4.2 CS FOR A VOLUNTARY CHANGE 29
3.4.3 COMPARING THE IMPOSED AND VOLUNTARY CS 29
3.5 COMPLEMENTARITY 29

3.5.1 Weak Complementarity 30
3.5.2 Weak Complementarity and Non-use Values 31
3.6 Free-Riding, Economic Theory and Willingness to Pay 33
3.6.1 Citizen vs. Consumer 33
3.6.2 Altruism 34
3.6.3 Types of Altruism 35
3.7 Compensating Surplus, Donations and Altruism 36
3.7.1 Deriving CS in the Case of Donations 37
3.8 Compensating Surplus and WTP for Certification 38
3.9 Summary 39

4 ENVIRONMENTAL VALUATION METHODS 40

4.1 Introduction to Environmental Valuation 40
4.2 Valuation Methods 40
4.2.1 Revealed or Stated Preference Methods 41
4.3 Stated Preference Methods 42
4.3.1 The Contingent Valuation Method (CVM) 42
4.3.2 Choice Modelling Methods(CMM) 43
4.4 CVM or CMM 44
4.5 Different Choice Modelling Methods 45
4.5.1 Choice Experiments (CE) 45
4.5.2 Contingent Ranking Method (CRM) 46
4.5.3 Contingent Rating (CR) 46
4.6 Choice Experiments, Contingent Ranking or Contingent Rating 47
4.6.1 Contingent Rating vs. Contingent Ranking 47
4.6.2 Choice Experiment vs. Contingent Ranking 47
4.7 Summary 48

5 POTENTIAL BIASES, VALIDITY AND RELIABILITY 49

5.1 Potential Biases Associated with the CRM 49
5.2 Incentives to Misrepresent Responses 50
5.2.1 Strategic Bias 50
5.2.2 Yea-saying 50
5.3 Implied Value Cues 50
5.3.1 Starting Point/Range Bias 51
5.3.2 Importance Bias 51
5.4 Scenario Misspecification Bias 51
5.4.1 Amenity Misspecification Bias 51

5.4.2 Context Misspecification Bias 53
5.5 Hypothetical Bias 54
5.6 Validity 55
5.6.1 Content Validity 56
5.6.2 Criterion Validity 57
5.6.3 Construct Validity 57
5.6.4 Convergent Validity 57
5.7 Reliability 58
5.8 Summary 58

6 THE QUESTIONNAIRE 60

6.1 Steps in Stated Preference Experiments 60
6.2 Identification of Wood Products (Step 0) 60
6.3 Identification of Attributes and Attribute Levels (Step 1) 61
6.4 Number of Alternatives, Attributes and Attribute Levels (Step 2 and 3) 62
6.4.1 Cognitive Burden 63
6.4.2 Choice of Attributes and Attribute Levels 63
6.4.3 Toilet Paper 63
6.4.4 Cutting Boards 64
6.4.5 Table Tops 64
6.4.6 Prices 64
6.4.7 Summary 66
6.5 The Experimental Design (Step 4) 66
6.5.1 Full Factorial Design 66
6.5.2 Fractional Factorial Design 67
6.5.3 Design Efficiency 67
6.5.4 Construction of Fractional Factorial Designs 69
6.6 The Questionnaire 69
6.6.1 Introduction 70
6.6.2 First Part of the Questionnaire 70
6.6.3 The Hypothetical Scenario 70
6.6.4 The Last Part of the Questionnaire 73
6.7 Summary 74

7 MODELLING DISCRETE CHOICE DATA 75

7.1 Discrete Choices 75
7.2 The Random Utility Model 76
7.3 The Functional Form of the Utility Function 78

7.4 THE MULTINOMIAL LOGIT MODEL 78
7.5 THE RANK LOGIT MODEL – OR THE EXPLODED LOGIT MODEL 79
7.6 LIMITATIONS OF LOGIT MODELS 83
7.6.1 THE IIIA ASSUMPTION 84
7.6.2 THE IID ASSUMPTION 84
7.7 MIXED LOGIT 85
7.8 MAXIMUM LIKELIHOOD ESTIMATION 85
7.9 DERIVING ESTIMATES OF MARGINAL WILLINGNESS TO PAY 86
7.10 SUMMARY 87

8 THE DATA SET 88

8.1 SAMPLE SIZE AND SELECTION 88
8.1.1 SAMPLE SIZE 88
8.1.2 SAMPLE SELECTION 88
8.1.3 SAMPLE SIZE AND MODEL RELIABILITY 89
8.2 RESPONSE RATES AND EFFECTIVE SAMPLE SIZES 89
8.3 REPRESENTATIVITY OF THE SAMPLE 90
8.3.1 REPRESENTATIVITY IN TERMS OF AGE 91
8.3.2 REPRESENTATIVITY IN TERMS OF GENDER 92
8.3.3 REPRESENTATIVITY IN TERMS OF REGION 92
8.3.4 REPRESENTATIVITY IN TERMS OF INCOME 93
8.3.5 REPRESENTATIVITY IN TERMS OF EDUCATION 94
8.3.6 REPRESENTATIVITY IN TERMS OF FOREST VISITS 95
8.3.7 REPRESENTATIVITY IN TERMS OF RESIDENCE TYPE 95
8.4 REPRESENTATIVITY OF THE DANISH POPULATION 96
8.5 ORIGINAL SAMPLE VS. EFFECTIVE SAMPLE 96
8.6 SUMMARY 96

9 THE ANALYSIS 97

9.1 THE STEPS OF THE ANALYSIS 97
9.1.1 STEP 1: EQUALITY OF BETA-COEFFICIENT VECTORS ACROSS INFORMATION SCENARIOS 97
9.1.2 STEP 2: ATTRIBUTE-ONLY MODELS 98
9.1.3 STEP 3: UNIVARIABLE ANALYSIS OF POTENTIALLY RELEVANT VARIABLES 101
9.1.4 STEP 4: DERIVING THE FINAL CHOICE MODEL 107
9.1.5 STEP 5: COMPARISON OF MODELS 108
9.1.6 STEP 6: TEST FOR VIOLATION OF IIA 109
9.1.7 STEP 7: ESTIMATION OF THE MODEL EXCLUDING UNCERTAIN RESPONDENTS 109
9.2 THE SOFTWARE 110

9.3 Analysis of Toilet Paper Data 111
9.3.1 Step 1: Toilet paper 111
9.3.2 Step 2: Toilet Paper 112
9.3.3 Step 3: Toilet Paper 114
9.3.4 Step 4: Toilet Paper 114
9.3.5 Step 5: Toilet Paper 116
9.3.6 Step 6: Toilet Paper 116
9.3.7 Step 7: Toilet Paper 117
9.4 Analysis of Cutting Boards Data Set 117
9.4.1 Step 1: Cutting Boards 117
9.4.2 Step 2: Cutting Boards 117
9.4.3 Step 3: Cutting Boards 119
9.4.4 Step 4: Cutting Boards 120
9.4.5 Step 5: Cutting Boards 121
9.4.6 Step 6: Cutting Boards 122
9.4.7 Step 7: Cutting Boards 122
9.5 Analysis of the Table Tops Data Set 123
9.5.1 The Data Set 123
9.5.2 Step 2: Table Tops 123
9.5.3 Step 3: Table Tops 125
9.5.4 Step 4: Table Tops 126
9.5.5 Step 5: Table Tops 128
9.5.6 Step 6: Table Tops 128
9.5.7 Step 7: Table Tops 129
9.6 Summary 130

10 ESTIMATION OF WILLINGNESS TO PAY 131

10.1 Mean vs. Median WTP 131
10.2 WTP for Certification, Toilet Paper 131
10.3 WTP for Certification, Cutting Boards 132
10.4 WTP for Certification, Table Tops 134
10.5 Comparison of WTP across product types 135
10.6 Willingness to pay for other product attributes 136
10.7 Sample WTP vs. Population WTP 137
10.7.1 Toilet paper 137
10.7.2 Cutting boards 138
10.7.3 Table tops 139
10.8 Summary 141

11 VALIDATION OF THE RESULTS 142

11.1 CONTENT VALIDITY 142
11.2 CONSTRUCT VALIDITY 142
11.2.1 GENERAL ISSUES 142
11.2.2 TOILET PAPER 145
11.2.3 CUTTING BOARDS 146
11.2.4 TABLE TOPS 146
11.2.5 SUMMING UP CONSTRUCT VALIDITY 147
11.3 CRITERION VALIDITY 148
11.4 CONVERGENT VALIDITY 148
11.4.1 OTHER STUDIES OF WTP FOR CERTIFICATION 148
11.4.2 WTP FOR ECO-LABELLED NON-WOOD PRODUCTS 150
11.4.3 SUMMING UP ON CONVERGENT VALIDITY 151
11.5 BIASES 151
11.5.1 YEA-SAYING 151
11.5.2 IMPORTANCE BIAS 152
11.5.3 PROBABILITY OF PROVISION BIAS 153
11.5.4 HYPOTHETICAL BIAS 153
11.5.5 ORDERING BIAS 154
11.5.6 FRAMING/INFORMATION 154
11.5.7 DISREGARD OF BUDGET CONSTRAINT 154
11.6 RELIABILITY OF WTP ESTIMATES 155
11.7 VALIDITY, BIASES AND CRM 155
11.8 SUMMARY 155

12 DISCUSSION 157

12.1 ESTIMATED VS. EXPECTED COMPENSATING SURPLUS 157
12.2 IMPLICATIONS OF WEAK COMPLEMENTARITY 157
12.2.1 CHOKE PRICE 157
12.2.2 CONSTRAINTS ON PREFERENCE EXPRESSION 157
12.3 POTENTIAL EFFECT OF ALTRUISTIC PREFERENCES 158
12.4 SUMMARY 159

13 CONCLUSION 160

REFERENCES 162

1 Introduction

Since ”Our Common Future” was published (WCED, 1987), and certainly after the Rio Summit in 1992, environmental issues have increasingly gained attention, and the perception that current exploitation of the earth’s natural resources is unsustainable appear to be widespread (Worldwatch Institute, 2003). In this context, reports of rapid deforestation and forest degradation, combined with accounts of the potential detrimental consequences of such excessive exploitation, have lead to concerns being raised regarding the sustainability of Conventional Forest Management (CFM) practices (FAO, 1997).

Traditionally, forest management has focused on the production of timber (Price, 1989; Klemperer, 1996). However, it has increasingly been recognized that forests give rise to other benefits than those associated with the extraction of timber and other readily marketable goods. Thus, seen from a welfare economic perspective the total value of forests also encompasses values associated with e.g. carbon sequestration, biodiversity and recreation (Pearce & Warford, 1993; CBD, 2001). Characteristic for many, if not most, of these additional values is that they fall within the category of public goods. Since no market exists for these goods, they represent no private economic value to the forest manager (Lampietti & Dixon, 1995, Adamowicz *et al.*, 1996). Consequently, seen from the forest managers point of view, the optimal management regime is that, which serve to maximize the profits accruing from timber extraction activities (Price, 1989; Klemperer, 1996).

The public goods and timber can only be produced jointly to a certain extent; that is, above a certain threshold the extraction of timber has a negative impact on the forests ability to e.g sequester carbon and support biodiversity (CBD, 2001; Pearce *et al.*, 1999; Lampietti & Dixon, 1995). Consequently, CFM regimes with focus on timber production in most cases entail an inefficient supply of public good forest benefits. Therefore such a management regime will in most cases not be optimal seen from a welfare economic point of view (Kahn, 2002). In this context, Sustainable Forest Management (SFM) may be defined as a management regime where the value, which the public good benefits from forests represent to society, is taken into consideration. Thus, SFM may be perceived to represent a management regime that is optimal seen from a welfare economic point of view (Pearce *et al.*, 1999)

In response to the concern regarding the sustainability of CFM, several NGO’s joined forces and founded the Forest Stewardship Council (FSC) in 1993 (Upton & Bass, 1995). FSC was the first organisation to launch certification as a means to promote sustainable forest management.

Certification is an example of an environmental labelling scheme. The basic idea of certification is that SFM can be promoted by creating a market for sustainably produced - i.e. certified – wood products (Kiker & Putz, 1997; Upton & Bass, 1995). More specifically, three of the main assumptions underlying the approach are that (Richards, 1999):

1. Consumers have positive preferences for certified products, and they are willing to pay a premium for such products.
2. The premium that certified products can be traded at will provide sufficient incentive for forest managers to adopt SFM practices.
3. The development of a market for certified products is hampered by the unobservable nature of the environmental public good attributes, which distinguish certified from conventionally produced products.

In this context, the basic purpose of certification is to enable consumers to distinguish between certified and conventional products. This is accomplished by attaching a label to products originating from a SFM forest, thereby turning the otherwise unobservable environmental attribute into a readily observable attribute (Nunes & Riyanto, 2001; Teisl & Roe, 2000)

The environmental attribute, which distinguishes certified from conventional products, can be characterised as a public good attribute (Karl & Orwat, 1999). Assuming that consumers are willing to pay a premium for certified products is therefore basically equivalent assuming that consumers are willing to voluntarily contribute to the provision of a public good. Such an assumption is incompatible with the free-rider theory of traditional economics (Cornes & Sandler, 1996; Gravelle & Rees, 1992). However, there exist alternatives to the traditional free-rider theory. These alternatives suggest that consumers based on completely rational motives may choose to contribute to the provision of public goods (Andreoni, 1989; Andreoni & Miller, 2002; Johansson-Stenman, 1997). That such motives may be relevant in relation to consumers' propensity to exhibit a positive WTP for certified products is supported by empirical studies. Thus, several studies find that consumers have a positive WTP for different kinds of eco-labelled products, among these certified wood products (Bjørner *et al.*, 2002; Andersen, 2002b; Jensen *et al.*, 2002; Veisten, 2002; Ozanne & Vlosky, 1997).

The focus of the present study will be on the consumer behavioural assumption underlying certification (i.e. assumption 1 mentioned above). More specifically, the aim is to assess Danish consumers' willingness-to-pay (WTP) a premium for FSC-

certified, as opposed to conventional, wood products. The study will be based on a survey of Danish consumers using the Contingent Ranking Method (CRM).

One of the primary reasons for choosing CRM is that, in actual purchase situations, the importance attached to products' environmental attributes may naturally be outweighed by the importance attached to other product attributes (Rametsteiner, 1999; Spinazze & Kant, 1999; Sriram & Forman, 1993). It is therefore considered relevant to extend the focus to include other attributes, rather than focus on certification in isolation. Using the CRM, a frame where respondents are asked to consider all attributes jointly can be created (Adamowicz *et al.*, 1998a; Hanley *et al.*, 2001). This way, respondents may be less inclined to put undue emphasis on certification. Likewise, assessment of the relative importance of certification compared to other attributes is made possible.

Previous studies of consumer's WTP for eco-labelled products indicate that WTP may vary across different types of product and/or different price ranges (Ozanne & Vlosky 1997; Pickering *et al.*, 2001). In order to accommodate such potentially different WTP's, and thereby get a more nuanced picture of individuals' preferences for certification, the survey will include three different wood product types representing different price levels.

Lastly, it is expected that consumers' WTP for certified products is contingent upon their knowledge about the merits of certification. Thus, it is expected that consumers' WTP will depend on the contents of the information that is made available to them (Caswell, 1997; OECD, 2001). In order to investigate the effect of information on consumers' WTP for certification, respondents will be divided into two subsamples, which are provided with different levels of information about certification.

1.1 Objectives

The overall objective of the study is to estimate Danish consumers' Willingness-to-Pay (WTP) for certified wood products, and explore the extent to which such WTP estimates reflect consumers' WTP for sustainable forest management.

The more specific objectives are:

- To assess the extent to which WTP vary across different types of products.
- To investigate if WTP is influenced by the level of information about certification given to the respondent.

- To assess the WTP for certification relative to the WTP for other product attributes.

- To evaluate the contingent ranking method as a means to elicit preferences for certification.

1.2 Limitations

- Several certification schemes exist. However we have restricted our survey to consider one scheme, the Forest Stewardship Council (FSC), which currently is the most well established scheme worldwide (FAO, 2002).

- This study only examines the welfare economic value of the benefits of certification seen from the consumer's point of view.

- This study only estimates a WTP for three specific wood products, why an aggregated WTP for sustainable forest management cannot be identified.

- The thesis does not discuss the credibility of forest certification. It is therefore assumed that the FSC-scheme is credible.

- Certification deals with all both the economic, social and environmental dimension of sustainability; in this study, however, focus is limited to the environmental dimension.

- The terms *forest* and *forest management* are used broadly in this study. That is, no distinction is made between the different types of forests and management regimes that are found worldwide.

1.3 Outline of Report

Chapter 2: The concepts of sustainable forest management and certification are introduced. In this context the goods and services subjected to valuation in the study are identified and characterised.

Chapter 3: The economic theory underlying valuation of environmental goods is presented. The relevant welfare measure to use in the study is derived, and its merits in relation to assessing WTP for certification and SFM are explored. Furthermore, motives for stating a positive WTP are discussed.

Chapter 4: Different valuation methods are presented, and the background for choosing the contingent ranking method in this study is discussed.

Chapter 5: Potential biases that may affect the validity of the results are discussed, and the different criteria for assessing the validity of the study are introduced.

Chapter 6: The design of the questionnaire and the experimental design are described.

Chapter 7: The econometric models used in the analysis of data are presented, and important aspects pertaining to modelling based on rank data are discussed.

Chapter 8: The representativeness of the effective sample compared to the Danish population is investigated.

Chapter 9: The chapter contains a thorough account of the analysis of data, leading to the derivation of the models from which estimates of WTP for certification are obtained.

Chapter 10: Estimates of sample mean and median WTP for certification for the different products are presented. Subsequently, these estimates are adjusted to arrive at estimates of the mean WTP for certification for the Danish population.

Chapter 11: The validity of the study is assessed against the criteria presented in chapter 3. There after, the presence of biases is discussed.

Chapter 12: Contains a brief discussion of the results in the context of the theory presented in chapter 3.

Chapter 13: Conclusion.

2 Forest Management and Certification

The focus of the present project is on assessing Danish consumers' Willingness to Pay (WTP) for certified wood products. As it will be explained in chapter 3, this WTP is supposed to reflect the value that consumers associate with Sustainable Forest Management (SFM) as opposed to Conventional Forest Management (CFM). The purpose of the present chapter is to clarify what distinguishes SFM from CFM and thereby identify and characterise the goods and services, which are being valued in the survey. Furthermore, it is the purpose to specify the role played by certification in relation to promoting SFM.

2.1 The Total Economic Value of Forests

The value concept encompassing the aggregate value of all the goods and services provided by an environmental resource, in this case forests, is referred to as the Total Economic Value (TEV) of the resource (Lampietti & Dixon, 1995).

The value components comprising TEV are presented in table 2.1 where it appears that TEV can be divided into use and non-use values[1]. Subsequently, Use Values (UV) can be divided into; Direct Use Values (DUV), Indirect Use Values (IUV) and Option Values (OV)/Quasi-Option Values (QOV). Non-Use Values (NUV) can similarly be divided into; Existence Values (EV) and Bequest Values (BV) (Georgiou *et al.*, 1997; Freeman, 1993; Pearce & Turner, 1990).

Table 2.1 The values comprising the total economic value of environmental goods (adapted from Georgiou *et al.*, 1997).

Total Economic Value				
Use Values			Non-Use Values	
Direct	Indirect	Option / Quasi-option	Existence	Bequest
Directly consumable outputs	Functional benefits	Future direct and indirect use values	Values of environmental legacy	Value from knowledge of continued existence

In principle, the components comprising TEV are additive. However, in practice some of the goods and/or services are mutually exclusive or can only be produced jointly to a certain extent. In connection to forests, maximising direct use values by clear felling is likely to be incompatible with e.g. maximisation of indirect climatic benefits (CBD, 2001).

[1] In the literature non-use values are often referred to as *passive use values* or *preservation values* (Georgiou *et al.*, 1997; Lampietti & Dixon, 1995).

2.1.1 Use Values

DUV are the values associated with the direct use of the forest. The direct use may take several different forms – e.g. timber harvesting activities, extraction of other goods or recreation. However, common for the direct uses of the forest is, that they, though with vastly different severity, have a physical impact on the forest environment.

In contrast, the IUV arise from the natural functioning of the forest ecosystem, and as such, appropriation of IUV's does not leave any physical impact on the forest. IUV's may be exemplified by the value of services such as carbon sequestration, watershed protection and biodiversity protection.(CBD, 2001; Sedjo *et al.*, 1997, Adamowicz *et al.*, 1996; Abt & Sills, 2003). It is important to note, that while some IUV's primarily are local – e.g. watershed protection – others, such as carbon sequestration, are global in nature.

OV and QOV do not originate from current use. They relate to benefits that may be derived from future uses of the forest. Both value concepts are most often exemplified by the potential future value of biodiversity in relation to the development of pharmaceutical products, called bioprospecting (Southgate, 1998).

OV may be seen as an expression of the subjective value that individuals assign to preserving resources, where availability in the future otherwise would be uncertain (Weisbrod, 1964). Thus OV may be seen to represent, the premium that individuals are willing to pay to preserve a resource that they want the option of using in the future Lampietti & Dixon (1995).

QOV, on the other hand, relates to the actual, but uncertain, value to society of expected benefits (Arrow & Fischer, 1974). Thus, contrary to OV, QOV is based on the more or less specific expectation that increases in knowledge will enable the present – or future generations – to realise the true value of the resource (Pearce & Turner, 1990).

2.1.2 Non-Use Values

As the term indicates, NUV are based on the notion that individuals may assign value to goods, which they neither do nor intend to make use of.

EV are related to the mere existence of the good (Krutilla, 1967). That is, it represents the value derived from knowing that the resource exists. It is often described as the value attached to the minimum viable size of the resource, and thereby the values that would be lost if a species were extinct or a resource depleted (Freeman, 1993).

BV refer to the value derived from preserving a resource for future generations to enjoy – that is either one's heirs or future generations in general (Freeman, 1979). Thus, BV are based on altruistic motives and may include both use and non-use values.

2.2 The Relative Importance of the Value Components

During the past decades increasing attention has been focused on estimating the magnitude of the different value components comprising TEV of forests (Adger *et al.*, 1995). Several reviews of different valuation studies have shown that TEV, and the relative importance of the different value components comprising TEV, varies from forest to forest (CBD, 2001; Lampietti & Dixon, 1995; Southgate, 1998)[2].

However, it is possible to draw at least three general conclusions concerning the composition of TEV (CBD, 2001).

- Values related to the competing services/functions of timber extraction (DUV) and carbon storage (IUV) appear to be the dominating values.

- It is only in relation to unique[3] forests, that other values become more important – i.e. uniqueness tends to be associated with high non-use values.

- Non-use values for "general" or "average" forests are modest.

2.3 TEV and Public Goods

As with other environmental resources, many of the goods and services derived from forests can be characterised as public goods (Hanley *et al.*, 1997).

Public goods, in opposition to private goods, are characterised by being non-rival and non-excludable in their consumption (Samuelson, 1954; 1955). Non-rivalry implies that one person's consumption of a good does not affect the availability of the good to others. Therefore, the opportunity cost of consumption is zero, and, at a given level of public good provisioning, the marginal social cost of letting an additional person consume the good is zero. Non-excludability implies that, once a good has been provided, no one can be excluded from consuming the good.

[2] Both TEV, and the components comprising TEV, may vary from being negative to several thousand $/ha/year (Lampietti & Dixon, 1995; CBD, 2001)

[3] Forests may be unique in themselves; however, they may also be considered unique if they serve as habitat for unique species or if their location is unique, e.g. proximity to towns/cities or important catchment areas with regards to watershed management (CBD, 2001).

Due to non-rivalry, the marginal value of an additional unit of the public good is determined, by the sum of individuals' valuations of the good implying vertical aggregation of marginal demand curves (Gravelle & Rees, 1992). Consequently, the optimal level of public good provisioning is determined, as the level where the sum of N individuals' marginal valuation (MB_i) of the good is equal to its marginal cost of production (MC):

$$\sum_{i}^{N} MB_i = MC$$

Accordingly, the price that each individual should pay to secure the efficient supply of the public good is equal to the individual's marginal valuation of the good.

However, non-excludability implies that the individual cannot be excluded from enjoying the benefits of the public good. Moreover, the contribution of the individual to the good is likely to have only a negligible effect on the total supply of the good. This means, that the rational individual is left with a strong incentive not to contribute to the provision of the public good or to contribute less than their marginal valuation (Wiser & Pickle, 1997). That is, to *free-ride* on others' contributions. It is such free-riding behaviour that rational individuals are expected to exhibit according to traditional economic theory (Cornes & Sandler, 1996). If too many people chose to free-ride, producers are provided with no - or insufficient - incentive to supply the public good[4]. Consequently, the emergence of an efficient market for public goods is hampered.

In reality, markets for public goods are in most cases not just inefficient. They are generally non-existent, implying that the values associated with public goods are not easily and readily appropriated by private agents. With reference to the different value components comprising TEV of forests, it appears that all values – except from the DUV – are derived from public goods. Of these, some can most appropriately be characterised as local public goods – e.g. IUV derived from watershed protection – whereas others are of global character – e.g. IUV derived from carbon sequestration and all the non-use-values (Hanley *et al.*, 1997). This implies that it is only the DUV of forests - i.e. the value associated with private and readily marketable goods - that can be realised by the forest owner.

[4] Depending on the extent of free-riding, that is a zero contribution compared to a minor contribution for the provision of the public good, the term *easy-riding* may be more appropriate (Cornes & Sandler, 1996). However for the ease of presentation, the term *free-riding* will be used in this study.

2.4 Maximising Forest Revenue – SFM vs. CFM[5]

In terms of the goods and services provided by forests, the distinction between private and public goods, which was made in the previous section, is, roughly speaking, equivalent to the distinction between market and non-market goods. This latter distinction is central in relation to illustrating the differences between SFM and CFM.

SFM may be defined as a management regime where focus is on both the marketed and non-marketed goods and services produced by the forests (Pearce *et al.*, 1999). This definition of SFM is rather loose, thus it does not specify the extent to which natural capital can be substituted with man made capital. That is, it is not specified whether SFM should be based on the criteria of weak or strong sustainability (Dubgaard *et al*,. 1999). This is certainly an important aspect of SFM, but not one that will be discussed in this study. In the present context, the essence of SFM is that when compared to CFM it takes all values into account.

Seen from society's point of view SFM is desirable since its focus is on maximising the value of the forest resource to society with regards to all the components comprising TEV. Seen from the private point of view, however, SFM is not that desirable. It most often is less profitable than CFM, since private agents are only able to capture the value associated with the DUV (Pearce *et al.*, 1999; Klemperer, 1996). Consequently, for private agents it is rational to adhere to CFM where focus is on maximising profits accruing from the private marketable goods.

As mentioned previously, some of the goods and services produced by forests are mutually exclusive or can only be produced jointly to a certain extent. Thus, it is often the case that the extraction of marketable goods from forests have a negative impact on the non-marketable services provided by forests. That is extractive activities give rise to environmental externalities. As a consequence of this reciprocal relationship, maximization of private profits is often found to be incompatible with maximisation of social value. The nature of the problem causing discrepancy between what is optimal seen from the private and the social perspective is illustrated in figure 2.1.

[5] SFM and CFM are not uniquely defined management regimes. Thus, there is a wide spectrum of different management regimes. Since the purpose of the present chapter is to illustrate the principal differences between sustainable and unsustainable management regimes the distinction made here is rather crude.

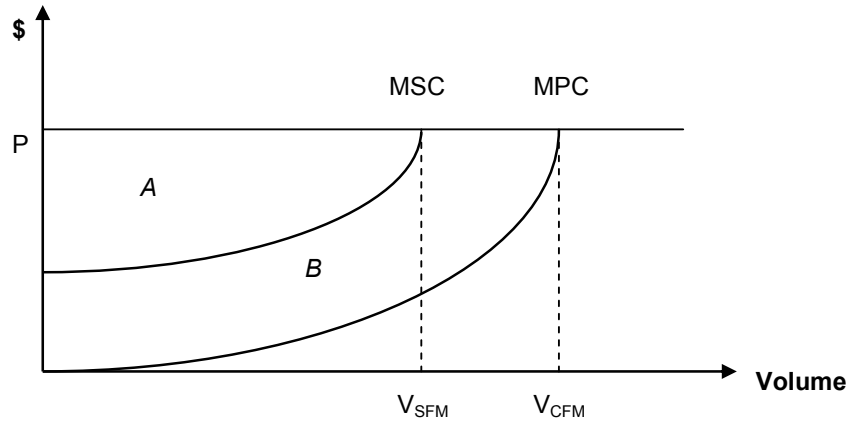


Figure 2.1 Optimal harvest level under CFM and SFM regimes (adapted from Kahn, 2002).

In figure 2.1 V_{CFM} denotes the level of extractive activities that is optimal under a CFM regime, and is determined as the level of production where the marginal private benefits of extraction (P) equal the marginal private costs of extraction (MPC). V_{SFM}, on the other hand, specifies the optimal level of extractive activities under a SFM regime – i.e. the level for which the marginal social benefits of extraction (P) equals the marginal social costs of extraction (MSC). As it is seen, V_{CFM} is greater than V_{SFM}. It is worth noting that the social and private benefits associated with extraction are identical, implying that the difference in optimal harvest levels solely can be attributed to the difference in marginal costs caused by the inclusion of externality costs in MSC[6].

The area *A* + *B* in figure 2.1 illustrates the private resource rent[7] associated with harvesting the trees at a given point in time (Kahn, 2002). As figure 2.1indicates, the private resource rent associated with harvesting V_{CFM} at a given point in time is significantly greater than the rent associated with harvesting V_{SFM} (*A*). Consequently, the transition from CFM to SFM will be associated with private costs (*B*) (Cerda & Lira, 2002), suggesting the reason as to why private agents do not voluntarily choose to operate at the social optimum.

[6] The cost of externalities included in MSC is the dual of the benefits associated with the non-market goods and services provided by the forest. Thus, it represents the non-market value that is foregone by harvesting beyond V_{SFM}.
[7] In this case, the term resource rent refers to the value of the extracted timber, above the cost of production including capital cost and exclusive of the normal return on investment.

2.5 Internalising the Externalities of CFM

Evidently, it is the existence of environmental externalities that drives a wedge between the socially optimal regime (SFM) and the privately optimal regime (CFM). Subsequently, from an economic point of view, promoting SFM is a question of devising methods for the effective and efficient internalisation of the externalities.

With reference to figure 2.1, there are three strategies by which the externalities associated with CFM can be internalised, thereby making the social and the private optima coincide (Baumol & Oates, 1988; Hanley *et al.*, 1997):

Strategy 1: Increase the private costs associated with exploiting forests – i.e. make MPC and MSC coincide (e.g. by imposing taxes on the revenue from timber extraction).

Strategy 2: Compensate the private agent for the profits he/she foregoes by constraining his efforts at the socially optimal level (e.g. by subsidising SFM). The required minimum compensation represented by the area *B*.

Strategy 3: Impose standards for forest management practices – i.e. stipulating that extractive activities should be confined at the level V_{SFM}.

The choice between strategies depends on how the externality problem is perceived[8]. If the externality is seen as a positive side-effect of SFM practices, then strategy 2 is most appropriate, since it may be seen as a way of paying the private agent for the public goods he/she produces. On the other hand, if the externality is perceived to be a negative side-effect of CFM practices, then strategy 1 or 3 are likely to be more appropriate.

According to traditional economic theory, the market forces cannot be relied upon to provide efficient outcomes when externalities of public good nature are concidered. Therefore, governmental intervention is considered a prerequisite for securing efficient internalisation of externalities such as those associated with CFM.

The traditional means of governmental intervention are implementation of regulatory instruments, such as standards (Strategy 3), or economic instruments, such as taxes (Strategy 1) and subsidies (Strategy 2) (Baumol & Oates, 1988; Börkey *et al.*, 1999). In relation to promoting SFM, the implementation of the traditional intervention measures is

[8] The choice between strategies may be seen from a property rights oriented perspective. If society is perceived to be entitled to the public good benefits accruing from SFM, then strategy 1 and 3 may be considered most relevant. If, on the other hand, the private agents are entitled to manage his forest as he/she wished to, then strategy 2 is more relevant.

complicated by the global nature of many of the externalities. Therefore, the problems associated with unsustainable forest management practices may be perceived as a transboundary pollution problem. That is, a problem where a solution is difficult to identify, due to the unequal distribution of benefits and costs among stakeholders (Hanley *et al.*, 1997).

2.6 Certification

On an international level, the widespread occurrence of deforestation and other kinds of excessive exploitation of forests – especially in the tropics – emerged as an important issue in the late 80's/early 90's (FAO, 1997). Since then, several initiatives related to the promotion of SFM practices have been initiated. However, due to incompatible interests of different stakeholders there has not been an international consensus upon the design of intervention schemes.

As a response to the lack of actual intervention on behalf of the established political community, several NGO's joined forces in 1993 and founded the international Forest Stewardship Council (FSC). FSC was the first organisation to launch environmental labelling – i.e. FSC certification - as a means to promote SFM (Upton & Bass, 1995).

The basic idea of FSC certification is, that the externalities associated with CFM partly can be internalised by creating a market where the private and public goods produced jointly under a SFM regime can be traded jointly. Underlying the approach is the assumption that there exists a latent demand for the public good for benefits associated with SFM, i.e. it is assumed that consumers have preferences for the public goods. However the consumers are not able to express their preferences, since they cannot identify SFM products. On a more practical level, the role of certification is to enable consumers to distinguish between products originating from a sustainable managed forest and from a conventionally managed forest. This is accomplished by attaching a label – i.e. a FSC label -, which serves to verify that the given product does indeed stem from a sustainably managed forest (Nunes & Riyanto, 2001; Upton & Bass, 1995, Teisl & Roe, 2000).

Basically, the certification approach to promoting SFM builds on Strategy 2 mentioned in the previous section 2.5. However, in contrast to traditional theory, where governmental intervention is seen as a prerequisite for internalising public good externalities, the certification approach builds on the principle of voluntary participation. The assumptions underlying the certification approach that promotes SFM are (Kiker & Putz, 1997; Richards, 1999):

- That certified products can be traded at a premium compared to conventionally produced products.

- That a sufficient number of consumers do not free-ride.

- That this premium will provide private agent with sufficient incentive to adopt SFM practices.

Stated differently, it is expected that the private economic benefits associated with participating in the FSC scheme will be sufficient to compensate the private agent for the profits he/she foregoes by constraining his efforts at the social optimum[9].

2.7 Value Components and Type of Goods Affecting WTP

Other things equal, the only difference between certified products and conventionally produced products is the amount of public good benefits that is associated with them. Consequently, the price premium on certified products should reflect the value that consumers derive from the forest being managed according to the principles of SFM rather than CFM. Therefore, an assessment of individuals' WTP for certified products may indirectly be used as an indication of the value they associate with SFM.

As mentioned, SFM is not a uniquely definable concept. Likewise, the difference in the amount of public goods associated with SFM and CFM, is not uniquely determined, since the requirements that need to be fulfilled in order to get FSC certified vary from country to country, and from forest to forest. Acknowledging these facts, it is evident that it is more or less impossible to specify the exact nature and extent of the benefits associated with buying a given certified product. Due to this intangible nature of the benefits, the preferences and thereby the WTP for certification are likely to be the result of more general beliefs about the benefits of SFM. People are furthermore expected to combine these beliefs with general attitudes towards forest management and the balance between environmental and economic considerations.

In relation to the nature of the goods influencing individuals' WTP for certification, it is important to recognize that the individuals buying certified products are most probably spatially distanced from the forest where the given product originates. This means that it is the TEV components referring to regional and global public goods that are most likely to influence individuals' WTP for certification.

[9] In addition to private economic benefits in terms of price premiums, improved market shares and/or access, and PR-value are also commonly mentioned as potential benefits associated with participating in the FSC scheme. In the present study, however, focus will be on the price premiums as they represent the most tangible of the potential benefits.

In relation to information on the values of forests, it is expected, that people in general primarily have been exposed to issues related to forest management in the form of either:

- More or less dramatic reports about the potentially catastrophic consequences of ongoing deforestation and other kinds of excessive exploitation.

- Reports about the important roles that forests play in relation to global warming and as habitats for plants and animals.

People are therefore most likely familiar with global public goods such as carbon storage/sequestration along with biodiversity preservation. Subsequently, people are believed to have these "specific" benefits in mind when they construct their preferences for SFM and thereby determine their WTP for certification. Stated differently, it is expected that it is the IUV associated with carbon, along with OV and NUV – both in the form of BV and EV – that give rise to any potential WTP for certification.

As mentioned in section 2.2, the value of the benefits associated with carbon sequestration/storage often account for a significant proportion of forests' TEV whereas the value associated with other IUV's or NUV's only make a significant contribution to TEV for unique forests. It is expected that many individuals may not recognize this, and therefore attach substantial symbolic values to other forest characteristics, such as biodiversity. Thus, when thinking about forests in general, individuals may be inclined to focus on characteristics, which in effect are specific to unique forests. As a consequence of this, the relative importance of the different values in relation to determining WTP for certified products may differ from that, which may be uncovered from assessment of the value of a specific forest.

Due to the spatial distance, combined with the intangible and symbolic nature of the values associated with certification, the WTP for certified products is likely to bear resemblance to donations. Thus the WTP for a given certified product is probably not that dependent on any specific expectations about a predetermined change in the level of the public good caused by the purchase of that given product. Stated differently it is unlikely that a consumer expects a net return of e.g. x species saved per z DKK paid. As such, the WTP is more likely to be based on the expectation that the premium contributes to the promotion of SFM. The implications of this in relation to using WTP for certified products as a mean to assess the welfare economic gain associated with the transition from CFM to SFM will be discussed in the next chapter.

2.8 Summary

Since many of the services and goods associated with forests are public goods for which no market exists, CFM regimes do not take these into account when the rent from the forest resources is maximised. From a welfare economic point of view, this is not a sustainable utilization of the forest resources, since the supply of the public goods produced by forests is inefficient. Certification of forests is a tool to internalise the externalities associated with CFM. This is accomplished by developing a market for the public goods, where consumers with preferences for SFM can identify and purchase wood products originating from sustainably managed forests. Such a market will only work, if consumers do not free-ride. Thus, consumers have to be willing to pay a premium for certified products that is high enough to compensate the forest manager for constraining the production of private goods at the socially optimal level. Subsequently, it is discussed that consumer's willingness to pay for certification may bear resemblance to donations. That is, the willingness to pay is likely to be motivated by a wish to contribute to the promotion of sustainable forest management rather than specific expectations of changes in the level of public goods provided.

3 Economic Theory of the Consumer

To establish the economic foundation of the present study, the neo-classical economic frame of consumer choice and of economic valuation of environmental goods is presented. This is followed by an introduction of the different welfare measures and a discussion of the welfare measure relevant to this study. Since most of the goods and services ensured by certification are public goods, traditional theory suggests that consumers will not be willing to pay for their provision through individual voluntary contributions. In this connection motives that may induce respondents to refrain from free-riding are explored in the context of the present study.

3.1 Economic Theory of the Consumer

The central concept in neo-classical economic theory is that of the rational utility maximising consumer (who is able to rank a feasible set of consumption bundles and choose the one preferred). The relevant aspects of neo-classical consumer theory are presented below. Where nothing else is referred, it is based on Deaton & Muelbauer (1980), Gravelle & Rees (1992) and Varian (1992).

3.1.1 Preference Relations

The neo-classical economic theory takes its point of departure in the consumer's preferences for consumption bundles, in this case wood products. These bundles are in the literature referred to as quantities of goods, but can be applied to qualities or attributes of goods as well (Small & Rosen, 1981; Lancaster, 1966). A consumption bundle can be denoted by a vector:

$$x = (x_1, x_2, \ldots, x_n)$$

where x_i, i=1,2…n, is the quantity of the i'th good.

Presented with two different consumption bundles the consumer is indifferent between bundle x' and x'' if they yield him/her the exact same utility, notated as:

$$x' \sim x''$$

If x' at least yields the consumer the same or more utility than x'' the notation is as follows:

$$x' \succeq x''$$

The symbols $\sim$ and $\succeq$ are called *preference relations*.

In order to consistently verify the choice of the consumer between a feasible set of wood products, and to establish an ordering of preferences, the preferences must satisfy certain standard properties. These are described in the following.

Completeness: For any pair of bundles x' and x'' either $x' \succeq x''$, $x' \preceq x''$ or $x' \sim x''$.

This assumption ensures, that there are no "holes" in the preference ordering. More specifically, it implies that the consumer can *always* express his/her preferences for a given good, rank the bundles and choose the one preferred.

Transitivity: For any three bundles x', x'' and x''' if $x' \succeq x''$ and $x'' \succeq x'''$ then $x' \succeq x'''$.

This assumption ensures that no bundle can belong to more than one indifference set - implying that the indifference sets have no intersections. Transitivity is the requirement for consistency of choices. The assumption of transitivity is fundamental for choices among bundles and therefore the ranking of bundles (Ben-Akiva & Lerman, 1985).

Reflexivity: For any bundle $x' \succeq x'$.

This assumption is rather trivial, stating that a bundle or a wood product is as good as itself. With reflexivity and transitivity a bundle of goods belongs to at least one and no more than one indifference set.

Non-satiation: Any consumption bundle $x' \succ x''$, if x' contains more of at least one good and no less of any other goods compared to x''.

The non-satiation assumption defines that a consumer always prefers a little more of anything everything else being equal. The assumption is easily applied and interpreted for quantities. In the case of e.g. qualitative and attitudinal goods it may be difficult *a priori* to conclude if the good is a positive good and therefore what "a little more" is. But a key point is, that the consumer knows, what "a little more" is, and thereby satisfies the assumption of non-satiation. A last implication of the assumption is that an indifference set can never be wider than a single point.

Continuity: The indifference set represents a continuous surface or curve.

This implies, that the curve of the two-dimensional indifference set has no gaps or breaks at any point. From the consumer's point of view, continuity ensures that a decrease in one good always can be compensated by an increase in another good. That is, trade-offs

between goods are always possible to make the consumer as well off. The assumption rules out the existence of lexicographic preferences where the consumer always chooses the bundle with more of one specific good regardless the levels of the other goods. As such the consumer is unwilling to make trade-offs, that are associated with a decrease in the amount of the good for which he/she has lexicographical preferences. It also rules out the existence of pseudo-lexicographical preferences where consumers are only willing make trade-offs once a certain threshold level of the good has been obtained.

Strict Convexity: Given any consumption bundle x', its better set is strictly convex.

Due to the convex indifference curve, it illustrates the neo-classical assumption of diminishing marginal rate of substitution. Strict convexity means, that the consumer always prefers a mixture of two consumption bundles, which the consumer is indifferent to.

Based on the above-mentioned assumptions, a utility function can be derived. The utility function assigns a value to the indifference curve, such that a higher value represents a higher positioned indifference curve. The utility level provided by a product bundle thereby becomes comparable with the utility of another product bundle, which enables the consumer to rank the bundles. The ranking of bundles thereby represents an ordering – corresponding to the *ordinal* utility function[10].

Definition of a specific utility function requires that, the level of utility u is identical for bundles in the same indifference set: $u(x') = u(x'')$ if and only if $x' \sim x''$, and that bundles on a preferred indifference set has a higher u: $u(x') > u(x'')$ if and only if $x' \succ x''$.

3.1.2 The Utility Function

In this section, the demand for goods in relation to changes in prices/quantities is described. Where nothing else is referred this section is based on Freeman (1993) and Varian (1992).

The consumer derives utility from both the consumption of market goods and environmental goods, e.g. such as taking a walk in the forest and going to the theatre. The different categories of sources of utility are confined in the utility function below.

$$u = u(X, Q, T) \tag{3.1}$$

[10] That the utility function is ordinal implies that the utility of different bundles can be compared on a relative scale, but no absolute values – i.e. cardinal measures of utility - can be identified.

Where X is a vector of quantities of *market goods*, Q is a vector of *public goods* like environmental and resource services, with fixed quantities and qualities for the consumer, and T is a vector of the time spent on various activities yielding utility to the consumer. In the present section, the base case where U is solely a function of market goods will be considered. In section 3.2.7 the model will be extended to include Q, whereas T will not be commented on any further.

Given the prices of X and a budget constraint M, the consumer is assumed to maximize his/her utility, expressed as:

$$\text{Maximise} \quad u = u(X) \tag{3.2}$$

$$\text{subject to} \quad \sum_i p_i \cdot x_i = M$$

The solution to the maximization problem gives the ordinary or the *Marshallian* demand function for the i'th good as a function of the price, P and the budget constraint, M:

$$x_i = x_i(P, M) \tag{3.3}$$

Where P is the vector of prices for the i'th good, $P = (p_1, p_2, \ldots, p_n)$. Substituting the ordinary demand function into the direct utility function gives the *indirect* utility function, which is a function of P and M:

$$u = v(P, M) \tag{3.4}$$

Using Roy's Identity, the ordinary demand function can be expressed by the derivatives of the indirect utility function:

$$x_i(P, M) = -\frac{\partial v / \partial p_i}{\partial v / \partial M} \tag{3.5}$$

Stating that the demand for x_i is a function of the change in utility caused by a change in price, relative to the change in utility caused by a change in income.

The dual to the utility maximising problem is that of minimisation of expenditure at a given level of utility (u^0):

$$\text{Minimise} \quad e = \sum_i p_i \cdot x_i \qquad (3.6)$$
$$\text{subject to} \quad u(X) = u^0$$

Where u^0 is the maximum utility achievable given the primal expenditure problem and e is the expenditure. The solution to the minimisation problem provides the expenditure function, which identifies the consumption bundle that achieves the targeted level of utility at minimum expenditure:

$$e = e(P, u^0) \qquad (3.7)$$

This expression is identical to the Hicks-compensated demand function. The Hicks-compensated demand function illustrates the change in demand for goods as a function of changes in prices holding utility constant (u^0). The compensated demand function can also be expressed by the derivative of the expenditure function in with respect to the different prices:

$$\partial e / \partial p_i = h_i(P, u^0) \qquad (3.8)$$

The Hicks-compensated demand function is not directly observable. The change in the compensated demand in response to a change in price can be expressed by the associated change of the Marshallian demand in response to a change in price. This relationship between the Marshallian- and the Hicks-compensated demand function is illustrated in the Slutsky equation:

$$\frac{\partial x_j(P,M)}{\partial p_i} = \frac{\partial h_j(P,u)}{\partial p_i} - \frac{\partial x_j(P,M)}{\partial M} x_i \qquad i, j = 1,\ldots,n \qquad (3.9)$$

The equation expresses how the demand for a good changes as prices change. The change in demand that can be observed on the market is be comprised by two elements; a s*ubstitution effect* and an *income effect*. The substitution effect specifies the change in demand caused by changes in the relative prices holding utility constant. The income effect specifies the change in demand caused by change in welfare, resulting from the change in price. The welfare economic implications of the differences in the Marshallian and Hicks demand functions will be discussed in section 3.2.

3.1.3 The Lancaster Approach to Economic Theory

In the presentation of neo-classical consumer theory, individuals were assumed to maximise utility with regards to bundles of goods *per se*. However, the focus in this project is on preference ordering with respect to the attributes or characteristics, of wood products, especially whether the product is certified or not.

Lancaster (1966) proposes that it is not goods *per se* but rather the bundle of characteristics that they consist of that give utility to the consumer. Consequently, demand for goods is derived from the demand for the characteristics that the good consists of (Lancaster, 1966; Rosen, 1974). According to the Lancaster Approach (LA), the transformation of goods into characteristics is accomplished by an objective "consumption technology" matrix, and subsequently, utility is defined as a function of the objective characteristics (Louviere *et al.*, 2000).

According to LA the same characteristic may be associated with several goods, just as each good may contain several different characteristics. Goods, which largely contain the same characteristics, are likely to be close substitutes whereas goods with no common characteristics are unlikely to be so (Lancaster, 1966).

It is assumed that characteristics are objectively assigned to goods. That is, it is assumed that all consumers perceive the characteristics of a given good the same. This implies that a common efficiency frontier representing the technologically determined efficient combinations of characteristics available at any given expenditure can be constructed independently of the individual consumer's preferences (Gravelle & Rees, 1992; Nicosia, 1974; Ratchford, 1975). Any rational consumer will choose to consume a combination of characteristics that is on the efficiency frontier; choosing otherwise would be irrational since it would imply that the same bundle of characteristics could be obtained at a lower cost elsewhere.

The choice to consume along the efficiency frontier is followed by a private choice. Here, consumers act on account of their individual and subjective, but still rational, preferences for different characteristics, and choose the combination that matches their private preferences. That is, the combination where the marginal utilities of the characteristics are equal to their marginal prices (Ratchford, 1975). Thus, the determinants of choices and the choice process in the LA are equivalent to those of traditional consumer theory. The only difference being that preferences, utility and prices relate to characteristics – i.e. attributes - of goods rather than to goods as such.

3.2 Welfare Measures

When an individual's current consumption possibilities are increased by a decrease in price or increase in quality, he/she experiences a welfare gain. In this context an example of the latter would be the possibility to purchase certified wood products. Similarly, when an individual is constrained from their present consumption patterns by an increase in price or decrease in quality, he/she experiences a welfare loss. Five different welfare measurements can be used to estimate such welfare changes. These measures will briefly be presented below and their relevance in relation to this study will be discussed.

3.2.1 Marshallian Consumer Surplus (M-CS)

The ordinary consumer surplus has its origin in the work by Alfred Marshall. The Marshall Consumer Surplus (M-CS) is defined as the area under the ordinary demand curve and above the horizontal price line (Marshall, 1920). Given a constant marginal utility of income the M-CS is proportional to the change in utility for any price change and can be used to estimate welfare changes. There are two problems though with the M-CS:

Marginal utility of income is seldom constant (Varian, 1992) or as stated by Marshal (1920); the marginal utility of goods is different between the rich and the poor (marginal utility of income). In that case M-CS will not correctly estimate welfare changes.

M-CS is sensitive to the simultaneous changes in prices also known as the *path dependency problem* (Johansson, 1987). Depending on the order of the price changes, different changes in surplus can be estimated. This is particularly evident when the multiple price change involves both increasing and decreasing prices (Johansson, 1991).

Due to these problems, the M-CS is not considered to be an appropriate welfare measure.

3.2.2 Hicksian Welfare Measures

In connection with the M-CS, the consumer adjusts the demand of the *i*'th good following a price change in relation to the original and constant budget constraint *M*. Referring to the Slutsky equation (3.9), the derivates of the Hicks-compensated demand function represents the pure substitution effect caused by a price change. That is, the consumer is kept on the original level of utility. Based on the Hicks compensated demand functions it thereby becomes possible to derive measures that will identify actual welfare changes. Depending on the ability to substitute between goods, two types of Hicksian welfare measures exist; the variation and surplus measure (Hicks, 1943).

3.2.3 Variation Measure

The variation measure is typically used to assess welfare changes in relation to price changes. Confronted with a price change, the individual is thus assumed to be able to substitute between goods, in order to equalise Marginal Rate of Substitution (MRS) to

relative prices, and thereby obtain the highest level of utility. Depending on the rights to the good the variation measure can be divided into *compensating variation* and *equivalent variation*.

In the Compensating Variation (CV) the individual is admitted the right to consume at the initial price of the good. For a price decrease of the good, the CV assesses the maximum amount the individual would be willing to pay to purchase at the new prices. Equivalently, in the case of a price increase, CV measures what lump sum payment the individual as minimum would require to be just as well off.

In Equivalent Variation (EV) the individual is admitted the right to the new price of the good. For a price decrease EV thus measures the minimum lump sum payment an individual will require to forgo the opportunity to purchase at the new lower prices, which he/she has the right to. Equivalently EV measures the maximum willingness to pay to avoid having to purchase at the new higher price.

3.2.4 Surplus Measure

The surplus measure is used when welfare changes result from quantity changes, where the good is available only in fixed quantities and therefore only can be consumed in fixed proportions. The surplus measure can also be divided into *compensating surplus* and *equivalent surplus*

The Compensating Surplus (CS) measure assumes that the individual has the right to the *initial* quantity of the good. For a quantity increase of the good, the CS specifies the maximum amount the individual would be willing to pay to obtain the new higher level of the good. Similarly CS measures what the consumer requires in compensation to accept a decrease in the quantity of the good to be just as well off.

In the Equivalent Surplus (ES) the individual has the right to the *new* quantity of the good. For a quantity increase ES states how much extra income, which is necessary to make the individual indifferent between obtaining the higher and settling for the lower initial quantity. Similarly ES measures what the consumer would be willing to pay to avoid a decrease in the quantity of the good.

3.3 Choice of Welfare Measure

As it appears from the previous sections the correct welfare measure to apply in a given situation depends on whether the change relates to prices or quantities, and on whether the individual has the right to the initial situation or the change.

3.3.1 Variation vs. Surplus Measure

As already mentioned, many environmental goods are only available in fixed proportions, exogenously determined to the individual. Compared to conventional wood products, certified wood products offer a higher, but *fixed*, level of environmental services and goods provided by forests, as discussed in chapter 2. The consumer is therefore prevented from choosing the exact level of services from forests, which maximises his/her utility.

On a theoretical level, the consumer can substitute between certified products and conventional products, in order to maximise utility, i.e. (Gravelle & Rees, 1992):

$$\frac{u_{cert}}{u_{conv}} = \frac{P_{cert}}{P_{conv}} = MRS_{conv,cert} \tag{3.10}$$

But in reality, it would only be possible if the consumer could adhere to one of the following strategies:

Strategy 1: The consumer chooses between various levels of certification, such as 10%, 50% or 100% certified wood products, or

Strategy 2: The consumer purchases different proportions of certified and conventional wood products, and thereby simulates the different percentages of certification presented above. If utility for example is maximised by a 10 % certification, the consumer purchases one certified product for every 9 conventional products.

On the market a given type of wood product is either certified or not, therefore strategy 1 cannot readily be applied. Consumers are therefore limited to substitute by purchasing different proportions of certified and conventional wood products. However, looking at one product type in isolation, this would probably only be a realistic strategy for a commonly purchased product, such as toilet paper[11]. Thus, for more lumpy goods, such as cutting boards and table tops, mixing between certified and conventional products does not appear to be a viable strategy. Consequently, the environmental benefits associated with certification in relation to these products are available only in fixed quantities. We therefore consider the most relevant measure to be a surplus measure.

[11] Pickering *et al.* (2001) explores preferences for eco-labelled fish products. In Pickering *et al.* (2001) it is assumed that people actually purchase proportional quantities of eco-labelled fish products and thus are able to substitute. As a consequence the welfare measure is a compensating variation measure.

3.3.2 Compensating vs. Equivalent

Whether the individual has a right to the higher level of the environmental good, q^1, or not, determines which of the surplus measures that should be used to estimate the welfare change associated with certification of wood products, and thereby indirectly with Sustainable Forest Management (SFM). In the current situation, legislation does not ensure the high level of sustainable management on a global scale associated with q^1. It is therefore concluded, that the point of origin is the present situation, with the present level of q^0. The correct welfare measure is thus a CS measure.

3.4 Derivation of the Compensating Surplus Measure

In the present section the basic model presented in section 3.1.2 will be extended to include two types of environmental goods; imposed environmental goods Q and voluntary environmental goods Z. The CS can be derived by the use of a utility function:

$$u = u(X, Q, Z) \tag{3.11}$$

Where X is the vector of private good quantities, $X=(x_1, x_2, ..., x_n)$, Q is a vector of imposed environmental goods, $Q=(q_1, q_2, ..., q_n)$, and Z is a vector of voluntary environmental goods, $Z = (z_1, z_2, ..., z_n)$, available to the individual. Z deviates from Q, since the consumer can decide whether or not he/she wants to consume the environmental good. However, it should be noted that Z, as in the case of certification, most likely only is available in fixed quantities[12].

Some of the imposed environmental goods are free (the air etc.) while others might have a price associated with the good. Let $R \geq 0$ be a price vector of Q, $T \geq 0$ be a price vector of Z and P a price vector of X. The present model is an extension of the model presented in Freeman (1993). The individual's utility maximising problem can be expressed as:

$$P \cdot X + R \cdot Q + T \cdot Z = M \tag{3.12}$$

where M is the budget constraint (income). Let $X + Z = C$. Thus, C ($C = c_1, c_2 ... c_n$) represents a vector of the goods, which the individual can decide whether or not he/she wants to consume. The fact that some imposed environmental goods are associated with a cost influences the form of the demand function. The conditional demand function for market and voluntary environmental goods is given by:

$$c_i = c_i(P, T, M - R \cdot Q, Q) \tag{3.13}$$

[12] If the consumer was free to choose the quantity of Z, and thereby able to substitute between goods to equalise MRS, then the relevant welfare measure would be a variation measure.

Where $R \cdot Q$ is the expenditure on imposed environmental goods with a price. This expenditure is conditioned on the level of imposed environmental goods $c_i(\ldots , Q)$.

Inserting the conditional demand function into the utility function gives the conditional indirect utility function:

$$v = v(P, T, M - R \cdot Q, Q) \tag{3.14}$$

Inverting the utility function for the term $M - R \cdot Q$ gives the conditional expenditure function. The conditional expenditure function identifies the minimum expenditure on market and voluntary environmental goods that, given P, T and Q, is required to yield a given level of utility u:

$$e^* = M - R \cdot Q = e^*(P, T, Q, u) \tag{3.15}$$

The dual to the maximization problem is the minimization of the expenditure subject to u^0. The solution to this problem is the restricted expenditure function:

$$e = e(P, T, R, Q, u^0) \tag{3.16}$$

The conditional and restricted expenditure functions are related as follows:

$$e = e^* + R \cdot Q \tag{3.17}$$

A change in q can affect the consumer's welfare in several ways. The change in q can be associated with a similar change in the price r, where r is the obligatory payment for the currently state of the environmental good q^0. In this case the consumer experiences a change in the level of q, but also in the lump sum payment $\Delta r \cdot \Delta q$. On the other hand, a change in q does not necessarily lead to an implicit change in price. Finally, the cost of providing q might increase independently of q. (Freeman, 1993)

As mentioned, the transition from Conventional Forest Management (CFM) to SFM is associated with an increase in the amount of environmental benefits produced by forests. This increase in environmental benefits may be perceived either as a change in q or in z. If the practice of SFM is made mandatory by law, the change is in q. On the other hand, if the transition is to be market-driven, such as is the case with certification, the voluntary change made available is in z. In the following, CS will be derived for both the case where the change is in q and where the change is in z.

3.4.1 CS for an Imposed Change

The transition from CFM to SFM is associated with a change in q from the initial level q^0 to the higher-level q^1. It is assumed that there are no costs associated with obtaining q^0, $r(q^0) = 0$. However it is assumed that, there will be costs associated with moving from q^0 to q^1. The compensating surplus (CS) with respect to a change in q can be expressed by the conditional indirect utility function, specifically:

$$v(P, T, M, q^0) = v(P, T, M - r\cdot(q^1-q^0) - CS, q^1) \tag{3.18}$$

The CS can also be defined using the restricted expenditure function:

$$CS = e(P, T, q^0, u^0) - e(P, T, r, q^1, u^0) \tag{3.19}$$

which, with reference to (3.16) describing the relationship between the conditional and restricted expenditure functions, is equivalent to:

$$CS = e^*(P, T, q^0, u^0) - e^*(P, T, q^1, u^0) - r\cdot(q^1 - q^0) \tag{3.20}$$

As it can be seen in (3.19), holding utility constant at u^0, CS is given by the difference in the conditional expenditure at q^0 and q^1, Δe^*, adjusted for the extra expenditure associated with the increase in q, $r(q^1 - q^0)$, see figure 3.1.

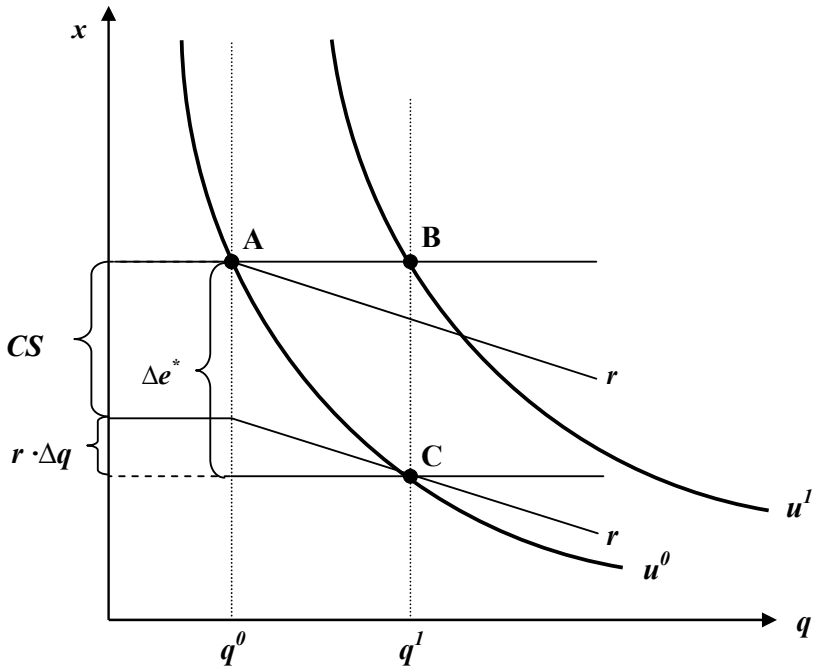


Figure 3.1 The compensating surplus.

In figure 3.1 the consumer is initially on u^0 at point **A** (q^0). This utility level is the one associated with CFM. Receiving the higher level of the environmental goods (q^1) associated with SFM, the consumer moves from **A** to **B**, on the new and higher level of utility u^1. CS is the individual maximum WTP for the change from CFM to SFM. This is illustrated by moving from point **B** to **C** minus $r \cdot \Delta q$.

3.4.2 CS for a Voluntary Change

In the voluntary case, the transition from CFM to SFM is associated with a change in z from the initial level z^0 to the higher-level z^1. That is the consumer can decide whether he/she wants to consume at z^0 or z^1. As in the imposed case, it is assumed that there are no costs associated with obtaining z^0, $t(z^0) = 0$. The CS with respect to a change in z can be expressed by the indirect utility function, more specifically:

$$v(P,\ T,\ M,\ z^0) = v(P,\ T,\ M - CS,\ z^1) \tag{3.21}$$

The CS can also be defined using the expenditure function:

$$CS = e(P,\ T,\ z^0,\ u^0) - e(P,\ T,\ z^1,\ u^0) = e^*(P,\ T,\ z^0,\ u^0) - e^*(P,\ T,\ z^1,\ u^0) \tag{3.22}$$

This is identical to the compensating surplus for a quantity constrained market good. Remembering that the change in q is identical to the change in z, and noting that $e = e^*$ due to the expenditure not being conditional of z, the CS for the voluntary change is equal to $\Delta \boldsymbol{e}^*$ in figure 3.1.

3.4.3 Comparing the Imposed and Voluntary CS

Comparing the CS measures obtained in (3.19) and (3.21) it is seen that the only difference is the adjustment in the imposed case for the extra expenditure associated with the increase in q, $r(q^1 - q^0)$. This leads to the conclusion, that the CS measures in the imposed and voluntary case are identical, if and only if $r = 0$. Seen from a welfare economic point of view, the CS associated with the special case of a voluntary public good is identical to the CS of a traditional imposed but free of charge public good.

In the remaining part of the study the environmental benefits associated with certification and thereby indirectly the transition from CFM to SFM will be denoted as q.

3.5 Complementarity

In a stated preference study, the assessment of the welfare economic impacts of a change in an environmental good, CS, can usually be derived directly as shown in section 3.4. However, in this study an estimate of CS is derived indirectly by respondents stating their preferences for certification, and thereby for SFM, through a complementary market good.

This means, that despite the fact that q^1 is voluntary, the consumer cannot go out and purchase q^1 directly. Thus, q^1 is only available as an attribute of wood products. This complementarity has implications both in relation to how an estimate of CS is obtained and also to how the estimates can be interpreted. More specifically it implies, that the CS measures derived in the previous sections, do not apply to our special case. In the following the theory of complementarity of goods will briefly be presented and CS in the case of complementarity will be derived.

3.5.1 Weak Complementarity

Complementarity between an environmental good – e.g. environmental quality – and a private good implies that the consumption of the private good depends on the on the level of environmental quality q. As an example, if individuals have preferences for the benefits associated with SFM, a FSC label increases the quality of the wood product and thus the overall probability that it is chosen. More specifically, if an increase in q increases the demand for the private good x_i, then there is *weak complementarity* between the demand for the private good and the environment. Weak complementarity requires that two conditions are satisfied (Freeman, 1993):

1. that x_i is non essential, why there exists a *choke price* $p^*_{x_i}$, where the compensated demand for x_i is zero[13], that is:

$$h_{x_i}\left(p^*_{x_i}, q, u^0\right) = 0 \tag{3.23}$$

2. that for values above the choke price the derivative of the expenditure function with respect to q is zero:

$$\frac{\partial e\left(p, p^*_{x_i}, q, u^0\right)}{\partial q} = 0 \tag{3.24}$$

The latter condition means that at or above $p^*_{x_i}$ the marginal WTP for q is zero. Changes in q therefore has no welfare economic significance, unless the price of x_i is low enough, so that the compensated demand for x_i is positive (Hanemann, 1995; Bockstael & McConnell, 1999).

Provided that these two conditions are met, CS can be derived by the use of the compensated demand curves for x_i at different levels of q. In this case, where the only

[13] In relation to discrete choice modelling, where each alternative is characterised by attributes, an attribute combination, which independently of price makes the alternative unattractive compared to the other alternatives, may have the same effect on the demand for q as a *choke price*.

possible levels of q are given by q^0 and q^1, the relevant compensated demand curves are given by:

$$h_{x_i}^*\left(q^0\right) = e\left(p_{x_i}^*, q^0, u^0\right) - e\left(p_{x_i}, q^0, u^0\right) \tag{3.25}$$

$$h_{x_i}^*\left(q^1\right) = e\left(p_{x_i}^*, q^1, u^0\right) - e\left(p_{x_i}, q^1, u^0\right) \tag{3.26}$$

Subsequently CS is defined as the area between the two compensated demand curves, that is:

$$\begin{aligned} CS &= e\left(p_{x_i}^*, q^1, u^0\right) - e\left(p_{x_i}, q^1, u^0\right) - e\left(p_{x_i}^*, q^0, u^0\right) + e\left(p_{x_i}, q^0, u^0\right) \\ &= e\left(p_{x_i}, q^0, u^0\right) - e\left(p_{x_i}, q^1, u^0\right) \end{aligned} \tag{3.27}$$

In the present study, weak complementarity arises due to certification being one among several attributes of a private good. This creates the foundation for deriving an estimate of the CS that consumers obtain from the benefits associated with SFM.

3.5.2 Weak Complementarity and Non-use Values

The derivation of the compensating surplus in the previous section is done without differentiating between Use Values (UV) and Non-Use Values (NUV). Traditionally, welfare measures based on observations of weak complementarity between market goods and environmental goods are expected to reflect only the UV of the environmental goods (Bockstael & McConnell, 1999). In this context, Hanemann (1995) explores the composition of the Hicksian welfare measure in connection to UV and NUV[14].

According to Hanemann (1995) the CS can be partitioned into a use, CS^{UV} and non-use component, CS^{NUV}, so that:

$$CS = CS^{UV} + CS^{NUV} \tag{3.28}$$

This particular definition of CS is derived from the following direct utility function

$$u(X, q) = T[u(x, q), q] \tag{3.29}$$

where X is a vector of private goods, and q is a public environmental good. In this model the marginal rate of substitution of Xs therefore solely depends on the level of q in $u(\cdot)$. If

[14] In this context non-use values refer to all values that do not entail contact between the individual and the forest.

$v(P, q, M)$ is the indirect utility function of $u(X, q)$ then CS for UV, NUV and UV+ NUV, respectively, is given by:

- $v(P, q^1, M\text{-}CS^{UV}) = v(P, q^0, M)$.
- $T[v(P, q^0, M - CS^{NUV}), q^1] = T[v(P, q^0, M), q^0]$ (3.30)
- $T[(P, q^1, M\text{-}CS^{UV}\text{-}CS^{NUV}), q^1] = T[v(P, q^1, M\text{-}C^{UV}), q^0]$.

The welfare change associated with NUV, CS^{NUV}, is determined independently of v. That is, it does not depend on the preferences for or use of private goods X. In contrast, CS^{UV} reflects that part of the preferences for q that is weakly complementary to the preferences for X. This suggests that the NUV of an environmental good cannot be derived through observations of relationships of weak complementarity between the environmental good and private goods.

Referring to chapter 2, it is primarily NUV that are associated with SFM, and thereby certification. Thus, the traditional distinction between UV and NUV suggests that the value derived from certification cannot be assessed by observing relationships of weak complementarity. It is in this context important to emphasise that, in the forests, the NUV are produced jointly with the marketable goods. Thus, there is a natural link between the NUV and wood products, implying that a relationship of complementarity is established between the NUV and the market goods produced by forests. This type of complementarity is quite different from the traditional way of perceiving complementarity. Usually the complementarity is only related to the consumption of the complementary good, and is as such not linked to the production of the good.

The resulting interaction between the market good (wood product) and the NUV is characterised by weak complementarity. As already mentioned, this may have implications in relation to how the obtained estimate of CS can be interpreted. With reference to the previous section, the assumption of weak complementarity implies that:

1. If the price of the complementary wood product exceeds the choke price of the good, the demand for the wood product, and thereby also the demand for NUV, is zero.

2. If other attributes, or attribute combinations, of the wood product independently of price makes the wood product unattractive it may have the same effect as a choke price. That is, the demand for NUV is zero.

3. Individuals can only express their preferences for the NUV ensured by certification of forests by purchasing a certified wood product.

4. Only the part of NUV which the individual consumer perceives as complementary to the wood product will affect the demand for the wood product.

All of the four implications have the potential to reduce the obtained CS estimate. That is, the obtained CS might underestimate the total true CS of SFM. More specifically, the first two implications point out that, even though an individual has preferences for the NUV associated with SFM, he/she will decline to buy a certified product if the price or attribute combination exceeds the choke level. The third implication means that individuals with no preferences for the wood product are prevented from expressing their preferences for the NUV. Finally the last implication refers to the fact, that individuals may not associate all NUV values with certification.

3.6 Free-Riding, Economic Theory and Willingness to Pay

In chapter 2 it was discussed how the goods and services associated with certification primarily are public goods. In this connection it was also discussed how traditional theory predicts, that consumers will free-ride, implying that consumers are expected to have a low or zero WTP for public goods, such as the ones associated with certification. Nevertheless, a significant number of empirical studies on individuals' propensity to contribute to the provision of public goods appear to reject the hypothesis of large-scale free-riding (Anderson & Holt, 2002). In relation to eco-labels, several studies all based on market data have found that consumers have a WTP for the public goods associated with the eco-labels (Bjørner *et al.*, 2002; Andersen, 2002b; Wier *et al.*, 2002; Roe *et al.*, 2001)[15].

The results of these studies suggest that the choice of the consumer is determined by other factors, which to a large extent outweigh the incentives to free-ride. The factors that are considered most relevant to the present study are presented in the following sections.

3.6.1 Citizen vs. Consumer

One possible explanation for the absence of free-riding could be that people hold different preferences for private and public goods.

In this context Sagoff (1988) makes the distinctions between the dual role of the individual as a citizen and as a consumer. Sagoff (1988) defines the concept of *citizen* and *consumer* as follows:

[15] The results of some of these studies will be presented in chapter 11.

- As a consumer the individual makes choices based on the egoistic motives and acts in accordance with economic theory.

- As a citizen, the individual makes choices based on motives related to what is in best interest to society and is thus guided by an "ethical rationality".

Several studies have been conducted to verify the dual role of the individual as proposed by Sagoff (1988). In this connection both Stevens *et al.* (1994) and Blamey *et al.* (1995) identify WTP functions for public goods (existence values for wild life protection and preservation values for forests), which entail variables representing citizen characteristics, such as; fairness and attitudes towards society's prioritisation of resources[16]. The results from these studies suggest that individuals might be guided by citizen preferences when valuing public goods.

Curtis & McConnell (2002) also examine the properties of citizen preferences.
The study focuses on individuals' willingness to contribute to wild life management by using a tax as payment vehicle. They propose, and find support of, the hypothesis, that a sharp distinction between consumer and citizen preferences cannot be made.
Moreover Curtis & McConnell (2002) emphasise that such a distinction may especially be problematic in relation to eco-labelled products where individuals are forced to combine their preferences for private and public goods. In this relation they argue that altruistic preferences may serve as a better explanation of why otherwise economically rational individuals make choices that appear to be more beneficial to the society than to themselves.

3.6.2 Altruism

As discussed above the term citizen preferences might not be relevant in relation to explaining consumers' preferences for public goods associated with eco-labels.
Thus as pointed out by Curtis & McConnell (2002) and supported by Andreoni (1989; 1990), altruistic preferences are likely to provide a better explanation of consumers' WTP for public goods associated with eco-labels.

Altruism refers to the fact that individuals may attach a positive value to payoffs accruing to others (Becker, 1981). Within the field of environmental economics and especially economic valuation of environmental goods, altruism is a well-known phenomenon. Johansson (1992) states that as much as 50-70% of the WTP for NUV may be motivated by altruistic behaviour.

[16] Blamey *et al.* (1995) actually found that *citizen variables* had greater explanatory power than *consumer variables*.

However, altruistic motives have been accused by several economists for not being valid motives in economic valuation of the environment due to the problem of double counting of benefits in cost-benefit analysis (Milgrom, 1993). Similarly, Kahneman & Knetsch (1992) argue that altruistic motives give rise to values based on moral satisfaction, and does not reflect the inherent value of the public good. In contrast, Carson *et al.* (2001:177) argue that: "*The answer from economic theory is clear: it is the utility whatever its source that matters for total value. Motives are essentially irrelevant from the perspective of economic theory...* "

Referring to the findings of Curtis & McConnell (2002), we believe that altruistic preferences might play an important role in relation to explaining individuals' WTP for certified products. Agreeing with the arguments advanced by Carson *et al.* (2001), it is believed that such preferences do not impinge on the validity of the study.

3.6.3 Types of Altruism

Depending on the specification of the altruistic preference structure, different types of altruism can be defined (Andreoni, 1989; Johansson-Stenman,1997; Becker, 1981). In the following sections, *pure* and *impure altruism*, which are considered to be the most relevant in relation to the present study will be explored and discussed.

In the following U_i is a vector of the utility of other individuals, z_i is the contribution that each individual adds to the total externality, $Z = \sum z_i$. In our case Z[17] is a positive externality (SFM) influencing U_i, and x_i is the consumption of other goods by individual i. The individual utility is labelled u_i.

3.6.3.1 Pure Altruism

Pure altruism refers to the case, where one individual's utility level is dependent on other individuals' utility levels, and accordingly that individual's maximisation of own utility cannot be seen in isolation from other individuals' wellbeing (Becker, 1981).

The utility function of a pure altruist is given by:

$$u_i = u(x_i, z_i, Z, U_i) \tag{3.31}$$

where the utility of others enters as a direct argument in the individual utility function. Thus the individual maximises his/her utility with subject to market goods, his/her contribution to the externality, the total level of the externality and the utility of other people. It is though difficult to imagine, that the exact utility of others, U_i, can be identified

[17] Since N (the number of individuals) is assumed to be large and the individual contribution to Z (z_i) is small, Z is written without any individual index.

and subsequently directly enter the individual's utility u_i. As such pure altruism is an extremely special case (Andreoni, 1990) and it is not considered relevant in connection to this study.

3.6.3.2 Impure Altruism

Impure altruism refers to the case, where the individual gains utility by the act of doing something for others, independent of the actual utility experienced by other people. This type of altruism is equivalent with the concept of *warm glow of giving*[18]. The utility function for an impure altruist is given by:

$$u_i = u(x_i, z_i, Z) \tag{3.32}$$

where z_i affects the utility both directly (contribution to Z) and indirectly through the individual's impure altruistic concern. The latter refers to the fact, that the individual derives utility from the mere act of contributing.

The impure altruistic utility can be based on expectations concerning the impacts of the contribution on U_i as in the case of pure altruism. But the utility could equally be determined by the social norms in the society as in the citizen model by Sagoff (1988). In the latter case, the individual does not as such care for the environment, but purely responds to the effect on his/her own conscience.

Another reason for stating a positive WTP related to impure altruism could be a simplification of the choice process in relation to choosing between a certified and conventional wood product. Retrieving information on forest management and certification is cumbersome and it is even more difficult to form the information into a choice. In that case the choice to choose a certified wood product might become an expression of "it is better doing something than nothing".

3.7 Compensating Surplus, Donations and Altruism

In chapter 2, it was discussed, that individuals are likely to perceive the WTP for certification as a more general contribution to the promotion of SFM, rather than as payment for a specific environmental improvement. The WTP for certification will therefore probably bear resemblance to donations. Similarly, in the previous section, it was discussed how impure altruistic motives are likely to be an important cause of positive WTP. Based on these considerations, it is relevant to look at the extent to which individuals' impure altruistic voluntary donations to a public good are likely to reflect the true CS of the good.

[18]Warm glow of giving is used when individuals yield utility by the act of giving (Andreoni, 1989; Kahneman & Knetsch, 1992).

3.7.1 Deriving CS in the Case of Donations

In order to derive CS in the case where individuals receive utility both from the environmental good and from the act of giving, the utility maximization problem is given by (Champ *et al.*, 1997):

$$\text{Maximise} \quad U_i = U_i(X_i, \Gamma, \gamma_i, Q, q_i) \tag{3.33}$$

$$\text{subject to} \quad P \cdot X_i + u \cdot \gamma_i + q_i = Y_i$$

Where X_i is a vector of market goods, Γ is a vector of total contributions by all members of society for the provision of all public goods other than the ones associated with SFM, γ_i is a vector of donations for these public goods by consumer i, Q is the total donation by all members of society for the public goods associated with SFM, q_i is the donation of the i'th consumer for the public goods associated with SFM, Y_i is income, P is the price vector for market goods and u is a unit vector conformable to γ_i. Let Γ_{-i} and Q_{-i} represent the total donations less those of the i'th individual, and note that they are exogenously determined.

Referring to the impure altruism model in section 3.6.3.2, donating money to public goods bring utility to the impure altruist in two ways; 1) directly by the increase in Γ and Q and 2) indirectly through the "warm glow" of donating γ_i and q_i. Subsequently the indirect utility function is given by:

$$V_i(P, Y_i, \Gamma_{-i}, Q_{-i} | \gamma_i, q_i) \tag{3.34}$$

Let X_i^*, γ_i^* and q_i^* equal the i'th consumer's; optimal private good bundle, other donations bundle and the donation to the public good associated with SFM. Finally, let γ_i^{**} be equal to the optimal donation to other public goods, given that $Q_{-i} = q_i = 0$. The i'th consumer's CS_i associated with the transition from CFM to SFM can then be defined by:

$$V_i(P, Y_i - CS_i, \Gamma_{-i}, Q_{-i} | \gamma_i^*, q_i^*) = V_i(P, Y_i, \Gamma_{-i}, 0 | \gamma_i^{**}, 0) \tag{3.35}$$

In (3.35) the maximum WTP for public goods associated with SFM of the i'th consumer is: $CS_i + q_i$. However, it is not necessarily the case that the consumer's maximum Willingness to Donate (WTD) is identical to the maximum WTP. To illustrate this, let $\bar{\gamma}_i$ equal an optimal vector of other donations in the situation where the consumer is constrained to set $q_i = 0$, while others are allowed to donate, implying that $Q = Q_{-i}$, and let d_i equate the utility with and without the opportunity to personally donate:

$$V_i(P, Y_i - d_i, \Gamma_{-i}, Q_{-i} | \gamma_i^*, q_i^*) = V_i(P, Y_i, \Gamma_{-i}, Q_i | \bar{\gamma}_i, 0) \tag{3.36}$$

With reference to (3.36) the WTD to the provision of the public goods associated with SFM of the i'th consumer is then given by: $d_i + q_i^*$, where d_i represent the extra utility gained from the increase in public goods caused by the infra-marginal amount donated. Making the reasonable assumption that utility is non-decreasing in Q, then $d_i \leq \mathrm{CS}_i$ implying that an individual's WTD is likely the be less than his/her WTP.

3.8 Compensating Surplus and WTP for Certification

In economic valuation the values estimated for an environmental change normally represent the value of the welfare measure. Traditionally, respondents either express an annual WTP or a total WTP for the environmental change, which afterwards are aggregated over the total number of individuals to estimate the total welfare economic value of the change (Hanley & Spash, 1993):

$$Aggregate\ WTP = N \cdot \sum_{i=1}^{n} w_i \cdot WTP_i \tag{3.37}$$

where N is the total number of individuals in the population, w_i is the weight that is associated with each observation, i, of WTP. In this study the respondents express their WTP for certification of forests by stating the extent to which they are willing to pay a price premium for a certified wood product compared to a conventional product. This kind of WTP cannot readily be transformed into an aggregated estimation of the total WTP.

In theory, the aggregated WTP and subsequently the welfare economic value associated with certification can be defined as follows:

$$CS = \sum_{i=1}^{N} \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{WTP_{ic}}{(1+r)^t} \tag{3.38}$$

where N is the number of people in the population, C is the total number of certified products, T is the lifespan of individual i, c represents the purchase of a certified wood product, t is the time of the event of c and WTP_{ic} is the WTP for product c of individual i . As a result, data on the total purchase of certified wood products – i.e. the WTP for each product by each individual and the individual's frequency of purchase during a lifespan – would be needed to estimate the true CS of the Danish population associated with certification.

In this study only the WTP for three certified wood products are investigated, why an estimation of CS as presented in (3.38) is not possible. The WTP estimates in this study

therefore only measure the fraction of the individual's total CS of certification which is associated with one purchase of each of the three products included in the study.

In this connection, it is also important to emphasise, that the estimated CS of certification does not necessarily encompass the total CS associated with the transition from CFM to SFM. With reference to section 3.5.2 the assumption of weak complementarity, imply that obtained estimates of CS for certification are likely to be an underestimate of the CS associated with the transition from CFM to SFM. Moreover, with reference to section 3.7, where it was argued that the WTP for certification may be perceived similar to a donation, this also suggest that the CS estimated in this study will only be a lower bound estimate of the true CS.

3.9 Summary

Following a presentation of the economic theory of the consumer and different welfare measures, the Compensating Surplus (CS) was identified to be the relevant welfare measure in this study. The benefits associated with certification are found to qualify as a voluntary environmental good. Subsequently, the CS for this good is derived from the relationship of weak complementarity between certification and wood products. Traditionally, such a CS is only expected to reflect the use values of a good. In this case, however, the obtained CS is expected also to include non-use values. It is also pointed out that the weak complementarity between the benefits of certification and wood products has implications on CS. Thus, the CS associated with certification will probably underestimate the true CS associated with the transition from conventional to sustainable forest management.

Different types of altruism, which may contribute to CS, are explored and the impure altruistic model is identified to be relevant as an explanation of positive WTP in this study. Subsequently, CS is derived from a donation model based on the presence of impure altruistic preferences. In this connection it was found that a CS based on donations is likely to be a lower bound estimate of the true CS. Lastly, it was argued that the obtained CS only represents the welfare change associated with one purchase of each of the three certified wood products.

4 Environmental Valuation Methods

In the previous chapters it was discussed how consumers, motivated by preferences for the environmental benefits associated with Sustainable Forest Management (SFM), are expected to be willing to pay a premium for certified products. In this chapter, the valuation methods available for eliciting the Willingness to Pay (WTP) for certification will be presented, and the background for choosing the Contingent Ranking Method will be discussed.

4.1 Introduction to Environmental Valuation

Economic valuation of the environment was introduced in the 1940's by Ciriacy-Wantrup (1947) and Hotelling (Mitchell & Carson, 1989; Hanemann 1994). The proposed *valuation methods* – the Contingent Valuation Method (CVM) and the Travel Cost Method (TCM), respectively – distinguished themselves from the previously used *pricing methods* by being based on a latent demand, i.e. consumer preferences and behaviour, rather than on market prices (Hanley & Spash, 1993; Hanemann, 1995). This difference between valuation methods and pricing methods is relevant in relation to assessing the welfare economic value of environmental resources or services. Thus, since pricing methods are based on actual market prices rather than preferences, they fail to account for the consumer surplus. This implies that pricing methods do not capture the true welfare economic value of the good or service in question.

Significant progress has occurred within the field of environmental valuation during the past decades, and a variety of methods – each with their advantages and disadvantages – are now available (Garrod & Willis, 1999). In this chapter, the different valuation methods will be presented. The focus of the review will be on the methods relevant in relation to identifying individuals' WTP for certified products.

4.2 Valuation Methods

Broadly speaking, valuation methods can be divided into two groups – *Stated Preference Methods* (SPM) and *Revealed Preference Methods* (RPM). Within these distinct groups, the methods can be sub-divided into *direct methods* and *indirect methods*. This creates four distinct categories of valuation methods (Mitchell & Carson, 1989). These are shown in table 4.1, where examples of specific methods belonging to each category are also given.

Table 4.1 Valuation Methods (Adapted from Freeman, 1993).

	Revealed Preference Methods	**Stated Preference Methods**
Direct Methods	Competitive market prices	Contingent Valuation Method (CVM)
Indirect Methods	Travel Cost Method (TCM) Hedonic Price Method (HPM) Discrete Choice Modelling, of panel data (DCM)	Contingent Valuation Method (CVM) Choice Modelling Methods (CMM)

The distinction between RPM and SPM is made on account of the origin of the data. RPM are based on data derived from observed (i.e. actual) behaviour, and SPM are based on data derived from stated (i.e. hypothetical) behaviour (Freeman, 1993; Mitchell & Carson, 1989). The distinction between direct and indirect methods, on the other hand, is based on the nature of the obtained data. Direct methods, are methods, which yield monetary values directly, and indirect methods, are methods where monetary values are deducted indirectly from the data by using models of individual behaviour and choice (Freeman, 1993). As it can be seen in table 4.1 CVM can, depending on the format used, give both direct and indirect estimates for the environmental good, for a more detailed explanation see 4.3.1.

4.2.1 Revealed or Stated Preference Methods

SPM are often subjected to critique due to the hypothetical basis of the methods. It is often highlighted that there in many cases is likely to be a gap between stated intentions and actual behaviour (Cummings *et al.*, 1995; List & Gallet, 2001). RPM, on the other hand do not suffer from this *hypothetical* bias. Accordingly, it is often suggested that they are more likely to produce value estimates that reflect the true value of the good, or service, subjected to valuation (Louviere *et al.*, 2000). However, as the RPM are founded on the observed choices of individuals on the market, either of two requirements needs to be fulfilled in order for them to be used (Bockstael & McConnell, 1999):

- There exists a market for the environmental good. In this case a direct method can be used, and the valuation is fairly straightforward, or

- There exists a market for a complementary private good.

As mentioned in chapter 3, complementarity between an environmental good – e.g. environmental quality – and a private good implies that the consumption of the private good depends on the level of environmental quality, q (Hanemann, 1995).

The fact, that the use of RPM is contingent upon the availability of data describing actual behaviour, renders the methods irrelevant in relation to many situations involving environmental goods, for which markets often are non-existing. Likewise, they are often quoted not to be able to measure Non-Use Values (NUV) (Hanemann, 1995; Bockstael &

McConnell, 1999). However, Freeman (1993:158) addresses the issue somewhat differently: "*Unless there is a market for preservation of a resource or for enhancing its quality for non-users, there will be no transactions to reflect the non-use values of the individuals, and only direct methods of estimating non-use values will be feasible*". As such, it is not that RPM are unable to uncover NUV. It just means that they can only be used to estimate NUV for which a market exists or NUV that is complementary to market goods. An example of the latter might be certified products, where the labelling of products establishes a connection between the consumption of the private good and the NUV associated with SFM.

In relation to the present study, where the focus is on valuing an attribute of a marketed good, the use of a RPM seems to be relevant. However, FSC labelled products have only been on the Danish market since 1999, and the number of products is limited (Kristensen, 2003). The level of information available on actual purchases/behaviour is therefore insufficient to use a RPM. Consequently, the present study will be based on the use of a SPM exploring the hypothetical behaviour.

4.3 Stated Preference Methods

Within the class of SPM, the Contingent Valuation Method (CVM) is the most widely used (Batemann & Willis, 1999). Choice Modelling Methods (CMM), however, has become increasingly popular during the recent years (Bennett & Blamey, 2001). The SPM has the potential to measure both Use Values (UV) and NUV.

4.3.1 The Contingent Valuation Method (CVM)

In CVM the aggregate value of an environmental change or good is estimated holistically, by presenting the individual with a precise description/scenario of the hypothetical good – e.g. change in the environment. Based on information regarding the rules of provision, present and future access to the good and method of payment, the individual is asked to state his/her valuation of the good. Within the CVM frame, several mechanisms have been developed to elicit the individual's preferences, see table 4.2 (Hanley *et al.*, 1997; Mitchell & Carson, 1989).

Table 4.2 Typology of CVM elicitation mechanisms.

	Actual WTP amount obtained	**Discrete indicator of WTP obtained**
Single question	Open-ended, Payment card and Sealed bid	Dichotomous choice, Spending question offer, Payment ladder and Interval checklist
Iterated series of questions	Bidding game and Oral auctions	Double bounded dichotomous choice, Polychotomous choice

Table 4.2 illustrates that the WTP for the good can either be obtained directly in terms of a monetary unit (actual WTP obtained) or indirectly (discrete indicator of WTP obtained).

As an example, the approach of CVM in relation to assessing WTP for certified products could be to present respondents with the hypothetical situation where they had the possibility to buy either a certified or a conventional product. Depending on the format, the respondents would then be asked to either:

- State the maximum price premium that they would be willing to pay to get the certified, as opposed to the conventional, product (actual WTP obtained); or

- Accept or reject to buy the certified product at a given price premium (discrete indicator of WTP).

4.3.2 Choice Modelling Methods[19] (CMM)

CMM have their background in the paper on *conjoint measuremen*[t] by Luce & Turkey (1964) who established the theoretical foundation of the CMM used today. During the 1970s, the first attempts to model consumer preferences with CMM, at that time known as *conjoint analysis*[20], were carried out within the fields of marketing and transportation economics. The aims were to understand the structure of - and to elicit people's preferences for various goods (Green & Rao, 1971; Green & Srinivasan, 1978).

CMM build on the theory proposed by Lancaster (1966), which was presented in section 3.1.3. Accordingly, CMM focus on how preferences for goods or services are constructed, the goal being to identify the utility that individuals derive from the different attributes, which compose the good or service in question (Bennett & Adamowicz, 2001). This is accomplished by presenting respondents with a set of alternatives (A_i). The alternatives define the good or service in terms of their key attributes (a_{ij}), and different alternatives are described by varying levels of the attributes – see table 4.3 (Garrod & Willis, 1999).

[19] In the literature the different CMM are categorised differently. In this study the categorisation in Hanley *et al.* (2001) is used. For others see Adamowicz *et al.* (1999)

[20] The term conjoint analysis was invented by Green & Srinivasan (1978:103), and was defined as: "*models and techniques that emphasize the transformation of subjective responses into estimated parameters*"

Depending on the more specific nature of the applied method, respondents are then asked to either choose the alternative they prefer, rank or rate the alternatives (Louviere *et al.*, 2000; Hanley *et al.*, 2001). Thus, the methods have a high degree of resemblance to real market situations, where consumers are used to evaluating several products simultaneously, and subsequently to choose between them (Rolfe & Bennett, 2000; Adamowicz *et al.*, 1999). By examining the trade-offs between attributes/attribute levels, that are implicit in the choices made by respondents, it is possible to derive an estimate of the utility associated with the different attributes (Garrod & Willis, 1999). If one of the attributes is measured in monetary units (i.e. price), it is possible to derive estimates of respondents' WTP for the other attributes from the marginal rate of substitution between the monetary attribute and the other attributes (Louviere *et al.*, 2000).

As such the CMM are somewhat similar to the CVM based on a discrete indicator method presented in table 4.2. In CMM the respondent chooses among different alternatives (A_i), with different levels of the attributes (a_{ij}) describing the environmental change, and different prices. In CV the respondents chooses between two alternatives (A_i) with only two attributes: a price and an environmental change, see table 4.3.

Table 4.3 Comparison of CVM (discrete indicator of WTP) and CMM.

Contingent Valuation Method		**Choice Modelling Methods**		
A_1	A_2	A_1	A_2	A_3
Status Quo	Environmental Change	a_{11}	a_{21}	a_{31}
0 DKK	XX DKK	a_{12}	a_{22}	a_{32}
		a_{13}	a_{23}	a_{33}
		ZZ DKK.	YY DKK.	WW DKK.

4.4 CVM or CMM

From the previous sections it is evident that, CVM and CMM frame the choice situation differently. This difference influences the kind of values that the methods can estimate, and may also have an effect on the magnitude of the estimated values. In relation to the purpose of the present study we believe that the direct nature of the CVM may be a serious disadvantage. It creates a situation that does not resemble ordinary purchase situations, where all attributes of a product are taken into account and valued more or less simultaneously. More specifically, it is believed that if respondents only have to make a trade-off between price and certification, they may be inclined to attach undue weight to certification, which would result in an overestimation of the true WTP for certification. In this context, we consider the high degree of resemblance to real market situations of the CMM elicitation mechanism to be an important advantage. Furthermore, it is believed that the difference in elicitation mechanism makes CMM less susceptible to some of the biases for which CVM has been criticised, see chapter 5.

Another advantage of CMM over CVM is that it is possible to assess the relative weight associated with different product attributes rather than just assessing the WTP for certification in isolation (Bennett & Adamowicz, 2001; Hanley *et al.*, 2001). That is, by using a CMM rather than CVM it may be possible to get a more nuanced perspective on how individuals value certification. Thus, apart from getting an estimate of the trade-offs that individuals are willing to make between certification and money, it is also possible to disclose how individuals make trade offs between environmental considerations and more tangible considerations such as product quality and appearance (Adamowicz *et al.*, 1998a; 1998b; McFadden, 1986).

Based on these observations, we find that CMM is the most appropriate methodology to adopt in the present context.

4.5 Different Choice Modelling Methods

The term CMM refers to three specific methods: choice experiments, contingent ranking and contingent rating. Similar for these methods is that value estimates are derived from observations of individuals' choices between alternatives that are characterised by attributes. Thus, for all methods the identification of key attributes, design of alternatives and choice set play an important role in relation to the validity of the method. However, the three methods vary significantly in terms of the way that preferences are expressed, the nature of the obtained data and the amount of extractable information per respondent and the quality of the data (Hanley *et al.*, 2001; Louviere *et al.*, 2000). The three methods are briefly described below.

4.5.1 Choice Experiments (CE)

The CE approach is the most simple of CMM, and it appears to be both the most applied and widely recognised of the CMM (Hanley *et al.*, 2001; Adamowicz *et al.*, 1998a; 1998b; Louviere *et al.*, 2000). In CE the respondent is presented with a choice set containing two or more alternatives, preferably including a status quo option, and is asked to choose the alternative most preferred (Bennett & Adamowicz, 2001). Implicit in this choice is the observation, that the utility, which the respondent associate with the most preferred alternative is higher than the utility, which the respondent associate with the other alternative(s) in the choice set. Other things equal, the amount of information extractable from a single CE observation is less than the amount of information that can be extracted from the other CMM. However, an advantage of CE is, that the task, which respondents are asked to perform is relatively simple and the cognitive burden low. Furthermore, when dealing with marketed goods, CE bear very close resemblance to the choices that the respondents are used to make in the market place (Louviere *et al.*, 2000). Thus, intuitively the task is likely to make sense to respondents.

4.5.2 Contingent Ranking Method (CRM)

In CRM, the respondent is presented with a choice set consisting of three or more alternatives, but rather than being asked to just choose the most preferred alternative, the respondent is asked to rank the alternatives from the most preferred to the least preferred (Louviere *et al.*, 2000). The rank ordering of alternatives can be seen as a sequence of choices, which expresses the relative utility that the respondent associates with the alternatives contained in the choice set. This implies that CRM provides the analyst with much more information on the preference structure of the respondents than CE does[21] (Hanley *et al.*, 2001; Holmes & Boyle, 2003; McFadden, 1986). Parallel to the increase in information that can be extracted from a CRM compared to a CE is an increase in the cognitive burden that is placed on the respondents (Ben-Akiva *et al.*, 1991; Foster & Mourato 2002; Hausmann & Ruud, 1987). Thus, it is likely that respondents will find it more difficult and strenuous to provide a complete ranking of the alternatives than to just identify the most preferred. This increased task complexity is suggested to affect the reliability of CRM data (Louviere *et al.*, 2000). Thus, one potential, and often mentioned, consequence is, that CRM data may display inconsistency of preferences across ranks. This inconsistency is suggested to be caused by respondents changing decision protocols across ranks, why it is suspected that the complexity may induce people to adopt heuristic decision rules. Another potential consequence is that respondents may be less attentive, or more uncertain, when assigning ranks to less preferred alternatives than when assigning ranks to the most preferred alternative. Subsequently, the observations related to lower ranks will be associated with greater variances than higher ranks (Hausmann & Ruud, 1987; Chapman & Staelin, 1982).

The extent to which the advantages of the CRM associated with the increased amount of information are greater than the disadvantages arising from increased task complexity depends on the specific context. The specific context includes the number of alternatives in each choice set, the number of attributes and the number of levels of each attribute.

4.5.3 Contingent Rating (CR)

In CR, the respondent is presented with a choice set consisting of a number of alternatives, which he/she is asked to rate independently, from a predefined scale (Louviere, 1988). The rating approach to CMM is the one that has the potential to provide the greatest amount of information on respondents' preferences. Thus, in addition to providing information on the relative ordering of alternatives – as with CRM – the rating approach also provides information on the intensities of the ordering (Mackenzie, 1993). In practice, however, the cardinal properties of the individual ratings imply that it may be difficult to make use of this increased amount of information. Thus, the way that individuals use the ratings scale

[21] The difference in the amount of information that can be extracted from CE and CRM respectively is discussed in section 7.4.

may vary significantly across individuals. Where some individuals use the entire scale presented, others may only use specific proportions of the scale (Mackenzie, 1993). More specifically, this renders aggregation of ratings across individuals problematic (McFadden, 1986). In order to reduce miss-specifications and biases in relation to the proper interpretation of ratings across individuals Louviere *et al.* (2000) suggest a transformation of the rating into a ranking. But this is only advisable if the rating does not contain tied preferences (Boyle *et al.*, 2001). By following this recommendation, however, the additional information inherent in the rating compared to a ranking is lost, suggesting that CRM might just as well be used in the first place.

4.6 Choice Experiments, Contingent Ranking or Contingent Rating

Based on the discussion of the characteristics, advantages and disadvantages of the different CMM presented in the previous section, we have chosen to use the CRM method in the present project. The argumentation for our choice of the CRM is presented below.

4.6.1 Contingent Rating vs. Contingent Ranking

Comparing CR and the CRM, the rating method may be seen as having one advantage over CRM. By rating alternatives identically, respondents are given the opportunity to express indifference between two or more alternatives. This is not an option made available in the CRM, implying that the method may force respondents to make choices between alternatives that they in effect are indifferent towards (Mackenzie, 1993; Morrison & Boyle, 2001). Consequently, the information derived from such choice observations will not reflect the actual utility associated with the alternatives. However, if respondents by their own initiative choose to express indifference by assigning the same rank to more than one alternative, this can actually be accommodated for in the model, which will be used in the present study. Thus, considering that ratings would have to be transformed into ranks in order to overcome the cardinality problem associated with ratings, we believe that the CRM is more appropriate than CR.

4.6.2 Choice Experiment vs. Contingent Ranking

The choice of CRM over CE is less straightforward than the choice of CRM over CR. Thus, the simplicity and close resemblance to real market choices of CE immediately makes this the most intuitively appealing of the CMM. However, with reference to the purpose of the present study, where one of the objectives is to investigate to what extent/and how WTP varies across three product types, CRM appears more appropriate. More specifically, due to the increased amount of information inherent in a CRM observation it was judged that it would be sufficient to present one choice set (for each product type) to each respondent if using CRM. If CE was chosen several choice sets for each product type would have been required. Using CE to valuate three product types was considered to take up too much space, making the questionnaire too long, considering the

limited amount of time the respondents can be expected to spend on answering. The use of CE was believed to increase the risk of respondents feeling discouraged from completing it just by looking at the number of pages comprising it. Considering that the alternatives to be ranked in the present study are fairly familiar market goods, the task that respondents are faced with, is considered less difficult than if the goods in question had been more complex and not commonly traded goods.

4.7 Summary

In this chapter different valuation methods available for eliciting WTP for certification are presented. Due to the limited availability of market data, the use of revealed preference methods is not considered relevant. Of the stated preference methods, the contingent valuation method is considered inappropriate due to lack of resemblance to actual market situations. Of the choice modelling methods, the contingent ranking method is subsequently chosen, since it makes it possible to retrieve more information from each respondent than had choice experiments been chosen.

5 Potential Biases, Validity and Reliability

A number of biases may be associated with using the Contingent Ranking Method (CRM). The presence of such biases may impinge on the reliability and validity of the obtained results. Fortunately, most biases may be avoided by careful design of the survey instrument (Pearce *et al.*, 2002). The first part of the present chapter contains discussions of the biases that are commonly advanced in relation to Stated Preference Methods (SPM). Subsequently, the different validity criteria, against which the survey results will be evaluated, will be introduced.

5.1 Potential Biases Associated with the CRM

The presence of biases introduces systematic errors into the results of a study. More specifically, bias is defined as being present if the sample mean stated WTP is found to be systematically different from what it would be in an identical actual market (Green & Tunstall, 1999; Andersen, 2002a)

There are several different potential sources of bias, the relevance and importance of which, depend on the specific nature of a given study. Table 5.1 provides an overview of the biases that will be discussed in the chapter[22].

Table 5.1 Potential biases associated with using SPM (adapted from Mitchell & Carson, 1989).

Incentives to Misrepresent Responses	Strategic bias Yea-saying
Implied Value Cues	Starting point bias / Range bias Importance bias
Scenario Misspecification	Amenity misspecification bias Symbolic bias Part-whole bias Probability of provision bias Context misspecification bias Framing effect biases Ordering bias
Hypothetical Bias	

In the following sections the more specific nature of the different biases, along with ways to avoid them, will be discussed. Potential differences between CVM and CRM in relation to their susceptibility to the different biases will also be discussed.

[22] It should be noted that the terminology and classification of biases is quite diverse. The ones adopted in this study draws upon several studies and have been adapted to fit the present context.

5.2 Incentives to Misrepresent Responses

One source of bias may be that respondents for some reason feel inclined to misrepresent their true preferences.

5.2.1 Strategic Bias

Strategic bias is often mentioned as an important potential bias in relation to CVM, and it refers to situations where a respondent deliberately state an untrue WTP due to strategic motives (Garrod & Willis, 1999). Such motives may be a desire to affect the provision of the good – e.g. stating a high WTP may be seen as a way of increasing the probability that the good will be provided. Empirical evidence, however, points toward weak strategic behaviour in CVM (Morrison *et al.*, 1996). According to Adamowicz *et al.* (1999) the risk of strategic behaviour is even lower in CMM than in CVM. Thus, the presence of several multi-attribute alternatives implies that a "strategic best choice" is much more difficult for the respondent to identify.

In relation to the present study, strategic bias is not considered to be a problem. The focus of the survey and the nature of the specified scenario do not refer to any policy proposals or other initiatives, which public institutions could be likely to consider. Thus, it is probably not a context that invites respondents to exhibit strategic behaviour.

5.2.2 Yea-saying

Yea-saying refers to the situation where respondents say "yes" to the proposed scenario without really paying attention to the specific content of the scenario. A motive for yea-saying may be a wish to please, either the interviewer or those responsible for the survey (Mitchell & Carson, 1989; Pearce *et al.*, 2002).

Due to the fact that respondents are presented with several multi-attribute alternatives in CMM, rather than the *all or nothing* situation characterising CVM, CMM is believed to be less prone to suffer from yea-saying than CVM (Hanley *et al.*, 2002), given that the respondents put relatively proportional weight on the different attributes.

5.3 Implied Value Cues

Bias due to implied value cues refers to situations, where respondents perceive some elements of the constructed market situation to contain information about the "correct value" of the good (Mitchell & Carson, 1989). In this connection it is important to acknowledge, that stated preference methods are often used to value quite complex goods, with which the respondent is likely to have no prior market experience. Thus, it is goods, which people not necessarily are used to include in their preference structure, and consequently they may be quite susceptible to outside influence from the researcher through the presented scenario.

5.3.1 Starting Point/Range Bias

Starting point bias or range bias emerges, when the amount(s) presented in the scenario influences the WTP stated by the respondent (Sudgen 1999; Hanley & Spash, 1993). That is, it refers to situations where the suggested price in the scenario is perceived to indicate the "true" value of the good. A CMM study by Hanley *et al.* (2002) found that the WTP for an environmental attribute was insensitive to the changes in price vector, and thereby rejected the hypothesis of starting point/range bias.

In terms of the present study, where respondents are presented with four alternatives, described by varying levels of attributes – and where there is no correlation between price and certification – we believe that the risk of starting point/range bias is minimal. However, it may be present in the sense that respondents, when asked if they would consider buying the product ranked 1, have said yes despite the fact that the overall price level of the product is too high.

5.3.2 Importance Bias

The mere fact that a survey is being conducted on a given subject may give respondents the impression that the good or attribute in focus – in our case FSC certification – is especially important (Mitchell & Carson, 1989). This, in turn, may imply that respondents, who prior to participating in the survey have no well-defined preferences for the attribute, are inclined to put undue emphasis on the certification attribute relative to the other product attributes, including price. That is, they might attach more weight to the certification attribute in the survey, than they would in a real market situation (Price, 2000).

5.4 Scenario Misspecification Bias

When elements of the constructed market are perceived differently by the respondent, than intended by the researcher, the value estimates are biased.

5.4.1 Amenity Misspecification Bias

Amenity misspecification biases refer to when respondents perceive the good subjected to valuation differently than intended (Mitchell & Carson, 1989). More specifically, this implies that respondents value another good than the one the researcher thinks the respondent is valuing.

5.4.1.1 Symbolic Bias

If the respondent values a "symbolic" – and more encompassing - good rather than the specific good on which the survey is centred, it will result in symbolic bias (Blamey, 1996). Symbolic bias may be relevant in relation the present study. With reference to chapter 2, at least some of the benefits associated with certification may influence some values, which are quite symbolic in nature. More specifically, the prospects of contributing

to preservation of species or to reduced global warming, are likely to imply that some respondents may consider the act of buying certified products as a "symbolic" act. However, based on the discussion in chapter 2, we do not really consider such "symbolic" motives to be a bias in the present context. Instead, we consider "symbolic" motives to be a natural consequence of the intangibility of the good subjected to valuation and thereby a reasonable explanation for having a positive WTP, rather than a bias as such.

5.4.1.2 Scope or Part-whole Bias

Previous CVM studies have shown, that the WTP for a good varies inadequately with the quantity/scope of the change in the provision of the good. This is known as *scope* or *part-whole bias*. According to Carson *et al.* (2001) it emerges as a problem due to ill survey design or administration procedures in CVM studies. The CMM can address the problem directly (Adamowicz *et al.* 1999). When each respondent is presented with multiple alternatives containing varying levels of the good in question it can be perceived as an inbuilt internal test for scope effect. The sensitiveness to scope was examined in Foster & Mourato (2003) using CE in relation to charitable donations. They found that CE was significantly more sensitive to scope than CVM.

In relation to the present study, a test for part-whole bias could have been investigated in relation to toilet paper. Had one more level of the certification attribute been included – e.g. 50% certified, 50% non-certified – it would have been possible to see, if the estimated WTP would vary with the quantity of the good. Due to practical considerations in terms of the size of the fractional factorial design, this, however, was not possible. Similarly, the assessment of WTP for different products may indicate the sensitiveness to scope. Ideally, everything else being equal, the WTP for certification should be proportional to the amount of wood used in the production of the product.

On a broader level, the diffuse, non-quantifiable and intangible nature of the benefits associated with buying a certified product may imply that scope biases are likely to emerge. It appears that respondents, at least to some extent, may have created different perceptions about the nature and extent of the benefits derived from certification, based on the information they have been provided with. Thus, the provided information has by no means been exhaustive. However, providing sufficient information to eliminate potential scope effects is practically impossible in the present context. With reference to the link between scope effects and information, investigating the effect of information on WTP may be seen as a way of testing for scope effects.

5.4.1.3 Probability of Provision Bias

If the respondent for some reasons perceives the probability that the good in question will be provided to be different from what is described in the scenario, then the value stated by the respondent will not really apply to the specified scenario (Mitchell & Carson, 1989). In

relation to the present study, such biases may be relevant if respondents do not perceive the FSC label to be credible.

Unless respondents feel certain that they get what they pay for – i.e. that the information conveyed by the FSC certificate is correct – they are unlikely to be willing to pay for certification, despite the fact that they may have positive preferences for the environmental issue in question (Kiker & Putz, 1997; OECD, 2001). In relation to the credibility of the FSC label, it may be noted that its credibility – and apparently with due cause – recently has been questioned in the media (Jyllands-Posten, 2003). Fortunately, this did not occur until after the present survey was finalized, implying that the media's attention has not influenced the choices of respondents in the study. Therefore, the preferences that respondents reveal for certified products can only be interpreted as an expression of the value that they ascribe to the public good benefits provided by sustainably managed forests if it is assumed that the respondents perceive the label to be credible. That is, the validity of the valuation approach – or at least the interpretation of the results – adopted in the present study is contingent upon that respondents have considered the label to be credible[23].

In order to minimize this bias, potential relations between respondents' trust in eco-labels and the value associated with certification are investigated in the analysis.

5.4.2 Context Misspecification Bias

If respondents perceive the context of the constructed market differently than intended, then context misspecification biases may arise.

5.4.2.1 Framing Effect Biases

A framing effect occurs, when the individual is sensitive to the context – i.e. the frame – in which a particular trade-off is examined (Rolfe & Bennettt, 2000). Two core issues related to framing are substitute goods and budget constraints. If neither of these are taken properly into account by the respondents, then framing bias will occur. Arrows *et al.* (1993) recommend that the framing effect should be addressed by reminding the individuals of relevant substitutes and their budget constraints. Several studies though have demonstrated that reminders in most cases only have a small effect on the estimated values (Loomis *et al.*, 1994; Kotchen & Reiling, 1999).

CVM estimates have been criticised for being too sensitive to framing effects (Rolfe & Bennett, 2001). CMM has an important advantage over CVM with regards to framing

[23]In the questionnaire, people were asked if they generally have trust in eco-labels on products produced in Denmark, the EC and in LDCs. In order to test the extent to which the quite crucial assumption related to the assumed credibility of the FSC label may be violated, respondents' answers to these trust questions enter as variables in the preliminary analysis of data.

effects caused by neglect of substitute goods. Thus, the simultaneous presentation of several alternatives implies that the individual automatically is forced to take substitute goods into consideration (Rolfe *et al.*, 2002).

Basically, minimising framing biases is a question of constructing a hypothetical scenario that replicates the corresponding real-world situation as realistically as possible (Rolfe & Bennett, 2001). That is, to create a frame that individuals consider realistic. Thus, it is not only substitute goods and budget constraints that affect framing. Other issues related to the frame of the hypothetical market such as the payment vehicle, the method of provision and property rights may also be important. The frame of the present study does not appear to give rise to significant framing biases. More specifically, the CMM frame mimics the real market better than CVM. Also, the payment vehicle is quite uncontroversial, and it is one that respondents are familiar with from other eco-labelled products. Lastly, the set up where respondents are asked to buy a joint private and public good is not new either, since other somewhat similar products are already being sold (e.g. environmental labelled detergents, paper ect.).

5.4.2.2 Ordering Bias

If the orders, in which the elements of the survey instrument are presented, have an influence on respondents' WTP's, then an ordering bias is present. In our case, such a bias may be relevant in relation to the way that the alternatives were presented. More specifically, throughout the sample, price was always the last presented of the product attributes. In a CE study by Hart-Hansen & Kjær (2003) it was tested if it had an effect on estimated WTP, whether the price attribute was presented first or last. They found an effect; the estimated WTP increased when the price attribute was presented first. With reference to this finding, it is possible that the WTP estimated in the present study might be biased downwards.

5.5 Hypothetical Bias

Hypothetical biases arise if the constructed hypothetical situation is not consistent with reality (Pearce *et al.*, 2002). Somehow, all the biases discussed above may arise due the hypothetical nature of the problem. What is in focus here, however, is the bias that may emerge if the value elicitation questions fail to convey to the respondent, that his/her stated choice should be consistent with the choice he/she would make in a real market.

In several cases, the values estimated by stated preference methods have been shown to be significantly higher than real values observed on the market (Cummings *et al.*, 1995; List & Gallet, 2001). However conclusions on differences between hypothetical and real WTP are not uniformly alike. Thus, based on a sample of 83 studies of WTP for quasi-public

goods, Carson *et al.* (1996) concluded that CVM compared to various revealed preference methods, on average understated the WTP with a mean ration of 0,89.

In relation to CMM only a few studies on WTP_{hyp} and WTP_{real} have been conducted. Carlsson & Martinsson (2001) compared hypothetical and real[24] WTP for voluntary donations to an environmental organisation and could not reject the hypothesis of $WTP_{hyp}=WTP_{real}$. In a similar study Johansson-Stenman & Svedsäter (2003) however rejected this hypothesis. These results are supported by Lusk & Schroeder (2002) who examined the hypothetical and real WTP for beefsteaks with different attributes. WTP_{hyp} was significantly higher that WTP_{real}, but the differences were in general only on a 20% level. This difference is rather low compared to the estimated calibration factors (mean WTP_{hyp}/WTP_{real}) found in a meta-analysis of 29 studies in List & Gallet (2001).

In terms of minimising hypothetical bias, List (2002) conducted a CE on the influence of *Cheap Talk* on hypothetical bias; cheap talk being a reminder to the respondents, that in previous similar studies, people have had a tendency to overstate their WTP. The study focused on the difference between WTP_{hyp} and WTP_{real} in relation to both private and public goods, and it was found that CE in general performed very well in relation to hypothetical biases. The hypothetical WTP was only slightly higher than the real WTP, and the inclusion of the cheap talk eliminated the difference for both the private and the public goods.

Due to the frame that is adopted in this study, as will be described in the next chapter, the ranking is framed to take place in a store, we felt that it would not be suitable to subject respondents to cheap talk or any other reminders about their budget constraint. They should be used to taking their budget constraints into account when thinking in terms of buying things. In this connection, we found that such reminders had been more relevant, if respondents were asked to value a good, which they had no prior experience in buying – e.g. a National Park, or cleaner air.

5.6 Validity

An important part of performing a stated preference study is to assess the validity of the study. Validity refers to the extent to which a study succeeds in measuring the values that it is intended to measure. Consequently, assessing the validity of a study is basically equivalent to assessing the extent to which the biases discussed in the previous sections have been present (Edward & Zeller, 1994; Michell & Carson, 1989).

[24] Due to the incentive incompability in CMM the real payment should only be viewed as the real payment in the experiment and not on a market (Carlsson & Martinsson, 2001).

Dealing with non-market goods, there are no unique, definitive yardsticks against which the validity of a study can be determined (Pearce *et al.*, 2002). In practice, the validity of a study is therefore often assessed against several more specific criteria[25]. These criteria, relating to different aspects of validity, are presented in table 5.2. In the following sections, it will briefly be discussed how the validity of a study can be assessed using each of the distinct criteria.

Table 5.2 Different validity criteria, and the aspects covered by each criteria.

Criteria	**Aspects Covered**
Content Validity	Was the study framed appropriately? Were the right questions asked?
Criterion Validity	Are the estimates consistent with: Actual markets? Simulated market experiments?
Construct Validity	Are the estimates consistent with: Economic theory? Intuition? Prior expectations?
Convergent Validity	Are the estimates consistent with: Estimates obtained by using other valuation techniques?

5.6.1 Content Validity

Content validity, sometimes referred to as face-validity, basically refers to the extent to which the scenario has been described in a way that will not give rise to scenario misspecification biases (Garrod & Willis, 1999). Therefore, a study's content validity cannot be assessed objectively. Prior to administration of the survey focus-groups, pre-testing and reviews of the survey instrument by experienced researchers are elements that may contribute to securing the content validity of a study.

In relation to the present study, it was not possible to conduct focus-group sessions. Instead, producers and stores selling the kind of products included in the study were consulted in the process of identifying the relevant product attributes. Pre-testing of the questionnaire, and subsequent discussions with pre-test participants, were conducted. Likewise, the questionnaire was reviewed by two persons with experience within the field of postal research.

[25] The more specific classification and terminology related to the construction of validity criteria, which may serve as the base for assessing the validity of a given study, varies across studies/text books (Garrod & Willis, 1999; Mitchell & Carson, 1989; Pearce *et al.*, 2002). However, it is the same aspects that are covered.

According to Pearce *et al.* (2002), high survey non-response rates may be used as a post-survey administration indicator of a study's content validity. This could indicate that the questionnaire has been too difficult to understand, or offensive in some way.

5.6.2 Criterion Validity

With reference to the aspects covered by criterion validity, testing for criterion validity may - other things equal - be perceived equivalent to testing for hypothetical bias (Garrod & Willis, 1999). The availability of relevant data either from actual market or simulated markets is a prerequisite for assessment of a study's criterion validity. Had the market for certified products been more developed, it would have been possible to explicitly test the criterion validity of the WTP estimates obtained in the present study. In light of this, the scope for assessing criterion validity is limited.

5.6.3 Construct Validity

Construct validity, also termed theoretical validity, refers to the degree to which the results are consistent with expectations based on economic theory and intuition (Mitchell & Carson, 1989). Here focus is mainly on the composition of the model used to estimate WTP – i.e. do the explanatory variables make sense? Is the effect of price as expected? etc.

In the present study, hypotheses regarding the expected effect of each of the variables entering the analysis are set up prior to the analysis of data. The hypotheses are based on a combination of theory, experience from previous studies and intuition. The construct validity of the study will be based on comparisons between the final model and the hypotheses.

5.6.4 Convergent Validity

Assessment of convergent validity entails a comparison of the obtained results with results obtained in other valuation studies and/or by the application of other valuation methods (Garrod & Willis, 1999; Hanley & Spash, 1993). Thus, the criteria, against which the convergent validity of a study is assessed, are equally susceptible to biases as the result of the study itself. Accordingly, if the comparison of results shows that there is a significant difference, it does not necessarily imply that the results of the study in question are invalid. Rather, it indicates that one of the studies – and potentially both – are invalid.

Within the scope of the present study it has not been possible to apply more than one survey format, and at the moment no other Danish studies focusing on WTP for certified products are available. Accordingly, the scope for assessment of the convergent validity of the study is limited. However, studies from other countries of WTP for certified products will be used as indicators of the convergent validity of the results. Likewise, the results of this study will be compared to the results of different studies assessing Danish consumers' WTP for other eco-labelled products.

5.7 Reliability

As validity, reliability also affects the credibility of a study. However, where validity refers to the "truth-value" of the results, reliability is concerned with the accuracy of the survey instrument. More specifically, reliability is concerned with the replicability of the study – i.e. the extent to which the same results would be obtained if the survey was repeated. (Andersen 2002a; Edward & Zeller, 1994; Hanley & Spash, 1993)

Mitchell & Carson (1989) mention four factors that may serve to increase the reliability of a given study:

1. Familiarity with the good, the method of provision, the elicitation frame and the payment vehicle.

2. The good to be valued can be presented in a way that is easy to understand.

3. The scenario appears plausible.

4. The scenario does not challenge prevailing moral assumptions.

Assessing the survey instrument used in the present study against these 4 criteria, it is our opinion that none of the elements in the survey instrument will seriously jeopardize the reliability of the study. In terms of testing the reliability of the study in practice, the two sub-samples – where the only difference is the level of information provided to the respondents – may be compared. If they lead to similar results it may be taken as an indication that the reliability of the study is fairly high.

5.8 Summary

Despite the fact that CMM, and thereby also CRM, appear to be less susceptible to biases resulting from strategic behaviour, yea-saying, suggested prices and framing effects than CVM, the risk of biases cannot be eliminated. In the present context, biases are expected to result primarily as a consequence of the diffuse and non-quantifiable character of the good to be valued – i.e. scope or part-whole biases. Yea-saying, importance, probability of provision and ordering biases may, however, also be present in varying degrees. Furthermore, the results are likely to be affected by respondents attaching a symbolic value to the benefits derived from certification. In the present context, however, we do not consider this to be a bias.

The degree, to which the results of the study are affected by biases will, to the extent that it is possible, be assessed against the four validity criteria, namely; content, criterion, construct and convergent validity.

6 The Questionnaire

In this chapter the design of the questionnaire used in the survey will be presented.

Estimating people's willingness to pay for public goods using Stated Preference Methods (SPM) requires framing of a hypothetical market, where the individual can express his/her preferences. The basic requirements for setting up the survey are more or less identical for each of the methods. The sections below are to a great extent based on the theory of questionnaire design related to the implementation of Contingent Valuation Method (CVM), which is the most extensively investigated SPM (Bateman & Willis, 1999). Specific literature exploring design features in relation to Choice Modelling Methods (CMM) and CRM will be used when relevant/possible.

6.1 Steps in Stated Preference Experiments

According to Adamowicz *et al.* (1999) and Hensher (1994) the preparation for a stated preference experiment can be divided into 5 steps.

1. Identification of the set of attributes describing the good to be valued.
2. Selecting the measurement unit for each attribute.
3. Specification of the number and magnitude of attribute levels.
4. Experimental design.
5. Survey instrument design.

The first three steps are equivalent to the set up of the hypothetical scenario in CVM. The fourth step, the experimental design, is unique for CMM/CRM. The fifth step, survey instrument design, is identical to setting up the questionnaire in CVM. In the present study another step though is added, step 0. In this prior step the relevant wood products are identified. The different steps in relation to this study are presented on the next pages.

6.2 Identification of Wood Products (Step 0)

In chapter 3 the weak complementarity between the purchase of a wood product and preferences for Sustainable Forest Management (SFM) was discussed. The respondents can therefore only express their preferences for the benefits associated with SFM by purchasing a certified product. Asking an individual to rank four alternative wood products, which the individual has no demand for, should theoretically give a zero Willingness to Pay (WTP) for certification (and all the other attributes defining the wood product). However, this WTP would not reflect the value derived from certification. It would only reflect the fact that the respondent has no preferences for the product. Based on this consideration, it was of great concern, that the chosen products would be relevant for as many respondents as possible. Since the questionnaire was mailed to a complete random

sample of the Danish population, the choice of wood products should therefore represent ordinary Danish consumers' preferences. Based on these considerations, the following three wood products were chosen:

- Toilet paper
- Cutting boards
- Table tops

Thus, within the different types of wood product, these products are believed to represent some of the products that respondents are most likely to be familiar with, and have preferences for. However, both in relation to cutting boards and table tops, we are aware of the fact that some people, for different reasons, prefer products made of other materials than wood[26]. Furthermore, in relation to table tops, there is a risk that some respondents might not have preferences for the product. More specifically, it may be that respondents who are not homeowners would never consider buying a table top[27]. However, referring to the objectives, it was desired to assess the WTP for products representing different types of products in terms of price and frequency of purchase. In this context the chosen products were perceived to be the most relevant candidates.

6.3 Identification of Attributes and Attribute Levels (Step 1)

Identifying the relevant key attributes and levels of each attribute is an important task in CMM. Depending on the type of good in focus, attributes can be identified by interviewing policy makers, experts and by conducting qualitative surveys and focus groups sessions (Louviere, 1988; Adamowicz *et al.*, 1998b; Bennett & Adamowicz, 2001). According to (Keeney & Raiffa, 1993) the chosen attributes must have the following properties:

- Completeness: The attributes must represent all the important characteristic of the good.
- Operational: The attributes must be meaningful to the sample.
- Decomposable: The good evaluated can be broken down into smaller parts of value.
- Non-redundancy: The attributes must be mutually exclusive to avoid double counting.
- Minimum size: The set of attributes should be kept as small as possible.

[26] Respondents who would never consider buying the wood products in question cannot be expected to have well-defined preferences for them. Consequently, such respondents will be excluded from the analysis (see chapter 8). Thus, it is expected that they would be inclined to attach too much emphasis on certification, which might be the only attribute of the products that they have preferences for.
[27] To accommodate for this, it will be tested if respondents' type of accommodation influences WTP.

When CMM is conducted in relation to an environmental change the attributes most commonly define the range of the environmental change – like how many birds, levels of access to an area, number of wild plants etc. In this study the focus is on WTP for, and the relative importance of, the FSC label compared to the other attributes defining the wood products. Thus, the environmental change is expressed through the certification attribute, which defines the complete good to be valued on equal terms with other product attributes. This approach is identical to the approach used in similar studies on wtp for eco-labelled goods (Loureiro *et al.*, 2000; Pickering *et al.*, 2001; Jensen *et al.*, 2002).

The identification process in this study consisted of two phases:

1. Preliminary investigations:
   - Field trips to relevant shop and stores. All three products are available in different qualities and are therefore sold in both discount stores and specialised shops; both types of stores were visited. Potential attributes were identified by observing the attributes highlighted in the product description/ads and by interviewing the staff in the shops.
   - The attributes of the tabletops were furthermore investigated by interviewing selected producers. (Aalborg Træindustri, 2002)

2. Pre-test of the questionnaire, (33 respondents):
   - Potential attributes and attribute levels were tested in the pre-test of the questionnaire. The respondents were asked to comment on the chosen attributes and report if any attributes were irrelevant or missing.

The results of the pre-test verified the choice of attributes and their levels. None of the respondents viewed the presented attributes as irrelevant, and only a few respondents had complementary attributes to add. All in all, the pre-test of the questionnaire verified, that the chosen attributes of the all three products were complete and operational.

6.4 Number of Alternatives, Attributes and Attribute Levels (Step 2 and 3)

When choosing the final number of alternatives, attributes and attribute levels the following four aspects must be taken into consideration (Ryan; 1999; Deshazo & Fermo, 2002):

1. The levels must be realistic to the respondent.
2. The respondents must have preferences and demands for the chosen attribute levels.
3. The construction and presentation of the levels must represent a trade-off to the respondent.

4. The overall number of alternatives, attributes and the levels must be considered in relation to the cognitive burden of the respondent.

6.4.1 Cognitive Burden

In terms of the cognitive burden of the task, behavioural economists predict that an increase in information and thereby choice complexity increases the noise added to the error term (Dellaert *et al.*, 1999). DeShazo & Fermo (2002) investigated the impact on choice consistency as a function of the number of alternatives (2-9) and attributes (4-9) in a CE. They found that an increase in the number of attributes increased the variance on the error component in the utility model. In terms of the number of alternatives, it was found that increased variation in the error term emerged after the third alternative, but this number of course depends on the specific study. Similarly, Mazzotta & Opaluch (1995) and Swait & Adamowicz (1996) found that increased choice difficulties increase the variance. Mazzotta & Opaluch (1995) found in a binary CE that respondents started using simplified (heuristic or lexicographical) choice strategies when there were more than 4 attributes per alternative in the choice set. The maximum number of alternatives indicated by these results are quite different from the number of alternatives previously used in marketing surveys, where choice sets with of up to 16 alternatives have been used (Beggs *et al.*, 1981).

The above-mentioned studies point out that the number of alternatives and attributes significantly affect the overall cognitive burden imposed on the respondent and thereby exerting a large influence on the choice made by the respondents. The gains associated with increasing the number of attributes and thereby the accuracy of the description of the products should therefore be compared to the cognitive capabilities of the respondents.

6.4.2 Choice of Attributes and Attribute Levels

Based on the field trips, interviews of experts, pre-testing of the questionnaire and with regards to the four aspects mentioned in section 6.4 the following attributes and attribute levels for toilet paper, cutting boards and tabletops were chosen, see Box 6.1 . The levels of the physical attributes generally mimic the levels on the market, and are believed to fulfil the first two aspects in section 6.3. But in some cases a simplification or adjustment of the attribute levels has taken place. These are commented on below.

6.4.3 Toilet Paper

On the market few toilet paper brands have four layers of paper. Due to the apparently limited market share, this level of layers was not included. Recycled paper in this study has only two levels 100 % or none, which does not completely correspond to the market situation, where levels between 0-100 % recycled paper are available. In terms of quality, the softness attribute and the number of layers attribute are intended to reflect different quality related aspects of toilet paper.

6.4.4 Cutting Boards

Cutting boards are available in many designs. The designs can in general be divided into cutting boards with horizontally orientated wood sticks or vertically orientated wood sticks, why these two types of design were chosen. Cutting boards can be found in various wood species, but during the field trips it became clear that the main species are the ones mentioned in Box 6.1. It is expected that the perceived quality of a given cutting board is an important determinant of choice. However it was found to be difficult to identify a quality attribute for cutting boards. It is therefore assumed that the presented alternatives are of identical quality.

6.4.5 Table Tops

Table tops are available in various wood species (Junckers 2002; Spekva, 2002; Aalborg Træindustri, 2002). Talking with experts on the area, it became evident, that the different wood species are important determinants of the choice, but they are not necessarily substitutes in economic sense. The choice of wood species for a table top depends on the type of kitchen and especially the colour of the kitchen. People with a light coloured kitchen will most often choose a relatively dark wood species and vice versa. In that sense buyers of table tops may be perceived to have lexicographical preferences for the shade of the wood species, but are willing to substitute between wood species with the same type of shade. Due to the difficulties in mimicking such a market situation in the ranking, the choice of the wood species, was made prior to the actual ranking, so that the respondents could choose the wood species they preferred.

6.4.6 Prices

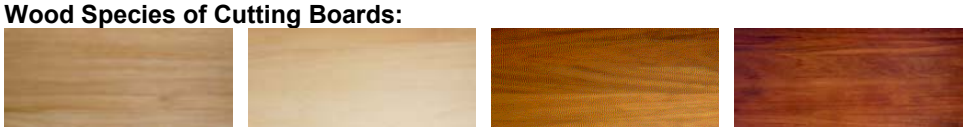
Prices for conventional toilet paper, cutting boards and table tops are observable on the market, why the prices in the scenarios are based on these. However, since certified products are expected to be traded at a price premium compared to conventional products, the prices in the three product scenarios are increased market prices for the specific product of different qualities. According to Gyrd-Hansen (2002), the use of the enhanced prices is furthermore necessary in order to make measurement of the maximum WTP possible[28].

[28] Measuring a maximum WTP, requires that the maximum simulated price is so high that the respondents rejects the alternative.

Toilet Paper		
Attributes	Description of attribute levels	Attribute levels
Softness	Standard soft/Soft/Very soft	3
Recycled fibres	New fibres/Recycled fibres	2
Layers	2 layers/3 layers	2
FSC certified	Not certified/Certified	2
Price in DKK.	9/13,25/17,50/21,50/25,75/30,00	6

Cutting Boards		
Attributes	Description of attribute levels	Attribute levels
Wood Species	Hevea/Beech/Teak/Rosewood, see below	4
Design	Vertically-/Horizontally orientated sticks, see below	2
FSC certified	Not certified/Certified	2
Price in DKK.	100/165/230/295/360/425	6

Wood Species of Cutting Boards:



a. ***Hevea*** b. ***Beech*** c. ***Teak*** d. ***Rosewood***

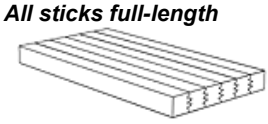
Design of Cutting Boards:

a. ***Vertically orientated sticks*** b. ***Horizontally orientated sticks***

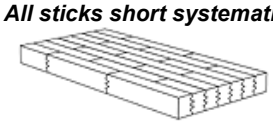
Table Tops		
Attributes	Description of attribute levels	Attribute levels
Design	a: All sticks full length b: Front stick full length-remaining sticks short c: All sticks short- systematically assembled d: All sticks short- randomly assembled, see below	4
FSC certified	Not certified/Certified	2
Variation in wood	Very uniform-perfect wood, few to no knots/Little variation- slight blemish, some knots	2
Price in DKK. per linear meter	1.400, 1.600,1.800, 2.000, 2.200, 2.400, 2.600, 2.800, 3.000, 3.200	10

Design of Table Tops:

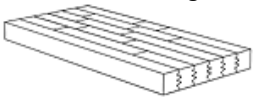
a. ***All sticks full-length***



c. ***All sticks short systematically assembled***



b. ***Front stick full-length remaining sticks short***



d. ***All sticks short randomly assembled***

Box 6.1 Attributes and attribute levels of toilet paper, cutting boards and table tops.

6.4.7 Summary

With references to Box 6.1 the number of alternatives and attributes are seen to be within the standards of state of the art CRM, see table 6.1.

Table 6.1 Number of alternatives and attributes in other CRM studies.

Ranking Studies	No. of Alternatives	No. Attributes
This Study	4	4-5
Foster & Mourato (2002)	3 + status quo	3
Hanley *et al.* (2001)	3 + no option	6
Foster & Mourato (1999)	4	4
Garrod & Willis (1998)	2 + status quo	5

6.5 The Experimental Design (Step 4)

The choice of how to design the experiment is crucial for the outcome of the subsequent discrete choice analysis. As such, the design of an experiment predefines which observations the discrete choice analysis is based on. Consequently the design also determines which effects that can be estimated from the data. Several methods are available to identify designs each with different abilities with regards to discrete choice modelling (Kuhfeldt *et al.*, 1994; Kuhfeldt, 2000; Bunch *et al.*, 1996). In general two types of designs exist:

- Full factorial design.
- Fractional factorial design.

In the following sections the design theory will be presented and discussed in relation to this study.

6.5.1 Full Factorial Design

A full factorial design is a design where each level of one attribute is combined with every level of all other attributes (Louviere, 1988). This implies that every possible alternative is presented in the design. As the product of the levels of the different attributes gives the number of possible alternatives, even small discrete choice problems have many possible alternatives. Accordingly using full factorial designs require large samples (Louviere *et al.*, 2000). Shown in Box 6.1, toilet paper has 5 attributes with (3,2,2,2,6) levels, cutting boards 4 attributes with (4,2,2,6) levels and tabletops 4 attributes with (4,2,2,10) levels. This gives 3 full factorial designs consisting of; $3 \cdot 2 \cdot 2 \cdot 2 \cdot 6 = 144$, $4 \cdot 2 \cdot 2 \cdot 6 = 96$ and $4 \cdot 2 \cdot 2 \cdot 10 = 160$ possible attribute combinations.

Full factorial designs have attractive statistical properties, and ensure that both main effects and all interaction effects between attributes can be estimated independently from each

other[29]. Empirical evidence suggests that main effects account for 70-90 % of the explained variance of choice, whereas interactions account for the remaining 10-30 %. Of these 10-30 %, two-way interactions accounts for half (Louviere *et al.*, 2000).

6.5.2 Fractional Factorial Design

A fractional factorial design only contains a subset of all possible alternatives. Depending on how the subset comprising the fractional factorial design is specified, it is possible to make a design, which allows pre-specified effects to be estimated from a reduced number of alternatives. A fractional design, though, may suffer from the impact of omitted interactions. This means, that the estimated effect of an attribute may be confounded with an omitted interaction. That is, the estimated effect will in fact reflect the effect of both the attribute and the omitted interaction. If the omitted interaction is significant, this implies that the estimated effect of the attribute is biased (Cochran & Cox, 1992).
As an example, the main effect "certification" might be confounded with the two-way interaction effect certification·softness. Such a confounding could have the following form;

certification = certification + 0,05·certification·softness.

If certification·softness is omitted from the fractional factorial design, then the certification parameter is biased by 0,05 times the parameter for certification·softness.

Due to the relatively large explainable variance of the two-way interactions (5-15%), fractional design should as a minimum encompass two-way interactions. Alternatively, if using a pure main effect design, which significantly reduces the size of the design, care should be taken to ensure, that main effects either are non-confounded or only confounded with three-way interactions or higher[30]. In most applications of CMM, main effects designs are used, which is also the case in this study.

6.5.3 Design Efficiency

Having decided to apply a main effect fractional factorial design, four criteria can be set up for identifying the most efficient design (Kuhfeld *et al.*, 1994; Zwerina *et al.*, 1996; Huber & Zwerina, 1996). The measurement of the efficiency is based on the *information matrix* **X´X** (A, D and G efficiency, which are relative measurements), see appendix I. The information matrix is proportional to the variance covariance matrix of the parameters in a least square analysis (Kuhfeld, 2000; Kuhfeld *et al.*, 1994). The efficiency of a design is thus a function of the variance and covariance of the parameters, such that efficiency

[29]A main effect can be defined as the pure effect an attribute has on the probability of choice. Similarly an interaction effect represents the joint effect of two or more attributes on the probability of choice.
[30] Three-way or higher levels of interactions totally only accounts for 5-15 % of the total variance, why the impact of an omitted variable is minimal if a main effect is confounded with a three-way – or higher level interactions

increases as variance decreases. The four criteria mentioned above are; level balance, orthogonality, minimum overlap and utility balance:

6.5.3.1 Level Balance

Level balance occurs, when each level of an attribute occurs with equal frequencies and when combinations of different levels across attributes also occur with equal frequencies. As an example, each level of a two level attribute, such as whether or not the wood product is certified, should occur in 50% of the alternatives. This ensures an equal weight to each level in the trade-off options of the respondents.

6.5.3.2 Orthogonality

Orthogonality implies, that the occurrence of an attribute level is independent of the levels of other attributes. Orthogonal designs therefore minimise correlation among the independent variables in a choice set/design. This ensures, that unique estimates of the parameters are not influenced by design properties. In orthogonal designs, dominating or implausible[31] alternatives might be present. Besides the relatively lower information obtained from dominating alternatives, the implausible alternatives might puzzle the respondent or even make the respondent disbelieve the choice set. Therefore, it is sometimes suggested, that dominant or implausible alternatives are removed from the design. But as it can be seen from the hypotheses in table 9.2, the expected sign of the utility related to the attributes "recycled", "design" and "wood species" cannot be defined a priori, why implausible and dominant alternatives are difficult to identify. Consequently, it was chosen not to remove any alternatives.

6.5.3.3 Minimum Overlap

Minimum overlap is satisfied, when the alternatives within the choice set have non-overlapping attribute levels. That is, if the alternatives are pair-wise different in all attributes. Since in discrete choice modelling, choices are modelled as a function of differences between alternatives, a minimum overlap maximizes the information from each choice set.

6.5.3.4 Utility Balance

Utility balance refers to, that choice sets should be defined in a way so that the utility across the alternatives are of a similar magnitude. Thus utility balance ensures, that the respondents actually do make trade-offs. As such a utility balanced design increases the significance on parameter estimates. But it can also increase the level of collinearity between attributes, and thereby reduce the efficiency of the choice set. Introducing utility balance in the choice set is therefore a trade-off between improved models and reduced orthogonality in the design. The gain in information in a utility balanced design/choice set

[31] A dominating alternative is an alternative, which in all attributes is superior to the other alternatives in the choice set. An implausible alternative is where the combination of attribute levels makes it unbelievable.

must also be viewed in relation to the trade-off with the increase in variance on the error component, due to the much harder task of choosing/ranking a utility balanced choice set (Swait & Adamowicz, 1996).

The criterion of utility balance is difficult to satisfy, since it demands prior knowledge of the true distribution of the parameters, which can be difficult to predict. Huber & Zwerina (1996) have found potential gains in design efficiency by the inclusion of such prior knowledge. In the present study prior knowledge was not available, why the criterion for creating a utility balanced design was not met.

6.5.4 Construction of Fractional Factorial Designs

Construction of efficient designs in this study was carried out in SAS using the macros %MKTRUNS and %MKTDES, which explore the experimental design size and construct efficient linear designs. The designs found in %MKTDES were blocked in PROC OPTEX, which ensures an orthogonal blocking in choice sets of 4 alternatives. The above-mentioned SAS programmes are able to meet three of the criteria; level balance, minimum overlap and orthogonality, but not utility balance.

For toilet paper, the fractional factorial main-effects design resulting from the above-mentioned procedures in SAS, contained 36 alternatives, which subsequently were blocked into 9 choice sets of 4 alternatives each. For cutting boards the design contained 24 alternatives, implying that 6 choice sets of 4 alternatives each were created. Lastly, for table tops the final design contained 40 alternatives, which were blocked into 10 choice sets of 4 alternatives. The designs, including information on their efficiency and confounding structure, are presented in appendix I.

6.6 The Questionnaire

Having defined the relevant key attributes, identified an efficient design and choice sets with the right properties to estimate the relevant effects unbiased, the framing of the choice situation is established through the development of the survey instrument i.e. the questionnaire.

As in CVM, a questionnaire generally obtains three types of information from the respondent:

- Attitudes towards and preferences for the good to be valued.
- The elicitation of preferences in relation to the good.
- Data on socio-economic and demographic characters.

In the following presentation, the different parts of the questionnaire are elaborated and discussed. For the full-length questionnaire see appendix II.

6.6.1 Introduction

On the first page of the questionnaire, the respondents are introduced to the issue under investigation, how and why the respondent has been sampled, who is conducting the survey etc. By this formal presentation of the survey, it is the goal to gain the respondents' confidence and trust, and thereby increase the number of valid answers (Bennett & Adamowicz, 2001).

6.6.2 First Part of the Questionnaire

Besides exploring the knowledge of the respondent and identifying possible determinants of the WTP (Arrow *et al.*, 1993) the first part of the questionnaire primarily serves the purpose of catching people's attention and making them tune into the context of the choices they are about to make (Bennett & Adamowicz, 2001; Loomis *et al.*, 1994). In the first part of the questionnaire people's knowledge of, experience with and trust in eco-labels on the Danish market is explored through fill-in matrices and single questions. To catch the respondent's interest, the first question includes a coloured presentation of different eco-labels.

Respondents are also asked to state their beliefs about forest management and to identify which characteristics they find most important of different wood products. Throughout the questionnaire the wording of the questions are chosen with care to avoid an unintended negative or positive effect on people's perceptions of the issue under investigation (Mitchell & Carson, 1989; Arrow *et al.*, 1993).

6.6.3 The Hypothetical Scenario

The hypothetical scenario consists of three interlinked elements (Loomis *et al.*, 1994); 1) the description of the hypothetical market and the products in focus, including description of the different attributes and their levels, 2) description of the payment vehicle and 3) finally elicitation of the respondents' WTP – that is, the ranking of the presented alternatives.

6.6.3.1 Presentation of Attributes and Attribute Levels

In the following subsections the presentation of the attributes is discussed.

Information

The level of information given in a valuation scenario is important, since the knowledge and perceptions of the environmental goods in focus most commonly varies among respondents (Mitchell & Carson, 1989). Identifying the appropriate information level is a delicate matter. The level of information must ensure that the environmental change is

understood and its' providence is accepted by the respondent (Carson *et al.*, 2001). On the other hand if the level exceeds some unknown level, *information overload* might cause the respondent to focus on *valuation neutral* information and not on the *valuation relevant* information (Mitchell & Carson, 1989). The level of information provided in surveys is a trade-off between giving the respondent sufficient information to evaluate the good and giving too much, so that focus changes to neutral information or too strong preferences for the good is created. The latter is also known as *importance bias*, which was discussed in section 5.3.

With reference to the objectives of the present study, two levels of information about certification were provided to different sub-samples, in order to assess the effect of information on WTP. The levels of information were chosen so that the low-level-information scenario reflects the level of information available on the real market. On the real market FSC labelled products have a small label attached, stating that the wood originates from a sustainably managed forest in accordance with the standards of FSC certification. This level of information corresponds to the information used and recommended in similar studies on WTP for eco-labelled products (Loureiro *et al.*, 2000; Pickering *et al.*, 2001, Jensen *et al.*, 2002). As a contrast, a brochure by the FSC was chosen as the alternative high-level-information scenario; the brochure is included in appendix II. The brochure gives more detailed information on the rules of certification and provides the respondent with both general and specific indications on what values the FSC label ensures in relation to forest protection/management.

Pictorial vs. Physical vs. Verbal Presentation of Products/Attributes
Generally, verbal presentation works well in describing goods, but with the increased use of computers, verbal presentation/stimuli of attributes has to some extent been replaced by pictures and figures. The use of pictorial stimuli is found to be crucial, when attributes like style (wood species) and design are important for the choice of the respondent (Vriens *et al.*, 1998).When appropriate, use of pictorial stimuli furthermore makes the frame of the choice situation more realistic and increases the reliability of the responses (Vriens *et al.*, 1998; Mitchell & Carson 1989). This potentially reduces some of the misspecification biases, which were discussed in chapter 5. In this study the perfect frame would naturally be a presentation of the physical products, but due to the mailing format of the questionnaire, only verbal and pictorial presentations of the products were possible. The difference in choice behaviour in relation to physical/pictorial stimuli is though not necessarily significant according to Jaeger *et al.* (2001). Pictorial stimuli can even in some cases increase the cognitive burden of the respondent (Vriens *et al.*, 1998), why they should only be used when appropriate. A physical stimulus was considered to verify the softness of toilet paper, but during the pre-test of the questionnaire respondents replied, that they did not use or need the included samples of toilet paper. The physical stimulus

was therefore excluded in the final questionnaire. Finally, it was decided not to present the quality attribute for table tops (number of grains and degree of symmetry) pictorially, since the visual difference varied too much between wood species. Instead a more neutral, verbal presentation was chosen, which can be interpreted as a conservative design solution, which is recommended by Arrow *et al.* (1993).

The pictures of the wood species, the designs of cutting boards and of table tops used for the pictorial presentation were either photographed or downloaded from the internet from producers' homepages.

6.6.3.2 Payment Vehicle

In SPM the payment vehicle and payment schedule has proved to have influence on the acceptance of the scenario and the following WTP exercise (Morrison *et al.*, 2000; Stevens *et al.*, 1997). According to Navrud (2001) there are four main types of payment vehicles:

- General taxes
- Special taxes
- Increased prices on goods related to the environmental good or service
- Earmarked fund for provision of the environmental good or service.

According to Mitchell & Carson (1989) and Bennett & Adamowicz (2001) the chosen payment vehicle must be plausible in relation to the proposed solution to the issue in focus. In connection to eco-labels, consumers are accustomed to associate the finance of an eco-label with the price of the product, why the "increased prices on goods" approach is the most relevant payment vehicle in relation to FSC certified products.

6.6.3.3 Elicitation of WTP

Prior to the ranking, the following scenario was constructed; it was attempted to create a situation that would resemble a real market situation as closely as possible. Respondents were asked to imagine that they were in a situation where they were going to buy the product in question. Subsequently, the attributes defining the available alternatives were presented. Afterwards, respondents were presented with a choice set of four alternative products, which they were asked to study closely with regards to the attributes. Finally, they were asked to rank the products, by assigning the number 1 to the most preferred product, the number 2 to the second most preferred, etc.

Choosing a No-option or Not

The ranking does not include a no-option[32]. If a no-option is included, it should be framed, so that it mimics real market behaviour and even so framing effects can occur. Banzhaf *et*

[32] A no-option provides the respondent with the opportunity to state if he/she does not want to choose any of the products (Batsell & Louiviere, 1991).

al. (2001) found that framing of the no-option, had a significant impact on the model parameters, proving that the definition of the no-option is very important.

Theoretically, the respondent should only use the no-option if the alternative entailed in the no-option yields a higher utility, than the alternatives in the ranking. But respondents in general have a tendency to use the no-option as an easy choice, when the perceived utility among alternatives in the choice set is of a similar magnitude (Huber & Pinell, 1994). In this case the no-option both reduces the effectiveness of the sample and gives false indications of the importance of an environmental good, in this case certification. On the other hand, a forced choice may overstate the likelihood of choosing specific attribute levels of the alternatives (Banzhaf *et al.*, 2001), and thereby bias the elicitation of preferences. It should though be mentioned, that the direction of such a bias depends on the difference in utility between alternatives and the domination of attribute levels (Dhar & Simonson, 2003).

The studies mentioned above are all conducted with CE, where the presence of a no-option seems natural, but in relation to ranking it is less obvious how a no-option could be included. Carson *et al.* (1994) suggests that the no-option should be included, when it increases the realism of the task. Following this concept in relation to CRM, a buy/no buy of the most preferred alternative is included at the end of the ranking. This type of no-option assumes that the respondent, presented with four alternatives, first evaluates the alternatives within the choice set and then decides whether to buy the product or not. It would have been interesting to see if inclusion of the no-option together with the alternatives would have changed the parameters in the derived model, but due to practical constraints this has not been possible.

6.6.4 The Last Part of the Questionnaire

The last part of the questionnaire consisted of both debriefing questions and socio-economic questions. The placement of socio-economic questions differs among surveys, but experiences have shown, that questions on personal characteristics are less rejected by respondents, if placed in the end of the questionnaire (Andersen, 2002a; Mitchell & Carson, 1989). Questions on age, education, occupation, accommodation, and income were asked.

Debriefing questions were asked immediately after the completion of the ranking, and focused on the respondents' consumer patterns. As recommended by Arrow *et al.* (1993), debriefing questions at the end of the questionnaire, should be used to check respondents' understanding and acceptance of the key elements of the environmental good in focus, and to test for irrational preferences. The debriefing questions are also appropriate to probe for motives for the respondents' choices (Hanemann, 1994). Using debriefing questions to identify irrational or inconsistent respondents in CRM is not as straightforward as in CVM.

An example: if a respondent in a CVM survey is willing to pay x DKK for a certified product, but in the debriefing of the questionnaire states, that he/she does not trust eco-labels or never buys such products, then the respondent does not express rational preferences. In CRM it is difficult on individual level to determine the motive of choice. Even though a certified alternative is ranked 1, this ranking could be motivated by the level of the other attributes of the wood product, such as the softness of toilet paper or design of the table top. Testing for irrational preferences would therefore require a specific design of the questionnaire. The design used in this study, does not facilitate a screening for rationality of expressed preferences.

6.7 Summary

The different steps in the design of the questionnaire used in the survey are presented. The motives for choosing toilet paper, cutting boards and table tops as the focus of the survey are outlined. The attributes characterising the products are identified, and the relevant presentation of the attributes are discussed. Subsequently, the experimental design of the survey is specified, and the frame of the ranking experiment is described.

7 Modelling Discrete Choice Data

In this chapter, the theory underlying the modelling of, as well as the characteristics defining, discrete choice data is presented. Furthermore, the Random Utility Theory (RUT), which is fundamental in discrete choice modelling, is described. Subsequently, the multinomial logit model and rank logit models[33] are introduced, and important aspects related to the modelling of rank data are discussed. The maximum likelihood estimation procedure is briefly presented, and lastly it is described how estimates of Willingness to Pay (WTP) for certification are obtained from Discrete Choice Models (DCM).

7.1 Discrete Choices

People are confronted with choices everywhere; at work, at home or in the leisure time. Inherent in these choices are trade-offs between time, money and other factors affecting individual utility. In this study, the primary objective is to assess individuals' WTP for certification by modelling their choices between certified and conventional wood products of different types. The data does not directly reveal the respondents maximum WTP for certification. Instead, it has to be deducted from the trade-offs inherent in the discrete choices that the respondents make between certification and price in the rankings. The choice set, and the alternatives included in the choice set, have to meet certain requirements in order to allow such indirect elicitation of WTP.

A discrete choice may be defined as a choice where an individual is faced with a set of alternatives from which he/she is only allowed to choose one – or in our case, to choose one a number of times. The choice set has to meet the following criteria (Train, 2003):

- The alternatives must be *mutually exclusive*.
- The choice set has to be *exhaustive*.
- The number of alternatives must be *finite*.

The criterion stating that the alternatives must be mutually exclusive implies that given alternatives A, B and C, choosing A means *not* choosing B and C. In relation to a ranking of alternatives this means that if A is ranked 1, B is ranked 2 and C is ranked 3, A is chosen and not B and C given only one choice etc. The criterion of exhaustiveness implies that all possible alternatives must be included. This may appear restrictive, but if respondents are provided with a no-option the criterion is met[34]. The last criterion, i.e. that

[33] A probit model would have been interesting to use in this study since it is not as limiting as the logit. But apparently multinomial probit or ranked ordered probit are only available in a demo package in SAS, why the probit model is left unattended.

[34] In the present study, the exhaustiveness requirement is not directly incorporated in the ranking question. Instead it is accounted for by asking respondents, upon completion of the ranking, whether or not they would actually consider buying the product ranked 1.

the choice set must be finite, can be restrictive. Thus, if the levels characterising an attribute are continuous – i.e. there are an infinite number of levels – rather than discrete, there is, in effect, an infinite number of alternatives. In the present study, an example could be the pattern of the grain of a table top. To model discrete choices as a function of such attributes, the continuous characteristics have to be converted into to discrete variables. Thus, the appropriateness of discrete choice modelling in a given context depends on the extent to which continuous characteristics can be transformed into a finite number of discrete variables in a meaningful way[35].

Provided that the choice sets satisfy the above-mentioned criteria, and that the alternatives included in the choice set meet the requirements presented in section 6.4, a model can be estimated, from which the Marginal Rates of Substitution (MRS) between different attributes defining the alternatives can be derived.

7.2 The Random Utility Model

As mentioned in chapter 3, individuals are assumed to exhibit utility maximising behaviour. The utility model underlying DCM is the RUT, which was introduced by Thurstone (1927), and further developed by Marschak (1960).

Modelling the choices of individuals requires (Louviere *et al.*, 2000):

- That the choice set available to decision maker, and the choice made by the decision maker from that choice set, is known.
- That the characteristics of the decision maker, which are relevant for the choice, are known.
- That a model of individual choice and behaviour is constructed.

Provided that these requirements are met, a Random Utility Model (RUM) describing the individual's choices can be derived as follows.

Assume that an individual, n, is confronted with a finite number of alternatives, J, where the individual is expected to obtain utility from each alternative, j. The specific utility is denoted as U_{nj}[36], $j = 1,...J$. Among the different j alternatives the individual chooses the alternative i, yielding him/her the highest level of utility, so that $U_{ni} > U_{nj} \forall j \neq i$.

[35] With reference to box 6.1, it is believed that the characteristics of the three types of products have been transformed into discrete numbers in a sensible matter.
[36] Marschak (1960) used the notation of utility U, which was a *real-valued random function* and is equivalent with U_{nj} used in this study.

The researcher engaging in discrete choice modelling of the individual's choice can observe neither U_{ni} nor $U_{nj.}$ However, the researcher has access to information regarding the attributes of the alternatives faced by the decision maker, labelled $x_{nj} \forall j$, and information regarding the attributes of the decision maker, labelled s_n. Based on this information the researcher can define a function, which relates the information on x_{nj} and s_{nj} to the utility of the decision maker. The function is denoted $V_{nj} = (x_{nj}, s_n) \forall j$. However, it is unlikely that all the information defining the utility of the decision maker is available for the researcher, why $V_{nj} \neq U_{nj}$.

This gives rise to the partitioning of the utility in two components $U_{nj} = V_{nj} + \varepsilon_{nj}$, where V_{nj} is referred to as the *systematic component* or *representative utility* and ε_{nj} is referred to as the *random component*. In this context it is important to note, that the inclusion of a random component does not imply, that the decision maker maximises utility randomly. Thus, he/she is still assumed to behave rationally. The inclusion of the random utility component is solely a consequence of the fact that that the variables known to the researcher – i.e. the variables included in the model – fails to capture all factors determining U_{nj}.

Based on the above specification of the utility function, the probability, P_{ni}, that individual n chooses alternative i over alternatives j is given by:

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj} \forall j \neq i) = \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i) \\ &= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) \end{aligned} \quad (7.1)$$

7.1 states that the probability of n choosing i, from a set of $J = 2$ alternatives is equal to the probability that the difference in random utility between j and i is less than the difference in systematic utility between i and j. Subsequently, P_{ni} can be expressed as a cumulative probability:

$$P_{ni} = \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \ \forall \ j \neq i) = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall \ j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (7.2)$$

Where I is the indicator function, which is equal to 1 when the expression is true – i.e. if the respondent chooses product i – and $f(\varepsilon_n)$ represents the density distribution of the random component. The researcher chooses the distribution of the random utility component, and different distributions result in different models.

7.3 The Functional Form of the Utility Function

The systematic component of the utility is a function of both alternative specific attributes (X_{nj}) and individual specific (S_n) attributes, denoted as $V_{nj} = (x_{nj}, s_n) \forall j$. The fit of the model not only depends on how well the attributes included in the utility function represent the choice of the individual, and which distribution is chosen for the random component. It also depends on the specification of V_{nj} in terms of how the attributes affect the utility of the individual.

The most commonly used and simple specification of the utility function is the *linear additive utility function* (Louviere & Woodworth, 1983; Louviere *et al.*, 2000). The linear additive function is denoted by (Train, 2003):

$$V_{nj} = x_{nj}'\beta + s_n'\delta \tag{7.3}$$

Where x_{nj} denotes the vector of attributes of alternative j facing individual n, and s_n denotes the vector specifying the characteristics of individual n, and β and δ are the coefficients of those vectors. By adopting this specification the total utility of an alternative is given by the sum of the utility associated with each of the components entering the utility function. In this study, the linear additive form is used in the modelling.

7.4 The Multinomial Logit Model

A variety of more specific logit model specifications exist (Greene, 2003; Train, 2003). In this section the most commonly used of these will be presented. This model specification – i.e. the Multinomial Logit Model (MNL) – builds on the logit specification formulated by Luce (1959), which was further developed by McFadden (1974). An inherent property of all conventional logit models is the Independence of Irrelevant Alternatives (IIA) property. The implications this assumption has on the applicability of the models will be discussed in section 7.6.

The MNL is used to model choices, where one alternative is chosen from a choice set consisting of $J \geq 2$ alternatives[37] (McFadden, 2001). It is based on the assumption that the ε_{nj} are Independently Identically Distributed (IID). The assumption of independence means, that each ε_{nj} is independent of the other $\varepsilon_{ni} \ \forall j \neq i$, why information related to ε_{nj} cannot explain ε_{ni}. More specifically, the ε_{ni}'s are assumed to follow a Gumbel or type 1 extreme value distribution. This distribution is characterised by a scale parameter μ and a location parameter δ, which in practice are usually set to 1 and 0, respectively, thereby obtaining the standard Gumbel distribution (Alpizar *et al.*, 2001). The scale parameter is related to the variance of the distribution such that $\text{var}\varepsilon = \pi^2/6\mu^2$. Compared to the normal

[37] In the case where J=2 the multinomial logit model collapses to the binary logit model.

distribution, the Gumbel distribution has a flatter tail, implying that it to a higher extend can accommodate aberrant choices, such as potential strong preferences for certified wood products.

Assuming that random components are Gumbel distributed, the MNL choice probability is given by (Alzipar *et al.*, 2001):

$$P_{ni} = \frac{e^{\mu V_{ni}}}{\sum_j e^{\mu V_{nj}}} \tag{7.4}$$

As it is seen from 7.4 the coefficients that are estimated for the model are confounded with the scale parameter. Usually, this correlation between the scale parameter and the obtained coefficient estimates is irrelevant when dealing with a specific model (Ben-Akiva & Lerman, 1985). As will be discussed in the next section, it may however have important implications in relation to the modelling of ranked data.

Inserting the linear and additive representative utility function $V_{nj} = x_{nj}'\beta + s_n'\delta$, , and applying the standard Gumbel distribution with $\mu=1$, the logit probability becomes:

$$P_{ni} = \frac{e^{x_{nj}'\beta + s_n'\delta}}{\sum_j e^{x_{nj}'\beta + s_n'\delta}} \tag{7.5}$$

7.5 The Rank Logit Model – or the Exploded Logit Model

Discrete choice analysis of binary and multinomial choice sets has been widely used within different branches of behavioural analysis since the 1970s (Domencich & McFadden, 1975; Louviere *et al.*, 2000; Punj & Staelin, 1978). Green & Srinivasan (1978) pointed out that the use of rank order data could give a higher efficiency per unit of time/respondent compared to Choice Experiments (CE). However, it was not until Beggs *et al.* (1981) derived the rank logit model that complete rankings could be analysed in practice. The Rank Logit Model (RL) is based on the concept of sequential choices, which was first investigated in relation to utility theory by Luce & Suppes (1968).

Luce & Suppes (1968) successively decompose the event of a ranking into two probabilities; a *choice event* and a *subranking event.* If the choice set contains four alternatives (a, b, c and d) the model can be expressed as follows:

$$\begin{aligned} &\Pr(a,b,c,d) = \Pr(a|a,b,c,d)\cdot\Pr(b,c,d) \Rightarrow \\ &\Pr(a,b,c,d) = \Pr(a|a,b,c,d)\cdot\Pr(b|b,c,d)\cdot\Pr(c,d) \Rightarrow \end{aligned} \tag{7.6}$$

$$\Pr(a,b,c,d) = \Pr(a|a,b,c,d)\cdot\Pr(b|b,c,d)\cdot\Pr(c|c,d)$$

Stating that the probability of the observed ranking (descending order), $\Pr(a,b,c,d)$ is the product of the probabilities of J-1 *independent* choice events (Chapman & Staelin, 1982). In our case this implies that the probability of the observed ranking, is given by the product of the following three probabilities:

- The probability of the "1st choice" – represented by $\Pr(a|a,b,c,d)$ – where the most preferred of the 4 alternatives is chosen.
- The probability of the "2nd choice" – represented by $\Pr(b|b,c,d)$ – where the secondly most preferred alternative of the three remaining alternatives is chosen.
- The probability of the "3rd choice" – represented by $\Pr(c|c,d)$ – where the most preferred of the two remaining alternatives is chosen

Applying the RUM to 7.6, the probability of the observed ranking becomes:

$$\Pr(U_{na}>U_{nb}>U_{nb}>U_{nb}) = \Pr(U_{nb} >U_{nbcd})\cdot\Pr(U_{nb}>U_{ncd})\cdot\Pr(U_{nc}>U_{nd}) \tag{7.7}$$

Assuming a standard Gumbel distribution for the error terms ε_{ij}, and adopting a *linear additive* specification of the utility function, the rank logit choice probability for individual n is subsequently given by (Beggs *et al.*, 1981):

$$P\left(U_{na} > U_{nb} > U_{nc} > U_{nd}\right) = \frac{e^{x_{na}'\beta+s_n'\delta}}{\sum_a^d e^{x_{nj}'\beta+s_n'\delta}} \cdot \frac{e^{x_{nb}'\beta+s_n'\delta}}{\sum_b^d e^{x_{nj}'\beta+s_n'\delta}} \cdot \frac{e^{x_{nc}'\beta+s_n'\delta}}{\sum_c^d e^{x_{nj}'\beta+s_n'\delta}} \tag{7.8}$$

Comparing 7.8 to 7.5 it is evident that RL choice probability for a rank of four alternatives is equivalent to the product of three MNL choice probabilities. Each of the choice probabilities thus describes one of the sequential choice events illustrated in 7.6. Modelling rank ordered data by the RL is therefore basically equivalent to exploding the full rank into J-1 independent multinomial choice sets, which subsequently are modelled by using the MNL. Accordingly, the term "rank logit" is often used interchangeably with the term "exploded logit" (Ben-Akiva *et al.*, 1991).

The multinomial choice observations resulting from the explosion of rank are, due to their indirect nature, often referred to as pseudo-choice-observations (Train, 2003). An explosion of a rank into J-1 pseudo-choice-observations is referred to as an explosion to the depth of J-1 or a full explosion. However, it is important to note that a rank can be

exploded to different depths. Instead of exploding the rank to the depth of *J*-1, the explosion process can thus be stopped at depth *J*-2, *J*-3 etc., depending on the number of alternatives in the choice set. The fact that a rank can be exploded to varying depths for the purpose of estimation implies that several models can be estimated from the same ranking data.

In the present study, where there are four alternatives in each choice set, three different models, based on different depths of explosion of the rank, can be estimated:

- Firstly, a Rank Logit (RL) model, where the rank is exploded to the depth of three – i.e. resulting in three independent pseudo-choice-observations – can be estimated. Estimation of this model implies that the information from the full rank is used in the modelling; consequently, this model will be referred to as the "RL - all ranks" model.

- Secondly, a RL model, where the rank only is exploded to the depth of two can be estimated. Consequently, this model is only based on the two pseudo-observations that result from respondents' choice of the most preferred, and subsequently the second most preferred, alternative; this model will be referred to as the "RL - 1^{st} + 2^{nd} Choice" model.

- Thirdly, a RL model where the explosion is confined to the depth of one can be estimated. This model is equivalent to the multinomial model that could be estimated from a CE. Since it solely relates to the choice of the most preferred alternatives of the four available alternatives, it will be referred to as the "1^{st} Choice MNL" model.

In relation to the presented models, it is important to note that, despite the fact that the models are based on data obtained from the same ranking, they are in effect based on different data sets. More specifically, as the depth of explosion decreases, the number of resulting pseudo-choice-observations, on which a model subsequently can be estimated, also decreases. Thus, moving from the "RL – all ranks" model, over the "RL – 1^{st} + 2^{nd} Choice" model to the "1^{st} Choice MNL" model, the data set is successively reduced.

Two factors have implications in relation to determining, which of the possible models is appropriate for a given data set; the amount of extractable information and the quality of the information.

In order to take full advantage of the information inherent in the ranking it should be exploded to the depth of *J*-1. Assuming that the pseudo-choice-observations resulting from

the explosion are independent, increasing their number will lead to a better model. The sampling variance of the parameter estimates of the model will be reduced, thereby increasing the accuracy of the obtained parameter estimates (Chapman & Staelin, 1982). More specifically, the advantage of CRM compared to CE in terms of the amount of extractable information, which in section 4.6 was mentioned as the primary reason for choosing CRM in the present study, is contingent upon the explosion to depths greater than one. Thus, limiting the explosion to depth one is, as it appears from the last of the presented models, the amount of information to be obtained from the ranking is no greater than what can be obtained from a CE.

However, with reference to section 4.5.2, the complexity of the CRM value elicitation mechanism is often suspected to have a potentially negative effect on the reliability of the obtained data. More specifically, two potential problems were mentioned: 1) respondents may not use the same decision protocol when assigning different ranks, implying that the determinants of choice vary across ranks, and/or 2) respondents may not be as attentive when assigning ranks to less preferred alternatives as they are when assigning ranks to the most preferred (Ben-Akiva *et al.*, 1991).

Both problems have implications in relation to the extent to which the ranking should be exploded. Consequently, if either of the problems is present, the beta-coefficient vectors – i.e. the parameter estimates for the explanatory variables – will be different across ranks[38]. Pooling of data across all ranks – i.e. exploding the rank in full – will therefore result in a "noisy" choice set, which will negatively affect the fit of the model (Chapman & Staelin, 1982). That is, the accuracy of parameter estimates may decrease rather than increase as a result of the explosion, and the significance/insignificance of variables as determinants of choice may be misrepresented. Prior to fitting models based on ranked data it is therefore important to test for potential instability of coefficient estimates across ranks.

In this connection, however, it is important to distinguish between the two potential causes of instability. If respondents pay less attention to lower ranks, or equivalently are less certain about the lower ranks, the inconsistency arise as a consequence of the fact that the estimated coefficients are confounded with the scale parameter of the error term distribution (Chapman & Staelin, 1982; Hausman & Ruud, 1987). More specifically, if respondents pay less attention to assigning lower ranks, the error terms for lower ranks will have larger variances than the error terms for the higher ranks. The difference in variances is expressed through the scale parameter, implying that the scale parameter for higher ranks is likely to be larger for higher ranks than for lower ranks. With reference to 7.4, this

[38] If people pay less attention to assigning lower ranks, the error terms for lower ranks will have larger variances than the error terms for the higher ranks. This implies, that the IID assumption is violated and that heteroscedasticity, leading to inconsistent parameter estimates, may be introduced in the model (Chapman & Staelin, 1982; Hausman & Ruud, 1987).

implies that the coefficient estimates obtained from models based on choice-observations related to high ranks are expected to be greater in absolute terms than the corresponding coefficient estimates obtained from models based on choice-observations related to lower ranks (Ben-Akiva *et al.*, 1991). Consequently, basing a model on such a data set where the variance varies across observations may imply that the explosion of the rank actually results in decreasing the accuracy of parameter estimates rather than increasing it.

According to Hausman & Ruud (1987) it is often the case that the numerical size of coefficient estimates decrease with increased depth of explosion. However, it is not a problem that necessarily precludes pooling of data across ranks. Thus, if focus is on estimating MRS between different product attributes – e.g. estimation of implicit WTP for certification – it is sufficient if the coefficient vectors are proportional, since this will ensure equality of MRS-estimates across ranks (Hausman & Ruud, 1987). Such inequality, but proportionality of coefficient vectors is likely to be the result of different scale parameters in data sets related to different ranks. In relation to pooling data, this implies that the problem of inequality of coefficient vectors can remedied by rescaling either of the data sets to reflect the scale of the other.

In contrast, if the instability is caused by actual inconsistency[39]- i.e. changing decision protocols across ranks – it implies that data from different ranks will reflect different choice processes. In this case, the problem cannot be remedied, and the explosion should be stopped just prior to the detection of such inconsistency.

Based on the difference between the implications of the different sources of inconsistency, it is also relevant to test for proportionality of beta-coefficient vectors across ranks prior to determining the extent to which the rank should be exploded. Both the test for equality of beta-coefficient vectors, and the test for proportionality of the beta-coefficient vectors can be undertaken by using a Likelihood Ratio test; the more specific procedure of the tests is explained in the chapter 9.

7.6 Limitations of Logit Models

As mentioned the logit model specification builds on the assumptions of Independently Identically Distributed (IID) error terms and Independence of Irrelevant Alternatives (IIA). The more specific implications of the two assumptions in relation to the applicability of logit models will be discussed in the following subsections.

[39] Inconsistent choices, among others, imply that the transitivity assumption presented in chapter 3 is likely to be violated.

7.6.1 The IIIA Assumption

IIA implies that the ratio of choices between two alternatives is independent of the presence or absence of other alternatives and/or attributes of the alternatives. More specifically, the axiom of IIA can be formulated as follows:

$$\frac{P_{ni}}{P_{nk}} = \frac{e^{V_{ni}} \Big/ \sum_j e^{V_{nj}}}{e^{V_{nk}} \Big/ \sum_j e^{V_{nj}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni}-V_{nk}} \tag{7.9}$$

7.9 shows that the ratio of choice probabilities only depends on *i* and *k*.

In some respects, the IIA may seem reasonable. It appears to be reasonable to expect that the ratio between choosing certified as opposed to a conventional pack of toilet paper is independent of the softness of the toilet paper. However, when the alternatives – or some of the alternatives – become too close substitutes, implying that the respondent is more or less indifferent between the alternatives, the assumption of IIA is likely to be violated (Lareau & Rae, 1989). Thus, in relation to the present study, and in equivalence with the much cited *red-bus-blue-bus* problem (Ben-Akiva & Lerman, 1985), a situation where IIA would be problematic, can be exemplified by the introduction of another eco-label in the choice set. If respondents perceive the two eco-labels to be identical, the IIA assumption implies that the model will overestimate the overall probability of choosing an eco-labelled product.

If IIA is violated it implies that the logit model specification is inappropriate in the given context. Thus, violations of IIA imply that the introduction of new alternatives – or the exclusion of alternatives – would change the estimated model. Consequently, the estimated model cannot necessarily be used as a basis for making inferences to other situations. Based on these considerations it is relevant to test for violations of IIA, when using logit models. In the present study, a test developed by Hausman & McFadden (1984) is used to test for violation of IIA. The test is presented in section 9.1.6.

7.6.2 The IID Assumption

A consequence of the IID assumption is that the logit model does not allow random taste variation, implying that it can only represent taste variations that can be linked to observed characteristics of the respondents (Train, 2003). Whether or not this is actually a problem depends on, how well the model is specified (Train, 2003). Thus, if the factors explaining taste variation is sufficiently well accounted for in the model, then the IID assumption is likely to hold. In connection to this study, the presence of random taste variations cannot

be dismissed. Thus, there may be random taste variations in relation to several product attributes, especially qualitative attributes, which the explanatory variables fail to capture. Such random taste was demonstrated in Andersen (2002b) in connection to purchase of non-battery eggs, where the mixed logit, compared to the multinomial logit, changed the estimated WTP for organic eggs drastically.

Apart from the inability of the MNL specification to incorporate random taste variation, the IID also implies that error terms cannot be correlated. In relation to the present study this may be especially important with respect to the table top data. Given the multiple observations per respondent in the ranking of table tops, it may be expected that there is correlation between the error terms for the same respondent. Moreover, with reference to the explosion of rank data into several independent choice observations, it may be argued that correlated error terms are equally likely to be present in the other data sets if the explosion is taken further than to the depth of one.

7.7 Mixed logit

As explained in the previous section, the applicability of the conventional logit model is limited in several ways, which may affect the appropriateness of the model in the present context. An alternative and relatively new logit specification, the mixed logit, allows for random taste variation – i.e. relaxing the IID assumption – and more flexible substitution patterns than those dictated by IIA. Unfortunately, a procedure to handle mixed logit model specifications is only available in a very preliminary, and practically undocumented, edition of SAS (PROC MDC). Consequently, a mixed logit specification cannot be applied in the present study, despite the fact that it would probably have been more suitable.

7.8 Maximum Likelihood Estimation

The Maximum Likelihood Estimation (MLE) method is used to estimate the different logit models. The maximum likelihood expression is different from model to model. In short MLE identifies the set of parameters, which generates the observed choice behaviour most often. The likelihood function for a general multinomial logit model is given by (Ben-Akiva & Lerman, 1985):

$$L = \prod_{n=1}^{N} \prod_{i \in C} P_n(i)^{y_{in}} \tag{7.10}$$

where C_n denotes the choice set from which i is chosen, and where y_{in} equals 1 if individual n choses alternative i, and zero otherwise. The likelihood for the whole sample is equal to the product of the likelihood's of the individual observations. This is a result of the assumption that the individual observations are independent.

With: $P_{ni} = \frac{e^{\beta' x_{in}}}{\sum_j e^{\beta' x_{jn}}}$ the log likelihood function becomes:

$$LL = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \left(\beta' x_{in} - \ln \sum_{j \in C_n} e^{\beta' x_{jn}} \right) \quad (7.11)$$

The log likelihood function is maximised by differentiating the function with respect to the β's and setting the partial derivatives equal to zero. That is:

$$\frac{\partial LL}{\partial \beta} = 0 \quad (7.12)$$

7.9 Deriving Estimates of Marginal Willingness to Pay

As explained in chapter 3, the relevant welfare measure is a Compensating Surplus (CS) measure. The maximum WTP for certification can therefore be defined as the amount of money (i.e. difference in price), which makes the individual indifferent between purchasing a certified product and an otherwise identical conventional product.

Following Hanemann's (1984) derivation of WTP in a RUM, the probability of choosing a certified product is given by:

$$P_1 (V_1 + \beta p_1 + \varepsilon_1 > V_0 + \beta p_0 + \varepsilon_0) = P_1 (V_1 - V_0 + \beta (p_1 - p_0) > \varepsilon_0 - \varepsilon_1) \quad (7.13)$$

Where V_1 represents the utility associated with the purchase of a certified product, excluding the price on the product p_1; V_0 represents the utility associated with an identical conventional product, excluding the price of the product p_0; β is the parameter of price (<0) and where $p_1 > p_0$.

When the probability in 7.13 is equal to 0.5, the individual is indifferent between the two products, and then $p_1 - p_0$ is equal to the maximum WTP. If $\varepsilon_0 - \varepsilon_1$ is symmetric around this WTP, then:

$$V_1 - V_0 + \beta (p_1 - p_0) = 0 \Rightarrow V_1 - V_0 + \beta\, WTP = 0 \Rightarrow$$
$$(V_1 - V_0)/-\beta = WTP \quad (7.14)$$

Since the only difference between $V_1 - V_0$ is whether or not the wood product is certified, the derived WTP is equal to the maximum WTP for certification.

Alternatively, the WTP can be derived by differentiating the estimated utility function, *V,* with respect to the variables; certification and price. Since the properties of CS imply that utility is held constant, the derivatives are set equal to zero. Subsequently, WTP is given by:

$$V = X + \alpha CERT + \beta PRICE \Rightarrow \partial V = \alpha + \beta = 0 \Rightarrow$$
$$\alpha / -\beta = WTP \qquad (7.15)$$

Where V is the observed utility function, X is the vector of other determinant variables, α is the coefficient estimate of certification (equivalent to the marginal utility of certification) and β is the coefficient estimates of price (equivalent to the marginal value of price/money). By dividing the marginal utility of certification with the marginal utility of price, an estimate of the value of certification in monetary units is obtained. Thus, the WTP for certification – expressing the MRS between certification and price – is derived by dividing the coefficient estimate for certification by the coefficient estimate of price. Similarly, the WTP for the other products' attributes can also be derived by dividing their coefficient estimates with that of price, and likewise the MRS between the different attributes can be identified by division of their respective coefficient estimates.

7.10 Summary

Upon the presentation of the random utility model and the multinomial logit model, the rank logit model is derived. Modelling ranked data by the rank logit model is shown to be equivalent to exploding the rank into several independent multinomial choice observations. In this connection, it is found that lower ranks may be associated with inconsistency and increasing variances, implying that it may not be appropriate to use all the information inherent in the full rank. Furthermore, the implications of the assumptions of independence of irrelevant alternatives and independent identically distributed error terms implicit in the logit model specification are discussed. Subsequently, the maximum likelihood estimation procedure is briefly introduced, and lastly, it is specified how estimates of willingness to pay can be derived from the chosen model.

8 The Data Set

In the present chapter, it will be described how the sample was identified. Thereafter, the extent to which the sample is representative of the Danish population will be assessed.

8.1 Sample Size and Selection

In the planning of the survey, the approximate variable costs per respondent were estimated to be approximately 30 DKK, and the fixed costs of carrying out the survey were estimated to be in the vicinity of 15.000 DKK. With 45.000 DKK available for the completion of the survey, the maximum total sample size was estimated to be approximately 1.000 respondents, equivalent to a maximum of 500 respondents for each sub-sample.

8.1.1 Sample Size

The choice of sample size was made with reference to the number of choice sets resulting from the blocking of the fractional factorial designs for the three products. Thus, each choice set should be represented an equivalent number of times both within the total sample and within each sub-sample. With 9 and 6 choice sets for toilet paper and cutting boards, and with 45 different combinations of choice sets for table tops[40], the numbers of respondents, which would secure equal representation of choice sets, was found to be multiples of 90. Since the maximum sample size for each sub-sample was set to 500, 450 was identified to be the relevant number of respondents for each sub-sample, equivalent to a total sample size of 900.

8.1.2 Sample Selection

It was a desire to obtain a sample, which as accurately as possible would reflect the Danish population. However, due to financial limitations, it was not possible to draw a completely random sample from the Danish population. Instead, a systematic stratified selection procedure was used, where strata were defined by the birth dates and the gender of individuals. More specifically, 5 birth dates were chosen for each sub-sample, and for each of these dates, 45 women and 45 men were identified with reference to their CPR-number. The birth dates chosen for each of the sub-samples are presented in table 8.1.

[40] With 10 different choice sets, and each respondent being presented with 2 different choice sets, this result in $(9 \cdot 10)/2 = 45$ different combinations.

Table 8.1 Birth dates determining the composition of the samples.

Birth dates for sample 1	Birth dates for sample 2	Representative of age group:
July 15th 1937	August 15th 1937	> 60 years
July 15th 1947	August 15th 1947	50 - 60 years
July 15th 1957	August 15th 1957	40 - 50 years
July 15th 1967	August 15th 1967	30 - 40 years
July 15th 1977	August 15th 1977	< 30 years

8.1.3 Sample Size and Model Reliability

In relation to establishing a link between sample size and reliability of the resulting model, Gyrd-Hansen (2002) recommended that, when using a main-effects design, as in this study, each alternative should be presented at least 30 times in order to secure the reliability of the estimated model. Theoretically - i.e. if the response rate is 100% - the chosen sample size implies that: a) each toilet paper alternative is represented 100 times in the total sample; b) each cutting board alternative is represented 150 times in the total sample, and c) each table top alternative is represented 90 times. Thus, for the total sample, the response rate has to drop below 30% before the recommendation of at least 30 representations of each alternative is violated[41], implying that the chosen sample size should be sufficiently large to allow for the estimation of a reliable model[42].

8.2 Response Rates and Effective Sample Sizes

Of the 900 questionnaires that were sent out on the 13th of February[43], 315 had been returned by the 7th of March where a reminder was mailed to those respondents who had not replied. Subsequently, an additional 82 questionnaires were returned, thereby increasing the total number of responses to 397, which is equivalent to a response rate of 44,1%.

This may appear to be a very low response rate, but according to Pearce *et al.* (2002) response rates of mail surveys are most commonly in the interval 25-50%. Seen from this perspective the response rate can only be considered most satisfactory. Considering the facts that: *a*) the questionnaire was mailed to "representative" Danes (as opposed to people with a special interest in the subject) and *b*) only one reminder (and no new questionnaire) was sent out the experienced response rate is thus perceived to be very good.

[41] That the recommendation will not be violated for any alternatives until the response rate drops below 30% is obviously contingent upon the fact that the non-responses are equally spread over all choice sets.
[42] In appendix III, the number of times that each block/alternative was presented in the effective sample is documented.
[43] The questionnaires were deliberately, planned to be sent out on a Thursday, implying that people would receive them just before a weekend. It was expected that this would be the time of the week where people were most likely to have time, and be mentally prepared, to fill out the questionnaire.

Of the received responses, a few were not filled out, and a few were discarded due to the question pertaining to income not had been filled out. This implies that the effective samples are a bit smaller than indicated by the response rates. In table 8.2 the response rates, and resulting effective sample sizes are specified. The table also specifies the number of respondents, who has provided valid responses for the separate rankings. The classification of some responses as "non-valid" has been based on inspection of reasons stated for not wanting to buy the product assigned rank 1. Respondents stating that they would not consider buying the product ranked 1 were singled out, and their reasons for not wanting to buy were scrutinized. Subsequently, respondents whose answer indicated that they had not understood, or accepted, the "rules" of the ranking were excluded[44]. Upon comparison of the size of the effective sample and the number of respondents providing valid responses, between 6 and 22 respondents were excluded from the rankings[45].

Table 8.2 Specification of response rates and effective sample sizes.

	Sample 1 High Info Level		**Sample 2 Low Info Level**		**Total Sample**	
	No.	%	No.	%	No.	%
Sent out	450	100	450	100	900	100
Returned	200	44,4	197	43,7	397	44,1
Rankings completed or partially completed	197	43,7	190	42,2	387	43,0
Income stated (= effective sample)	192	42,7	184	40,9	376	41,8
Proportion of effective sample, providing valid responses for each separate ranking:						
Rank A	186	96,9	178	96,7	364	96,8
Rank B	182	94,8	172	93,5	354	94,1
Rank C	-	-	170	92,4	-	-
Rank D	-	-	170	92,4	-	-
Rank E	-	-	161	87,5	-	-

8.3 Representativity of the Sample

A prerequisite for making inferences from our sample to the Danish population in general, is that the sample resembles the population. In the following sections, the representativity of the sample with respect to the socio-economic characteristics that are included in the modelling – i.e. age, sex, education, income, geographical location, frequency of forest visits and type of residence – will be investigated.

[44] This is more or less equivalent to the exclusion of protest-bidders in CVM surveys. Thus, the respondents that were excluded were those whose answers indicated an unwillingness – or inability- to make the necessary trade-offs when assigning ranks.

[45] In a few cases, the reductions from the effective sample to the number of usable responses are caused by the fact that some respondents have completed only 1, 2 or 3 of the rankings.

When the response rate is below 50%, Andersen (2002a) recommends that it is also investigated, if the respondents who have returned the questionnaire are representative of the original sample. A prerequisite for undertaking such analyses is access to the relevant background information on the respondents failing to respond. In our case, only information related to the age, gender and geographical location of the respondents in the original sample was available. Accordingly, it is only in relation to these characteristics, that it can be investigated whether or not the effective sample is representative of the original, full sample.

8.3.1 Representativity in Terms of Age

In table 8.3 the age distribution of the respondents in the effective sample is compared with the age distribution in the original sample and the Danish population. When comparing the effective sample with the original sample it appears that participation in the survey declines with age. However, as the test result shows, the age distribution in the effective sample does not differ significantly from the age distribution in neither the original sample nor the population in general.

Table 8.3 Comparison of the age distribution in the effective sample with the age distribution in the original sample and the Danish population*.

Age group (years)	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	CPR (%)	Danish Population (%)
20-29	24,0	22,8	23,4	20,0	20,0
30-39	19,3	21,2	20,2	20,0	23,0
40-49	17,7	20,1	18,9	20,0	21,0
50-59	19,3	20,7	19,9	20,0	22,0
60-69	19,8	15,2	17,6	20,0	14,0
Total	100	100	100	100	100
Test for statistically different distributions: Effective sample vs. CPR; $\chi^2 = 3{,}4930^{NS}$					
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 8{,}3624^{NS}$					

*Danmarks Statistik (2002a).

8.3.2 Representativity in Terms of Gender

In table 8.4 the composition of the effective sample in terms of gender is compared with the original sample and the Danish population. The test results in the last two rows of the table show that there are no significant differences – neither between the effective sample and the original sample, nor between the effective sample and the Danish population.

Table 8.4 Comparison of gender distribution in the effective sample with that in the original sample and that in the Danish population*.

Gender	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	CPR (%)	Danish Population (%)
Female	51,0	48,4	49,7	50,0	49,7
Male	49,0	51,6	50,3	50,0	50,3
Total	100	100	100	100	100
Test for statistically different distributions: Effective sample vs. CPR; $\chi^2 = 0{,}0106^{NS}$					
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 0{,}0009^{NS}$					

* Danmarks Statistik (2002a).

8.3.3 Representativity in Terms of Region

The regional distribution of respondents – i.e. determined by their area of residence – is compared to the regional distribution of respondents in the original sample and in the Danish population. The test results, which are presented in table 9.5, show that the effective sample can be considered representative of the original sample but not of the population in general. Compared to the population, there is an overrepresentation of individuals living in the Copenhagen area and a corresponding under representation of individuals living in rural areas.

Table 8.5 Comparison of regional distribution of respondents in the effective sample with that of individuals in the original sample and the Danish population*.

Region	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	CPR (%)	Danish Population (%)
CPH	22,9	22,8	22,9	23,9	11,5
Urban areas - the islands.	13,5	15,2	14,4	13,2	15,5
Urban areas.- Jutland	22,9	23,4	23,1	22,4	22,9
Rural areas	40,6	38,6	39,6	40,4	50,1
Total	100	100	100	100	100
Test for statistically different distributions: Effective sample vs. CPR; $\chi^2 = 0{,}675^{NS}$					
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 50{,}55^{***}$					

* Danmarks Statistik (2003a) and Danmarks Statistik (2002b).

8.3.4 Representativity in Terms of Income

In table 8.6 the percentage of households in the effective sample, and in the population, belonging to three intervals for yearly household income is presented. The table show that there is a strong overrepresentation of high-income households, and the test confirms that the difference is highly significant.

Table 8.6 Comparison of household income distribution in the effective sample with the household income distribution in the Danish population*.

Yearly Household Income interval	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	Danish Population (%)
0-199.999 DKK	15,6	13,0	14,4	32,5
2-400.000 DKK	28,1	31,0	29,5	29,9
> 400.000 DKK	56,3	56,0	56,1	37,6
Total	100	100	100	100
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 72{,}1471^{***}$				

* Danmarks Statistik (2003b).

Since income is expected to be a potentially important factor in determining WTP for certification, the overrepresentation of high-income households is unfortunate. It is expected that this might result in overestimation of true WTP. An observation that may serve to explain the overrepresentation of high-income households is that individuals have been sampled rather than households. This implies that households with two adults are

twice as likely to be represented in the sample as households with one adult[46]. Since the average yearly household income is higher for households with two as opposed to one adult (Danmarks Statistik, 2002c), the overrepresentation may therefore be seen as a natural consequence of the applied sampling procedure. Thus, the observed overrepresentation does not necessarily imply that respondents of high-income households have had a higher propensity to participate in the survey – it may just be a reflection of the composition of the original sample.

8.3.5 Representativity in Terms of Education

With reference to table 8.7, where the effective sample is compared to the population in terms of educational level, there is a strong – and statistically very significant – overrepresentation of individuals with a high level of education. This may be explained by the fact, that people with a high level of education has found it less demanding to complete the questionnaire. Similarly, they may have felt more compelled to contribute to the successful completion of a scientific project.

To some degree, educational level and income level are expected to be correlated. With this in mind, the overrepresentation of individuals with a high level of education may serve to explain the observed overrepresentation of high-income households. This, in turn, may be taken to indicate that the overrepresentation of high-income households is not just a consequence of the composition of the original sample, as suggested in the previous section.

Table 8.7 Comparison of the educational level of respondents with that of the Danish population* (people between 30 and 69 years of age).

Education	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	Danish Population (%)
Primary, or not stated	8,9	12,0	10,4	34,0
High school	4,7	4,9	4,8	4,1
Practical	30,2	29,4	29,8	38,5
1-2 years of sec. edu.	10,9	11,4	11,2	4,6
3-4 years of sec. edu.	29,2	30,4	29,8	13,3
5-6 years of sec. edu.	16,2	12,0	14,1	5,7
Total	100	100	100	100
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 227{,}8741^{***}$				

* Danmarks Statistik (2002c).

[46] This observation is supported by the fact that 60,1% of the households represented in our sample have two adults and 20% have one. These percentages can be compared with around 50% for each of the household types in reality (Danmarks Statistik, 2002a).

8.3.6 Representativity in Terms of Forest Visits

Part of a survey conducted by Aakerlund (1998) focused on the frequency of forest visits, and the reliability of respondents reported visit frequency. The results from the Aakerlund (1998) survey serve as the base of comparison for evaluating the representativeness of the effective sample in terms of frequency of forest visits. The comparison is presented in table 8.8, where it appears that the respondents in the effective sample can be considered to be representative of the Danish population in terms of frequency of forest visits.

Table 8.8 Comparison of forest visit frequency of respondents with the frequency found in Aakerlund (1998).

Frequency of visits	Visits per year	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	Danish population (%)
1-2 visits per week	110	23,4	20,8	22,4	24,3
1-2 visits per month	18	30,2	30,0	30,1	28,7
1 visit every 3rd month	4	18,2	21,9	20,0	22,9
1-2 visits per year	1,5	21,4	19,7	20,5	17,4
Almost never	0,5	6,2	7,7	6,9	6,7
Total	-	100	100	100	100
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 4{,}399^{NS}$					

8.3.7 Representativity in Terms of Residence Type

In table 8.9 the effective sample is compared to the population in general in terms of residence type. The respondents in the effective sample are not representative of the population, since there is an overrepresentation of individuals owning their homes whereas individuals living in rental homes are underrepresented.

Table 8.9 Comparison of residence type of respondents with that of the Danish population*.

Residence type	Sample 1 High Info Level (%)	Sample 2 Low Info Level (%)	Total sample (%)	Danish Population (%)
Owned	69,3	70,1	69,7	61,5
Rental	22,4	23,4	22,9	32,7
Co-op	8,3	6,5	7,4	5,8
Total	100	100	100	100
Test for statistically different distributions: Effective sample vs. DK; $\chi^2 = 16{,}9537^{***}$				

* Danmarks Statistik (2003d).

8.4 Representativity of the Danish Population

The results presented in the previous sections show that it is only in terms of age, gender and frequency of forest visits that the effective sample can be considered representative of the population in general. This observation, of course, has implications in relation to determining the extent to which inferences can be made from the sample to the population. More specifically, one should be careful when extending the result derived from the sample to conclusions regarding the entire Danish population.

8.5 Original Sample vs. Effective Sample

The representativity of the effective sample in relation to the original sample can only be investigated in relation to age, gender and region of residence. In all three cases, representativity was established. However, it is suspected that the effective sample may fail to be representative of the original sample in terms of education, and consequently, also in terms of income and type of residence due to correlations among these characteristics. In other words, it seems, that individuals with a low level of education have been less inclined to participate due to the perceived complexity of the task. However, since no tests can be made, this is merely speculation.

8.6 Summary

The sample size of the study was determined to be 900 respondents; of these 44,1% returned the questionnaire. The effective sample is found to be representative of the original sample with regards to gender, age and region of residence. Similarly, the effective sample is found to be representative of the Danish population in terms of gender, age and frequency of forest visits. However, in terms of region of residence, income, education and residence type, it is found that the effective sample cannot be considered representative of the Danish population.

9 The Analysis

The present chapter contains a thorough description of the analysis of data. This approach has been chosen, since no other studies we have come across have provided detailed descriptions of the different steps in the analysis of rank data. Therefore we find it most relevant to provide a coherent presentation of the analysis of data related to the models for choice of toilet paper, cutting boards and table tops. To provide an overview of the analyses from which the final models are derived, section 9.1 contains a detailed description of the 7 different steps of the analyses. Section 9.2 presents the software used for the analysis. The subsequent three sections, 9.3, 9.4 and 9.5, contain the analysis of the data pertaining to the three different product types.

9.1 The Steps of the Analysis

The analysis of the data can be partitioned into 7 steps. In the following 7 subsections the background along with the content and purpose of each step is presented.

9.1.1 Step 1: Equality of beta-coefficient vectors across information scenarios

The first step of the analysis is to test for equality of beta-coefficient vectors across the data sets pertaining to the two different information scenarios. The result of the test will determine whether or not the data from the two scenarios can be pooled in the further analysis. Thus, if the beta-coefficient vectors that can be estimated from the two sub-samples are significantly different, it will indicate that the determinants of choice are different in the two sub-samples. Consequently, it would be inappropriate to pool the data and estimate a joint model.

Since the ranking of table tops was only presented in one of the information scenarios, step 1 is only relevant in relation to toilet paper and cutting boards.

As mentioned in chapter 7, a Likelihood Ratio test (LR-test) can be used to test for equality of beta-coefficient vectors across data sets from two sub-samples (Ben-Akiva *et al.*, 1991). The test statistic, which test the hypothesis that $\hat{\beta}_{\text{data set 1}} = \hat{\beta}_{\text{data set 2}}$, is given by:

$$- 2 (L_R - L_U)$$

where L_R is the log-likelihood value of the model estimated from the pooled data set (implying that the beta-coefficient vector is restricted to be identical across the two datasets, i.e. $\hat{\beta}_{\text{data set 1}} = \hat{\beta}_{\text{data set 2}}$), and L_U is the log-likelihood value of the unrestricted model, which is given by the sum of the log-likelihoods of the models estimated separately from the separate data sets (in our case two) (Chapman & Staelin, 1982). The test statistic is χ^2 distributed with $K_U - K_R$ degrees of freedom, where K_U and K_R are the number of estimated parameters in the unrestricted (the sum of parameters estimated in each separate

model) and restricted models, respectively (Ben-Akiva *et al.*, 1991). For a brief review of the test used in Step 1, see box 9.1

Test:	Test for equality of coefficient vectors across data set
Purpose:	To test if data can be pooled across information scenarios
Hypothesis:	$\beta_{\text{data set low information}} = \beta_{\text{data set high information}}$
Test:	$-2(L_R - L_U) \Rightarrow$
	$-2\ [L_{\text{pooled data set}} - (L_{\text{data set low information}} + L_{\text{data set high information}})]$
	χ^2 distributed, with $K_U - K_R$ degrees of freedom,

Box 9.1 Test used in Step 1.

9.1.2 Step 2: Attribute-only models

The primary purpose of the second step is to determine the depth to which the ranking can be exploded. Thus, as mentioned in chapter 7, instability of coefficient estimates across ranks may imply that the extent of explosion should be limited to depths less than *J*-1.

Inspired by the approach adopted in Morrison & Boyle (2001) and Foster & Mourato (2002), five different models based on different sub sets of the full data set, and with product attributes as the only explanatory variables, are estimated. In addition to the three models presented in chapter 7, which represent explosions of the rank to varying depths, the estimated models are:

- a 2nd Choice MNL model estimated on data related to the choice of the most preferred alternative from the choice set, remaining after the 1st Choice have been made.

- a 3rd Choice MNL model estimated on data related to the choice between the two alternatives remaining after the 1st and 2nd Choice have been made.

The main purpose of estimating these two models is, that they are used in the tests for equality of beta-coefficient vectors across ranks. Thus, they are not by themselves considered relevant candidates for the final model.

In relation to all the models estimated in the analysis, it should be noted that the models are fitted without an intercept. The inclusion of an intercept would therefore be equivalent to assuming that one or more important attributes, with a constant contribution to utility

across all products, are not included in the model (Gyrd-Hansen, 2003; Bockstael, 2003). Moreover, in connection with toilet paper and cutting boards, the potential effect of information on the coefficient of the certification attribute is incorporated in the models by constructing an information dummy. The information dummy variable designates, which information level a given observation relates to and is entered in the model by interacting it with the certification attribute. Subsequently, the coefficient estimate for this interaction effect will reflect the effect of information on the utility derived from certification.

For each model, McFadden's Likelihood Ratio Index (LRI), ρ^2 and the adjusted LRI, $\bar{\rho}^2$, are calculated. These statistics are given by:

$$\rho^2 = 1 - \frac{L(\hat{\beta})}{L(0)} \quad \text{and} \quad \bar{\rho}^2 = 1 - \frac{L(\hat{\beta}) - K}{L(0)}$$

where $L(\hat{\beta})$ is the value of the log likelihood function at the estimated parameters, $L(0)$ is its value when all the parameters are set to zero and *K* is the number of estimated parameters. The LRI's are used as goodness-of-fit measures in discrete choice models analogous to the use of R^2 in conventional regression analysis (Ben-Akiva & Lerman, 1985; Greene, 2003). However, its interpretation is quite different; where R^2 specifies the percentage of variation explained by the estimated model, the LRI's specify the percentage increase in the log-likelihood function above the value at $L(0)$. Thus, the LRI's do not have the same clear meaning as R^2 since the value of the LRI is determined as a function of the estimated model rather than as a function of the data[47] (Train, 2003).

Generally discrete choice models must have a ρ^2 value higher than 0.1 to be perceived acceptable (Pearce *et al.*, 2002), and a value between 0.20 and 0.40 as an indicator of extremely good model fits (Louviere *et al.*, 2000).

There seems to be some disagreement concerning the applicability of the LRI's to assess the fit of models. Ben-Akiva & Lerman (1985) highlight the fact that ρ^2 suffers from the same shortcoming as R^2, namely that it always will increase when additional variables are added. Based on this observation, they recommend the use of the adjusted LRI, which takes the number of variables into account. According to Train (2003), the LRI can only be used to compare models that have the same $L(0)$ - i.e. models that are based on exactly

[47] Another goodness-of-fit measure sometimes used in connection with discrete choice models is the "percent correctly predicted". According to Train (2003), however, this measure should be avoided, since it is inconsistent with the way that the estimated probabilities should be interpreted. Thus, the "percent correctly predicted" measure assumes that, in a given choice situation, it is always the alternative with the highest probability that will be chosen. The relevant interpretation, on the other hand, is that if the given choice situation is repeated a number of times, then the alternative with the highest probability will be the one that is chosen the greatest number of times. Based on this, we have not found it relevant to make use of this goodness-of-fit measure.

the same dataset. Even in that case, however, the usefulness of the LRI is limited, since it provides no additional information to what can be obtained by comparing the $L(\hat{\beta})$'s of the different models. Chapman & Staelin (1982), on the other hand, suggest that LRI may be used to determine the depth to which a rank data can be exploded. They argue, that since LRI does not depend on the number of observations, the LRI should remain approximately constant as the depth of explosion increases. If the LRI starts to decrease, as additional pseudo-observations are added, it would imply that the added observations are "noisy". And, in that event, the explosion process should be terminated just prior to the last explosion.

In addition to comparisons of the LRI's for the different models, a LR-test identical to the one used in step 1 is also used to determine the appropriate extent of explosion. The LR-test is used to test if there is a significant difference between the coefficient vectors across ranks (Ben-Akiva *et al.*, 1991; Hausman & Ruud, 1987).

As discussed in chapter 7, the extent to which inequality of coefficient estimates preclude pooling of data depends on the cause of the instability. Consequently, if the LR-test leads to a rejection of the hypothesis of equal coefficient vectors across ranks, it is then relevant to test for proportionality of coefficient vectors. Thus, it is not until the hypothesis of proportionality also has been rejected that it can be concluded definitively, that pooling of data across ranks is inappropriate.

A LR-test can be used to test the hypothesis that the coefficient vectors are proportional (Ben-Akiva *et al.*, 1991). If they are proportional, equality of coefficient vectors can be obtained by rescaling one of the data sets. The hypothesis to be tested is given by: $\beta^{m+1} = \mu_m \beta^m$, where m specifies the rank that is modelled. Since data pertaining to m+1 is likely to be more "noisy" than the data pertaining to m, the parameters of data set m+1 are likely to be smaller in absolute magnitude than those of data set m[48]. Therefore, it is expected that the value of μ_m - the rescaling parameter - will be somewhere between zero and one.

The test is undertaken by multiplying all variables in data set m+1 are by μ_m, and subsequently, estimating a joint model for the m and m+1 data sets. The joint model is estimated with different values of μ_m, and the likelihood values of the resulting models are compared. The μ_m, which gives the maximum likelihood value is the relevant scale parameter. Subsequently, the LR-test statistics is obtained by comparing the rescaled joint model with the separate models estimated from the data sets m and m+1, respectively. For a review of the tests used in step 2, see Box 9.2.

[48] The difference in parameter estimates is caused by the confounding between the scale parameter and coefficient estimates mentioned in section 7.3.

Test: Test for equality of coefficient vectors across ranks

Purpose: Determine the depth of the explosion of the rank

Hypotheses:

a) $\beta_{1st\ choice} = \beta_{2nd\ choice} = \beta_{3rd\ choice}$
b) $\beta_{1st\ choice} = \beta_{2nd\ choice}$
c) $\beta_{2nd\ choice} = \beta_{3rd\ choice}$

Test stat.: $2(L_R - L_U) \Rightarrow$

a) $-2\ [L_{RL\ -\ all\ ranks} - (L_{1st\ choice} + L_{2nd\ choice} + L_{3rd\ choice})]$
b) $-2\ [L_{RL\ -\ 1st\ +\ 2nd\ choice} - (L_{1st\ choice} + L_{2nd\ choice})]$
c) $-2\ [L_{RL\ -\ 2nd\ +\ 3rd\ choice} - (L_{2nd\ choice} + L_{3rd\ choice})]$

χ^2 distributed, with $K_U - K_R$ degrees of freedom

Test: Test for proportionality of coefficient vectors across ranks

Purpose: Determine whether respondents have made consistent choices across ranks

Hypotheses:

a) $\beta_{1st\ choice} = 1/\mu_{12}\beta_{2nd\ choice} = 1/\mu_{13}\beta_{3rd\ choice}$
b) $\beta_{1st\ choice} = 1/\mu_{12}\,\beta_{2nd\ choice}$

where μ_{12} and μ_{13} refer to the scale parameters between $\beta_{1st\ choice}$-$\beta_{2nd\ choice}$ and $\beta_{1st\ choice}$ - $\beta_{3rd\ choice}$ respectively.

Test stat.: $-2(L_R - L_U)$

a) $-2\ [L_{RL\ -\ all\ ranks} - (L_{1st\ choice} + L_{rescaled\ 2nd\ choice} + L_{rescaled\ 3rd\ choice})]$
b) $-2\ [L_{RL\ -\ 1st\ +\ 2nd\ choice} - (L_{1st\ choice} + L_{rescaled\ 2nd\ choice})]$

χ^2 distributed, with $K_U - K_R$ degrees of freedom

Box 9.2 Tests used in Step 2.

9.1.3 Step 3: Univariable Analysis of Potentially Relevant Variables

Based on the hypotheses of potentially influential variables relating to socio-economic or attitudinal respondent characteristics, univariable analysis of each potentially relevant variable is performed as the third step of the analysis. The results of the univariable analysis determine, which variables to include in the multivariable analysis.

In step 2, the choices are modelled solely as a function of the attributes of the alternatives. However, it is likely that different characteristics of the respondent also may be important

factors in explaining choices. The respondent characteristics, which enter as variables in the analysis, are presented in table 9.1. All variables entering the analysis are constructed in the form of dummy variables. If the variable is represented by two different levels, e.g. the variable sex (male and female), one dummy variable is constructed; if the variable has three levels, e.g. income, two dummy variables are constructed, etc.

Table 9.1 List of respondent characteristics variables used in the analysis.

Variable	Definition	Contrast category	Ref. to question
Socio-Economic Variables			
INC1	Yearly household income between 200.000 dkr and 400.000 dkr.	HH income < 200.000	8.10
INC2	Yearly household income > 400.000 dkr.		
SEX	1 if male; 0 otherwise.	Female	8.1
AGE1	1 if respondent is between 40 and 60 years of age; 0 otherwise.	< 40 years old	8.1
AGE2	1 if respondent is older than 60 years; 0 otherwise.		
EDU1	1 if 2-4 years of secondary education; 0 otherwise.	< 2 yrs. Sec. Edu.	8.2
EDU2	1 if 5-6 years of secondary education; 0 otherwise.		
ACCO1	1 if owner of accommodation; 0 otherwise.	People living in rental home	8.4
ACCO2	1 if living in ”co-op” house/flat; 0 otherwise.		
KIDS	1 if there are children below the age of 18 in the household; 0 otherwise.	HH’s with no kids	8.9
REG	1 if resident of Copenhagen County; 0 otherwise.	People living in rural districts	Zip-code (cf. CPR)
REG1	1 if resident of a municipality on Zeeland or the Islands; 0 otherwise.		
REG2	1 if resident of a municipality in Jutland; 0 otherwise (Ditto).		
Attitudinal Variables			
Trust in Ecolabels			
TRUST1	1 if the respondent has trust in ecolabels on products produced in Denmark; 0 otherwise.	People without trust	2.1
TRUST2	1 if the respondent has trust in ecolabels on products produced in other EU countries; 0 otherwise.		2.2
TRUST3	1 if the respondent has trust in ecolabels on products produced in LDCs; 0 otherwise.		2.3
Perception of Forest Management			
FOR11	1 if agreeing in fully or partly to the statement that Danish forests are being managed sustainably; 0 otherwise.	People who disagree	3.1
FOR12	1 if “don’t know” whether Danish forests are being managed sustainably; 0 otherwise.		
FOR21	As FOR11, but for other European forests.	As FOR11	3.2
FOR22	As FOR12, but for other European forests.	As FOR12	
FOR31	As FOR11, but for forests in LDCs.	As FOR11	3.3
FOR32	As FOR12, but for forests in LDCs.	As FOR12	

Variable	Definition	Contrast category	Ref. to question
Interest in the environment			
ENV1	1 if the respondent has average interest in environmental issues; 0 otherwise.	People with little interest.	8.8
ENV2	1 if the respondent has great interest in environmental issues; 0 otherwise.		
Frequency of forest visits			
VFOR1	1 if visiting forests at least once a week; 0 otherwise.	People visiting less often.	8.5
VFOR2	1 if visiting forests at least once a month, but not quite once a week; 0 otherwise.		

Since respondent characteristics do not vary over the alternatives presented to each respondent they cannot enter the model directly. That is, since choices are modelled in terms of differences, they will factor out when transforming the choice observation data into differences. To overcome this problem, the effect of respondent characteristics on utility is entered in the model by interacting them with product attributes. Hereby, a model specification, where the effects of respondent characteristics vary over alternatives, is obtained (Gyrd-Hansen & Søgaard, 2001; Train, 2003). Subsequently, the coefficient of the interaction variable will express the influence that the characteristics have on the utility derived from the associated attribute. That is, it will express the additional (if positive) or reduced (if negative) amount of utility derived from the product attribute by individuals belonging to the category coded as 1. For the individuals belonging to the category coded as 0, on the other hand, the utility derived from the particular product attribute is equal to the coefficient of the attribute.

With several product attributes and an even greater number of socio-economic and attitudinal characteristics the possible number of interaction terms that can be defined is high. It is unlikely that all will be equally relevant. According to Hosmer & Lemeshow (2000), interaction terms[49] that do not make sense from a theoretical or common-sense perspective should not be included in the model. In the present study, focus is on uncovering preferences, and subsequently WTP, for certification. Therefore, emphasis will be on the interplay between respondent characteristics and the utility derived from certification. Based on these considerations, hypotheses of plausible and potentially important interactions have been identified. Following the approach adopted in Gyrd-

[49] It should be noted that Hosmer & Lemeshow (2000) refer to interaction terms between main effects. However, the conclusion is equally applicable in terms of interactions between main effects and respondent characteristics. Thus, the specification of respondent characteristics in terms of interactions with product attributes is actually equivalent to a specification where the respondent characteristics enter the model both as a "main effect" and through the interaction; however, the "main effect" factors out in the modelling as described in the text.

Hansen & Søgaard (2001), the hypotheses are presented in table 9.2, thereby providing a comprehensive overview of the interactions being considered in the model.

In table 9.2, the first column specifies the interaction effect in question. The letter preceding the underscore denotes the attribute, with which the respondent characteristic is interacted (C = the certification attribute, P = the price attribute) and following the underscore the respondent characteristic is specified (ref. to table 9.1). The second column specifies the label for the estimated coefficient, and the third and fourth columns specify the hypotheses mathematically and verbally, respectively. As it appears, it is not in all cases possible to hypothesize about the sign of the effects of both interaction effects and main effects on utility.

Table 9.2 Hypotheses of potentially important interaction effects.

Attribute / interaction effect	**Para-meter**	**Hypotheses - mathematically (Utility denoted U)**	**Hypotheses - in words**	**Significant in sub-sample**		
				A	**B**	**C**
General hypotheses socio-economic factors						
P_INC1	$\beta_{P\text{-}I1}$	dU/d(P_INC1) > 0	The disutility of cost is lower for middle income groups than for low income groups.	+	-	+
P_INC2	$\beta_{P\text{-}I2}$	dU/d(P_INC2) > 0	The disutility of cost is lower for high income groups than for low income groups.	-	-	+
P_SEX	$\beta_{P\text{-}S}$	dU/d(P_SEX) >0	The disutility of a price increase is less for women than for men.	-	+	-
C_INC1	$\beta_{C\text{-}I1}$	dU/d(C_INC1) > 0	People with a middle income may derive more utility from certification than those with a low income.	-	+	+
C_INC2	$\beta_{C\text{-}I2}$	dU/d(C_INC2) > 0	People with a high income may derive more utility from certification than others.	-	+	+
C_SEX	$\beta_{C\text{-}S}$	dU/d(C_SEX) > 0	Women derive more utility from certification than men.	+	+	-
C_AGE1	$\beta_{C\text{-}A1}$	dU/d(C_AGE1) >< 0	The utility of certification may vary with age.	-	-	-
C_AGE2	$\beta_{C\text{-}A2}$	dU/d(C_AGE2) >< 0	Do.	+	+	+
C_EDU1	$\beta_{C\text{-}E1}$	dU/d(C_EDU1) >< 0	The utility of certification may vary with the level of education.	-	-	-
C_EDU2	$\beta_{C\text{-}E2}$	dU/d(C_EDU2) >< 0	Do.	-	-	+

Attribute / interaction effect	Para-meter	Hypotheses - mathematically (Utility denoted U)	Hypotheses - in words	Significant in sub-sample		
				A	B	C
C_KIDS	β_{C-K}	dU/d(C_KIDS) > 0	People with children derive more utility from certification than people without children.	+	-	-
C_REG	β_{C-R}	dU/d(C_REG) > 0	People living in CPH derive more utility from certification than others.	+	+	+
C_REG1	β_{C-R1}	dU/d(C_REG1) >< 0	The utility of certification may vary with the level of urbanisation of residential area.	-	-	+
C_REG2	β_{C-R2}	dU/d(C_REG2) >< 0	Do.	+	-	-
General hypotheses - attitudinal factors						
Trust in Eco-labels						
C_TRUST1	β_{C-T1}	dU/d(C_TRUST1) > 0	The utility of certification is greater for people who trust eco-labels than for people who do not. (DK)	-	+	+
C_TRUST2	β_{C-T2}	dU/d(C_TRUST2) > 0	Do. (EU)	-	+	+
C_TRUST3	β_{C-T3}	dU/d(C_TRUST3) > 0	Do. (LDC)	-	-	+
Perception of Forest Management						
C_FOR11	β_{C-F11}	dU/d(C_FOR11) < 0	The utility of certification decreases if one perceives that forests are already being managed sustainably. (DK)	-	-	+
C_FOR12	β_{C-F12}	dU/d(C_FOR12) < 0	The utility of certification decreases if one has no opinion about the present level of forest sustainability.	-	-	-
C_FOR21	β_{C-F21}	dU/d(C_FOR21) < 0	As CERT*FOR1 (EU)	+	-	-
C_FOR22	β_{C-F22}	dU/d(C_FOR22) < 0	As CERT*FOR12 (EU)	+	-	-
C_FOR31	β_{C-F31}	dU/d(C_FOR31) < 0	As CERT*FOR1 (LDC)	-	+	-
C_FOR32	β_{C-F32}	dU/d(C_FOR32) < 0	As CERT*FOR12 (LDC)	-	+	-
Interest in the environment						
C_ENV1	β_{C-E1}	dU/d(C_ENV1) > 0	The utility of certification increases with increased interest in environmental issues.	+	+	-
C_ENV2	β_{C-E2}	dU/d(C_ENV2) >0	Do.	-	+	-

Table 9.2 Continued

Attribute / interaction effect	Para-meter	Hypotheses - mathematically (Utility denoted U)	Hypotheses - in words	Significant in sub-sample		
				A	B	C
Frequency of forest visits						
C_VFOR1	$\beta_{C\text{-}V1}$	dU/d(C_VFOR1) >< 0	The utility of certification may vary with the frequency of forest visits.	+	+	+
C_VFOR2	$\beta_{C\text{-}V2}$	dU/d(C_VFOR2) >< 0	Do.	-	-	-
Product specific hypotheses						
All 3 product types						
CERT	β_C	dU/dCERT > 0	If a product is certified, the overall utility of it increases.	+	+	+
PRICE	β_P	dU/dPRICE < 0	If the price of a product increases, the overall utility of it decrease.	+	+	+
Toiletpaper and cuttingboards						
C_INFO	$\beta_{C\text{-}I}$	dU/d(C_INFO) > 0	Increased information about certification increases the utility associated with certification.	-	+	
Toilet paper						
RECYC	β_R	dU/dRECYC >< 0	The utility of paper may depend on whether it is made from recycled or new fibres.	+		
LAY	β_L	dU/dLAY > 0	The utility of 3-layer paper is higher than that of 2-layer paper.	+		
SOFT	β_S	dU/dSOFT > 0	As the softness of paper increases, the utility associated with it increase.	+		
VSOFT	β_{VS}	dU/dVSOFT > 0	Do.	+		
Cuttingboards						
HEVEA	β_H	dU/dHEVEA >< 0	The utility of a product depend on its appearance.		+	
BEECH	β_B	dU/dBEECH >< 0	Do.		+	
TEAK	β_T	dU/dTEAK >< 0	Do.		-	
DESIGN	β_D	dU/dDESIGN >< 0	Do.		+	
Table tops						
DESIGN1	β_{D1}	dU/dDESIGN1 >< 0	The utility of a product depends on its appearance.			+
DESIGN2	β_{D2}	dU/dDESIGN2 >< 0	Do.			+
DESIGN3	β_{D3}	dU/dDESIGN3 >< 0	Do.			+

Attribute / interaction effect	Para-meter	Hypotheses - mathematically (Utility denoted U)	Hypotheses - in words	Significant in sub-sample		
				A	B	C
VAR	β_V	dU/dVAR < 0	If the quality of a product increases, the overall utility of it increases.			+
C_ACCO1	$\beta_{C\text{-}A1}$	DU/d(C_ACCO1) ><0	The utility of certification may depend on the type of accommodation.			+
C_ACCO2	$\beta_{C\text{-}A2}$	DU/d(C_ACCO2) ><0	Do.			-

In relation to model building, Hosmer & Lemeshow (2000) recommend that univariable analysis is used to identify potentially relevant main-effect variables. Subsequently, potentially important interaction effects are identified, by including the interactions one at a time. In this study, a modification of this strategy is used: univariable/bivariable analysis. Estimation of separate models for each potentially important interaction term[50] is used to determine whether or not the proposed hypotheses are significant. To avoid premature rejection of hypotheses – i.e. failing to identify potentially important variables – a p-value of 0,25 is used as a screening criterion. Variables with $p < 0,25$ are considered potentially significant whereas variables with $p > 0,25$ are deemed insignificant as determinants of choice (Hosmer & Lemeshow, 2000).

In table 9.2, the last column specifies whether or not the interaction was significant in the univariable regression, based on the data pertaining to toilet paper (A), cutting boards (B) and table tops (C).

9.1.4 Step 4: Deriving the Final Choice Model

In step 4 of the analysis, the final choice model including both product attributes and interactions between product attributes and respondent characteristics is derived.

Based on the results of the univariable analysis, a preliminary multivariable model is specified. The final model is subsequently derived by subjecting this preliminary model to the model selection procedures available in SAS[51] – i.e. backwards, forward, stepwise and

[50] More specifically, these separate models are estimated with the interaction term along with the relevant products attribute as the only explanatory variables.

[51] Hosmer & Lemeshow (2000) note that using the univariable approach to determine, which variables to include in the mulitivariable model may imply that collections of variables which are only significant when taken together are overlooked. Running the best-subset selection procedure on a data set including all variables – both those that were significant, and those that were insignificant in the univariable analysis – the results did not indicate that any collection of jointly important variables had been missed.

best-subset selection – stipulating that all variables included in the final model must be significant at the 5% level[52].

In terms of the significance of the variables included in the final model, Lovell (1983) proposes that exploratory data mining procedures tend to result in exaggerated claims of significance. Subsequently, a Rule of Thumb (RoT) for determining true significance levels is proposed (Lovell, 1983). With c variables in the candidate set, k variables in the final model, and a desired significance level of α, the relevant significance level to impose on coefficients is approximated by: $\hat{\alpha} = \frac{k}{c}\alpha$. The extent, to which the applied model selection procedures have resulted in the inclusion of "insignificant" variables in the final model, is investigated by applying the RoT to determine the adjusted significant levels of the variables included in the model.

9.1.5 Step 5: Comparison of Models

In order to determine if the inclusion of respondent characteristics improves the model, the model derived in step 4 is compared to a model based on attributes only, where potential insignificant variables have been excluded.

Depending on whether or not the two models to be compared are nested or not, different tests are used to investigate if the model in statistical terms have been significantly improved by the inclusion of respondent characteristics. For the comparison of nested models a LR-test similar to the one in step 1 and step 2 is used. For the comparison of non-nested models[53], a test developed by Horowitz (1983) is used.

The test, which is developed specifically for testing non-nested hypotheses of discrete choice models, is based on comparison of the $\bar{\rho}^2$'s for the two models, and is given by:

$$\Pr\left(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z\right) \leq \Phi\left\{-\left[-2zL(0) + \left(K_2 - K_1\right)\right]^{1/2}\right\}$$

where $\bar{\rho}_l^2$ = the adjusted likelihood ratio index from model l = 1, 2; K_l = the number of parameters in model l, and Φ = the standard normal cumulative distribution function (Ben-Akiva & Lerman, 1985). The test states the probability that the observed difference in adjusted LRI's is obtained by chance. The smaller this probability is, the more likely it is

[52] Hosmer & Lemeshow (2000) warn against the use of too stringent entry/removal criteria. This may result in a premature dismissal of important variables and/or failure to identify groups variables, that only are significant when considered jointly. To avoid this, the stringency criteria imposed on the selection of variables have been tightened incrementally, and it has been checked that the stringency of the initially imposed criteria has not influenced on the composition of the final model.

[53] Non-nested models are defined as models, where neither can be obtained as a special case of the other – i.e. given the parameters of one of the models, it is not possible to choose values for the other model's parameters in such a way that the models become identical (Horowitz, 1983).

that the difference can be attributed to a statistically significant difference in the fit of the models.

9.1.6 Step 6: Test for Violation of IIA

As mentioned in chapter 8, an inherent property of logit models is the assumption of Independence from Irrelevant Alternatives (IIA). IIA implies that the choice probabilities observed from a subset of alternatives only depend on the alternatives included in the subset. Thus, for the IIA to be satisfied it is required that the estimated relative choice probabilities are unaffected by the removal, or addition, of alternatives from/to the choice set (Ben-Akiva & Lerman, 1985).

A test developed by Hausman & McFadden (1984) can be used to test for violations of IIA. The test builds on the observation, that if the logit model is correctly specified, parameter estimates from a model estimated on full choice sets should be consistent with parameter estimates from a model estimated on reduced choice sets. The test statistic, which tests the hypothesis that $\hat{\beta}_C = \hat{\beta}_{\tilde{C}}$ is given by (Ben-Akiva & Lerman, 1985):

$$\left(\hat{\beta}_{\tilde{C}} - \hat{\beta}_C\right)' \left(\Sigma_{\hat{\beta}_{\tilde{C}}} - \Sigma_{\hat{\beta}_C}\right)^{-1} \left(\hat{\beta}_{\tilde{C}} - \hat{\beta}_C\right)$$

where $\hat{\beta}_C$ and $\hat{\beta}_{\tilde{C}}$ refer to the estimated coefficient vectors for the models estimated on the full and the reduced data sets, respectively, and where $\Sigma_{\hat{\beta}_C}$ and $\Sigma_{\hat{\beta}_{\tilde{C}}}$ refer to the associated covariance matrices. The test statistic is asymptotically χ^2 distributed with $\tilde{K}$ degrees of freedom, where $\tilde{K}$ is equal to the dimension of $\hat{\beta}_{\tilde{C}}$. Basically the test determines the extent to which the parameter estimates from the two models are the same (Hausman & McFadden, 1984). If they are found to be approximately the same – i.e. if the hypothesis cannot be rejected – it is taken to indicate that the IIA assumption is not violated, and consequently, that the logit specification can be accepted. On the other hand, if the hypothesis is rejected, this indicates that the IIA assumption is violated, implying that another model specification should be used.

The model derived in step 4 is tested for violations of IIA using two different approaches to choice set reductions. In the first approach, choice sets are reduced by eliminating alternatives A, B, C and D from the choice sets one at a time. In the second approach, choice sets are reduced on account of one of the attributes.

9.1.7 Step 7: Estimation of the Model Excluding Uncertain Respondents

In the last step of the analysis, the final model derived in step 4 is estimated on a data set where respondents having reported to be somewhat uncertain about their ranking have been excluded. The purpose is to investigate if the composition of the model changes,

when uncertain choices are excluded from the data set. Since the models are estimated on different data sets, it is not possible to test, which model is best.

Upon completion of the rankings in the questionnaires, respondents were asked on a scale from 0 to 10, with 0 denoting "very uncertain" and 10 denoting "completely certain", to state how certain they were about the rankings that they had just made. Until now, all respondents no matter their level of reported certainty have been included in the analysis. There are two reasons why it was chosen to include all. Firstly, it was judged that as long as people have felt capable of completing the ranking, then the information inherent in their rankings would be relevant. Secondly, it is expected that people may have used the scale quite differently[54]. If respondents with a reported level of certainty below a certain number are excluded from the data set, some of the excluded observations would be associated with less uncertainty, than some of the observations included in the reduced data set.

However, it is acknowledged that it might be inappropriate to include "uncertain observation" on equal terms with "very certain observation". Thus, it is judged relevant to re-estimate the model derived in step 4 on a data set where uncertain respondents are excluded. The identification of the threshold below which the level of certainty is unacceptable – i.e. too low – is bound to be arbitrary. Even though it was emphasised that "uncertain" answers were no less valuable to the survey than "certain" ones, it is expected that at least some have felt that they should report at least some minimum level of certainty. Based on this, we judge that the appropriate threshold at least should be well above the middle. Subsequently, 7 is chosen as the threshold.

9.2 The Software

All data processing – i.e. both analyses and preparation of data for analysis – is done using SAS. The more specific procedure that is used is the procedure PROC PHREG, which can handle Rank Logit (RL) and Multinomial Logit (MNL). In relation to RL, PROC PHREG is able to handle ties, which is necessary in cases where the respondent only has assigned a rank to the most preferred alternative (Allison, 2001). For the tests for violation of the IIA assumption, the SAS procedure PROC IML, which can handle matrix operations, is used.

The data sets along with the programming editors used in the analysis and modelling of data are documented on the enclosed CD-Rom.

[54] This observation is analogous to the problems associated with using ratings, which were mentioned in chapter 4.5.3

9.3 Analysis of Toilet Paper Data

In the present section the 7 steps of analysis presented in section 9.1 will be performed on the data obtained from the ranking of toilet paper, i.e. Rank A.

9.3.1 Step 1: Toilet paper

The results of the LR-tests for equality of beta-coefficient vectors across the two information scenarios are presented in table 9.3. The test results show that the hypothesis of equality of beta-coefficient vectors cannot be rejected[55]; neither for the rank order logit model nor for the multinomial logit model. Accordingly, there should be no problems associated with pooling the datasets for the subsequent analysis no matter which model is deemed most appropriate for further analysis in step 2.

Table 9.3 Results of the test for equality of coefficients across the datasets for toilet paper from the two information scenarios.

Model	Test-statistic	DF	$\chi^2_{0,95}$
Rank Order Logit	3,192	6	12,59
1st Choice MNL	0,738	6	12,59

[55] One explanation for this may be that, as is seen in table 9.2, the coefficient for the certification-information interaction is insignificant in all but one of the estimated models.

9.3.2 Step 2: Toilet Paper

The five different models for toilet paper, with attributes as the only explanatory variables, are presented in table 9.4. When comparing the parameter estimates for the different models, they appear to vary significantly between the models. Some parameters even change sign, which indicate that the models might not describe the same choice process. With reference to the first three models, the changes in coefficient estimates for the significant attributes conform to the expectation that the numerical size of coefficient estimates decrease with increasing depth of explosion of

Table 9.4 Models of toilet paper choice.

	RL - all ranks	**RL - 1st + 2nd Choice**	**1st Choice MNL**	**2nd Choice MNL**	**3rd Choice MNL**
Parameter	β	β	β	β	β
SOFT	0,2117* (0,0904)	0,1116 (0,1073)	0,0353 (0,1557)	0,1044 (0,1544)	0,3019 (0,1845)
VSOFT	0,0178 (0,0914)	0,0260 (0,1077)	0,1074 (0,1517)	-0,0840 (0,1536)	0,0014 (0,1862)
RECYC	0,2765** (0,0731)	0,4332*** (0,0851)	0,5750*** (0,1207)	0,2548* (0,1257)	-0,2629 (0,1635)
CERT	0,6356*** (0,1049)	0,8083*** (0,1234)	1,1327*** (0,1902)	0,5583*** (0,1693)	0,2451 (0,2152)
LAY	0,2540** (0,0734)	0,2121* (0,0848)	0,3758* (0,1254)	0,0458 (0,1178)	0,3890** (0,1605)
PRICE	-0,0705*** (0,0061)	-0,0789*** (0,0071)	-0,0998*** (0,0104)	-0,0555*** (0,0101)	-0,0447*** (0,0133)
C_INFO	0,1080 (0,1463)	0,0184 (0,1697)	-0,0970 (0,2534)	0,1313 (0,2345)	0,6652* (0,3091)
Log Likelihood	-996,61	-758,86	-400,57	-348,53	-217,17
ρ^2	0,1048	0,1339	0,2002	0,0723	0,0839
$\bar{\rho}^2$	0,0985	0,1259	0,1863	0,0537	0,0544
N	364	364	364	342	342
WTP for CERT	**9,02 DKK/pack**	**10,24 DKK/pack**	**11,35 DKK/pack**	**10,06 DKK/pack**	**5,48 DKK/pack**

* significant at the 5% level; ** significant at the 1% level; *** significant at the 0,1% level. Numbers in brackets are the standard errors. The implicit Willingness-To-Pay for certification presented in the last row is obtained by dividing the coefficient estimate for certification with the coefficient estimate for price. The effect of information has not been included since the coefficient estimate for the certification-information interaction is insignificant in all except one of the estimated models.

the rank. However, the changes in coefficient estimates do not appear to be proportional for the different attributes. In relation to PRICE and CERT, for instance, the change is greater for CERT than for PRICE implying that the implicit Willingness-to-Pay (WTP) for certification decreases, as additional ranks are included in the estimation. Comparing the

last three models, it is seen that the number of significant variables tends to decline from 1st to 3rd Choice. PRICE, CERT, RECYC and LAY, at least to some extent, all appear to be important determinants of choice, whereas SOFT, VSOFT and C_INFO appear to be insignificant.

With reference to the LRI values for the three first models presented in table 9.4 it is seen that the LRI by far is largest for the 1st Choice MNL model – i.e. the model where the ranking only is exploded to the depth of one. This indicates that the analysis should be based solely on the data pertaining to the first rank.

The results of the LR-test used to verify the conclusion drawn from the LRI's are presented in table 9.5. It appears that the hypothesis of equality of coefficient vectors across all ranks, and across rank 1 and rank 2 can be rejected. However, the hypothesis of equality of coefficients across rank 2 and rank 3 cannot be rejected.

Table 9.5 Likelihood ratio tests for pooling different levels of rank data for toilet paper.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice	60,68	14	23,68
1st Choice vs 2nd Choice	19,52	7	14,07
2nd Choice vs 3rd Choice	11,13	7	14,07

The results of the LR-tests for the rescaled models are presented in table 9.6. The hypothesis $\beta^{m+1}=\mu_m\beta^m$ cannot be rejected in either of the cases considered. This points out that the difference in coefficient vectors across ranks is ascribable to respondents finding it increasingly difficult to assign ranks the lower they get (Ben-Akiva *et al.*, 1991). Thus, had it been a consequence of choice inconsistency it is unlikely that the difference in coefficient estimates from different choice data sets could be represented by a single scale parameter.

Table 9.6 Likelihood ratio tests for pooling different levels of rank data for toilet paper after correcting for different scales.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice	22,52	14	23,68
1st Choice vs 2nd Choice	4,04	7	14,07

Note: μ_m was found to be 0,55 for the comparison of data set 1 and 2; and to be 0,5 for the comparison of data set 1 and 3, see appendix VII. These values conform to the expectation that the variance associated with lower ranks is higher than that of higher ranks.

Even though the rescaled LR-tests suggest that data from all ranks can be pooled implying that the RL – all ranks model can be used, we chose to base the further analysis on the 1st choice MNL model specification. The unscaled LR-tests show that the reliability of the data decreases with increased depth of explosion of the rank. Intuitively this seems reasonable. It is probably much easier to choose the most preferred alternative out of four than to choose the most preferred alternative out of two unattractive alternatives. The latter may be considered equivalent to conducting an analysis on information obtained from people that have been placed between the devil and the deep blue sea. This is a situation where at least some, if not most, people probably will find that the easiest way to make a choice is to flip a coin. The fact that only 2 attributes are significant in the third model indicates that something similar might be the case. Consequently, it is considered appropriate to exclude the data pertaining to the 3rd Choice, also it is only by a narrow margin, that the hypothesis of proportionality is accepted. A comparison of the LRI for the 1st Choice MNL in table 9.4 and the LRI of the rescaled 1st+2nd Choice RL model (see appendix IV) show that the LRI, despite the rescaling, drop significantly when the depth of explosion is increased from 1 to 2. This indicates that the 1st choice model is the most appropriate. With reference to the LRI's for the relevant models and considerations similar to the ones just discussed, it is therefore chosen to exclude the data pertaining to the 2nd choice[56].

9.3.3 Step 3: Toilet Paper

The results of the univariable regressions presented in table 9.2 show that 10 out of the 27 hypotheses regarding interaction terms are potentially significant for toilet paper (sub-sample A). These variables, along with the product attributes, will enter the multivariable analysis considered in the next step.

9.3.4 Step 4: Toilet Paper

Based on the results of the univariable analysis, a preliminary multivariable model for toilet paper is specified as follows:

$$
\begin{aligned}
U_{TP} = {} & \beta_{P_I1} \cdot P_INC1 + \beta_{C_S} \cdot C_SEX + \beta_{C_A2} \cdot C_AGE2 + \beta_{C_K} \cdot C_KIDS \\
& + \beta_{C_R} \cdot C_REG + \beta_{C_R2} \cdot C_REG2 + \beta_{C_F21} \cdot C_FOR21 + \beta_{C_F22} \cdot C_FOR22 \\
& + \beta_{C_E1} \cdot C_ENV1 + \beta_{C_V1} \cdot C_VFOR1 + \beta_{C} \cdot CERT + \beta_{P} \cdot PRICE + \beta_{R} \cdot RECYC \\
& + \beta_{L} \cdot LAY + \beta_{S} \cdot SOFT + \beta_{VS} \cdot VSOFT
\end{aligned}
$$

[56] By choosing to use the MNL model specification, problems arising from potential violations of the IID assumption are avoided (see section 7.5.2).

Subjecting this preliminary model to the model selection procedures in SAS, and stipulating that all variables included in the final model must be significant at the 5% level, the following model results[57]:

$$U_{TP} = 1{,}2189 \cdot CERT^{(***)} - 0{,}09827 \cdot PRICE^{(***)} + 0{,}5579 \cdot RECYC^{(***)} + 0{,}4092 \cdot LAY^{(***)} - 0{,}5951 \cdot C_VFOR1^{(*)}$$

with: $L(\hat{\beta}) = -398{,}895$, $\rho^2 = 0{,}2036$ and $\bar{\rho}^2 = 0{,}1936$.

It appears that the number of variables has been significantly reduced; among others the softness attributes have been removed from the model. The signs on the coefficient estimates for CERT, PRICE and LAY are as expected. Apart from this, people apparently prefer toilet paper made from recycled paper as opposed to paper made from new fibres – this may be seen as an expression of preferences for environmentally friendly paper. Furthermore, it appears that people who visit forests at least once a week have less strong preferences for certified products than other people. This could be due to the fact, that people who visit forests often, are more familiar with current forest management practices, and do not consider them to be unsustainable, thus decreasing the need for certification at least in Denmark. In terms of the implicit price of certification, people who visit forests less than once a week (i.e. the base case) are estimated to have a marginal WTP for certification equal to: (1,2189/0,09827) = 12,4 DKK/pack, whereas people, who pay visits more often are estimated to have a marginal WTP for certification of: ((1,2189-0,5951)/0,09827) = 6,34 DKK/ pack.

The preliminary model contained 16 variables, and the final model contains 5 variables. Imposing a significance level of 5% on all variables included in the model, and applying Lovell's RoT, the relevant significance level becomes: $\hat{\alpha} = (5/16) \cdot 0{,}05 = 0{,}0156 \approx 0{,}016$. Applying this significance level, the variable VFOR1 is no longer significant at the 5% level. Using the Rule of Thumb, the actual/deflated significance level for the variable VFOR1 is estimated to be: $\alpha = (16/5) \cdot 0{,}045 = 0{,}144 \approx 0{,}14$.

Despite this, VFOR1 is kept in the model. Thus, even though it may not be the most obviously relevant variable, it does not clash with common sense. Moreover, it satisfies the conventional way of defining significance levels.

[57] The asterisks in brackets denote the level of significance of the coefficient estimates; *** = significant at the 0,1% level, ** = significant at the 1% level and * = significant at the 5% level.

9.3.5 Step 5: Toilet Paper

In order to investigate if the inclusion of interactions between attributes and respondent characteristics has improved the model, the final model derived in step 4 is compared with the main effects model, where insignificant variables have been deleted, i.e.:

$$U_{TP} = 1{,}0777 \cdot CERT^{(***)} - 0{,}0988 \cdot PRICE^{(***)} + 0{,}5545 \cdot RECYC^{(***)} + 0{,}3904 \cdot LAY^{(**)}$$

with: $L(\hat{\beta}) = -400{,}8945$, $\rho^2 = 0{,}1996$ and $\bar{\rho}^2 = 0{,}1915$.

Comparing the $L(\hat{\beta})$'s and $\bar{\rho}^2$'s of the two models there does not appear to be any significant difference between the fit of the two models.

The LR-test statistics used to verify, if the addition of the C_VFOR1 variable significantly improves the model in statistical terms is given by:

$$-2\,(L_R - L_U) = -2\,(-400{,}8945 - (-398{,}895)) = 3{,}999$$

and is χ^2 distributed with $K_U - K_R = 5 - 4 = 1$ degrees of freedom. Compared with $\chi^2_{1,0.95} = 3{,}84$, it appears that the hypothesis can be rejected. This implies that including C_VFOR1 in the model constitutes a statistically improvement of the model. Thus, the test contradicts the intuitive conclusion drawn from just looking at the goodness of fit measures, though not by much.

9.3.6 Step 6: Toilet Paper

In addition to eliminating alternatives A, B, C and D, the toilet paper choice sets have been reduced by eliminating on account of the softness attribute.

The results of the tests for violations of IIA are presented in table 9.7, where it appears that the hypothesis cannot be rejected for either of the cases considered. Accordingly, there seems to be no need to question the use of the logit model specification to model toilet paper choice in this case.

Table 9.7 Hausman & McFadden's (1984) tests for violation of IIA for toilet paper.

Alternative removed	A	B	C	D	Standard soft	Soft	Very soft
Test Statistic	7,07	6,21	6,13	6,53	6,77	3,04	3,96
$\chi^2_{0,95}$	11,07	11,07	11,07	11,07	11,07	11,07	11,07

9.3.7 Step 7: Toilet Paper

Estimating the model derived in step 4 on the data set where uncertain respondents have been excluded – and excluding insignificant variables – the following model results:

$$U_{TP} = 1{,}0577 \cdot CERT^{(***)} - 0{,}1006 \cdot PRICE^{(***)} + 0{,}6125 \cdot RECYC^{(***)} + 0{,}4537 \cdot LAY^{(***)}$$

with: $L(\hat{\beta}) = -313{,}44$; $\rho^2 = 0{,}2102$ and $\bar{\rho}^2 = 0{,}2002$.

Thus, by excluding uncertain respondents, C_VFOR1 is no longer significant, and the coefficient estimates change a little. The implicit WTP for CERT is now estimated to be 10,51 DKK/pack for all respondents no matter their frequency of forest visits.

9.4 Analysis of Cutting Boards Data Set

In this section the 7 steps of the analysis will be performed on the data set obtained from the ranking of cutting boards, i.e. Rank B.

9.4.1 Step 1: Cutting Boards

The results of the LR-test for equality of beta coefficient vectors across the two information scenarios are presented in table 9.8 and they show that the hypothesis $\hat{\beta}_{data\ set\ 1} = \hat{\beta}_{data\ set\ 2}$ cannot be rejected. This implies that the data sets from the two scenarios can be pooled in the further analysis.

Comparing the sizes of the test-statistics for pooling the toilet paper data sets and the cutting board data sets it is seen that the test-statistics are significantly higher for the cutting board dataset. This suggests, that the effect of information is greater in relation to choosing between different cutting boards than choosing between different packs of toilet paper.

Table 9.8 Results of tests for equality of coefficients across the datasets for cutting boards from the two information scenarios.

Model	Test-statistic	DF	$\chi^2_{0,95}$
Rank Order Logit	8,415	6	12,59
1st Choice MNL	5,043	6	12,59

9.4.2 Step 2: Cutting Boards

The five different models estimated with product attributes (plus the certification-information interaction) as the only explanatory variables are presented in table 9.9. Looking at the models, some of the same trends that were seen for the toilet paper models are also present here; the coefficient estimates seem to vary significantly between the

models, and the number of significant variables decline from 1st to 3rd Choice. The last row in the table shows that the implicit WTP for certification also varies across models.

Table 9.9 Models of cutting board choice.

	RL - all ranks	**RL - 1st + 2nd Choice**	**1st Choice MNL**	**2nd Choice MNL**	**3rd Choice MNL**
Parameter	β	β	β	β	β
HEVEA	-0,4247*** (0,1006)	-0,4412*** (0,1255)	-0,6087** (0,1873)	-0,2796 (0,1748)	-0,4876** (0,1770)
BEECH	0,3820*** (0,1018)	0,4115*** (0,1149)	0,4114** (0,1493)	0,4028* (0,4028)	0,1504 (0,15036)
TEAK	0,1357 (0,0982)	0,0876 (0,1152)	-0,0576 (0,1609)	0,3119 (0,1702)	0,0173 (0,2048)
DESIGN	0,1827* (0,1827)	0,2215* (0,0912)	0,0350 (0,1351)	0,3855** (0,3855)	-0,0744 (0,1487)
CERT	0,4322*** (0,4322)	0,4502*** (0,1194)	0,4392** (0,4392)	0,4708** (0,4708)	0,2448 (0,2207)
PRICE	-0,00207*** (0,0003)	-0,0026*** (0,0004)	-0,0032*** (0,0006)	-0,0018** (0,0006)	-0,0010 (0,0007)
C_INFO	0,2008 (0,1447)	0,4157* (0,168)	0,5491* (0,234)	0,2714 (0,2714)	-0,4085 (0,3014)
Log Likelihood	-1001,34	-761,66	-420,14	-334,34	-217,89
ρ^2	0,074	0,107	0,142	0,078	0,045
$\bar{\rho}^2$	0,067	0,099	0,128	0,058	0,014
N	354	354	354	331	330
WTP for CERT	**208,79 DKK**	**173,15 DKK / 333,04 DKK**	**137,25 DKK / 308,84 DKK**	**263,02 DKK**	**244,8 DKK**

* significant at the 5% level; ** significant at the 1% level; *** significant at the 0,1% level. Numbers in brackets are the standard errors. The implicit WTP for certification presented in the last row, where the first estimate relates to the WTP in the reduced information scenario and the second relates to the "full information" scenario (only calculated for the models where the interaction is significant).

In terms of the significance of different variables it is seen that PRICE, CERT, HEVEA and BEECH all appear to be important determinants of choice. However, the importance of DESIGN and C_INFO varies, but generally appear to be quite low and TEAK appears overall to be insignificant[58].

Comparing the LRI for the 3 first models, it is indicated that explosion should not be taken further than to the depth of 1. A likelihood ratio test is used to verify this conclusion. The results of the test are shown in table 9.10, where it is seen that the hypothesis of equal coefficient vectors across data sets can be rejected in all the cases considered. For the 1st

[58] That TEAK is insignificant may point out that people are indifferent between TEAK and rosewood, which is the contrast category. Since both are dark types of wood this makes sense.

Choice and 2nd Choice data sets, however, it is only by a narrow margin that the hypothesis is rejected.

Table 9.10 Likelihood ratio tests for pooling different levels of rank data for cutting boards.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice	57,93	14	23,68
1st Choice vs 2nd Choice	14,35	7	14,07

The results of the subsequent LR-test for proportionality of the beta-coefficient vectors are presented in table 9.11, where the hypothesis of proportional parameter vectors cannot be rejected. Thus, as was the case with toilet paper inequality of parameter vectors is caused by different variances, rather than choice inconsistencies as such.

Table 9.11 Likelihood ratio tests for pooling different levels of rank data for cutting boards after correcting for different scales.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice	21,11	14	23,68
1st Choice vs 2nd Choice	9,81	7	14,07

Note: μ_m was found to be 0,75 for the comparison of data set 1 and 2; and to be 0,425 for the comparison of data set 1 and 3 (see appendix VII). These values conform to the expectation that the variance associated with lower ranks is higher than that of higher ranks.

Despite the fact that the results of the rescaled LR-test suggest that data can be pooled across ranks, it is nevertheless decided to base the further analysis on the data related to 1st choice only[59]. This decision is based on considerations similar to the ones advanced in section 9.3.2.

9.4.3 Step 3: Cutting Boards

The results of the univariable regressions for cutting boards (sub-sample B) are shown in table 9.2, where it is seen that 13 out of 27 hypotheses regarding interaction terms are potentially significant for explaining cutting board choice.

[59] The 1st +2nd Choice RL model is presented in appendix V. Here it appears that the LRI, despite the rescaling, drop significantly when the depth of explosion is taken further than the depth of 1.

9.4.4 Step 4: Cutting Boards

Based on the results of the univariable analysis, the preliminary multivariable model for cutting board choice is given by:

$$
\begin{aligned}
U_{CB} = {} & \beta_{P_S} \cdot P_SEX + \beta_{C_I1} \cdot C_INC1 + \beta_{C_I2} \cdot C_INC2 + \beta_{C_S} \cdot C_SEX \\
& + \beta_{C_A2} \cdot C_AGE2 + \beta_{C_R} \cdot C_REG + \beta_{C_T1} \cdot C_TRUST1 + \beta_{C_T2} \cdot C_TRUST2 \\
& + \beta_{C_F31} \cdot C_FOR31 + \beta_{C_F32} \cdot C_FOR32 + \beta_{C_E1} \cdot C_ENV1 + \beta_{C_E2} \cdot C_ENV2 \\
& + \beta_{C_V1} \cdot C_VFOR1 + \beta_{C_I} \cdot C_INFO + \beta_{C} \cdot CERT + \beta_{P} \cdot PRICE \\
& + \beta_{H} \cdot HEVEA + \beta_{B} \cdot BEECH + \beta_{D} \cdot DESIGN
\end{aligned}
$$

Subjecting the preliminary multivariable model to the backwards, forward, stepwise and best-subset model selection procedures, the following models results:

$$
\begin{aligned}
U_{CB} = {} & 0{,}6764 \cdot C_INC2^{(**)} - 0{,}6808 \cdot C_SEX^{(**)} + 0{,}4399 \cdot C_ENV1^{(*)} \\
& + 0{,}8462 \cdot C_ENV2^{(**)} + 0{,}4889 \cdot C_INFO^{(*)} - 0{,}0032 \cdot PRICE^{(***)} \\
& - 0{,}5818 \cdot HEVEA^{(***)} + 0{,}4474 \cdot BEECH^{(***)}
\end{aligned}
$$

with: $L(\hat{\beta})$=-411,1975, $\rho^2 = 0{,}1604$ and $\bar{\rho}^2 = 0{,}1441$

In this model the utility of cutting boards is determined by PRICE, type of wood (i.e. HEVEA, BEECH vs. darker woods) and certification. The coefficient estimates, for which specific sign were expected, are found to conform to the expectations. For the different types of wood it is interesting to note that the respondents apparently have positive preferences for Beech but negative preferences for Hevea.

In relation to certification it is interesting to note, that CERT as a main effect has disappeared from the model – instead the utility associated with certification is expressed through its interactions with respondent characteristics. Thus, the utility derived from certification depends on the income level of the household (C_INC2), the sex of the respondent (C_SEX), the respondent's interest in environmental issues (C_ENV1, C_ENV2) and the availability of information about certification (C_INFO). An implication of this specification is that people fitting the base case derive no utility from certification. This conclusion may seem a little harsh, however, considering that the model describes aggregate preferences rather than a specific individual's preferences, it does not appear that unreasonable.

In terms of the implicit WTP for certification, the model shows that[60]:

- 211 DKK/cutting board for respondents living in households with a yearly income >400.000 DKK.
- 153 DKK/cutting board for respondents who receive the high level of information.
- 137 DKK/cutting board for respondents with average interest in environmental issues.
- 264 DKK/cutting board for respondents with great interest in environmental issues.
- 213 DKK/cutting board for men.

The fact that the model describes aggregate preferences may also serve to explain why an attribute like DESIGN was found to be insignificant. Whether one prefers one design over the other is a matter of taste. Analogous to an example given by McCullough (2002), this implies that if half of the respondents prefer design#1, and the second half prefer design#2, then the aggregate model will show that respondents are indifferent to design[61] – i.e. that the attribute is insignificant.

Applying Lovell's RoT, the relevant significance level to impose on coefficients is approximated by: $\hat{\alpha} = (8/19)\cdot 0{,}05 = 0{,}021$. Applying this significance level, C_ENV1 and C_INFO are no longer significant at the 5% level; their respective significance levels are estimated to be 11,9% and 7,8%.

Once again, despite the conclusions to be drawn from the RoT, we choose to keep the variables in the model; thus, the model is intuitively appealing, and it satisfies the conventional way of defining significance levels.

9.4.5 Step 5: Cutting Boards

To investigate if the inclusion of respondent characteristics has improved the fit of the model, the model derived in step 4 is compared with the main effects model containing all significant attributes, i.e.:

$$U_{CB} = 0{,}5497 \cdot C_INFO^{(*)} - 0{,}0033 \cdot PRICE^{(***)} - 0{,}5883 \cdot HEVEA^{(***)} + 0{,}4391 \cdot BEECH^{(***)} + 0{,}4502 \cdot CERT$$

with: $L(\hat{\beta})$=-420,2295, $\rho^2 = 0{,}142$ and $\bar{\rho}^2 = 0{,}1318$.

Since the two models are non-nested they are compared by the test developed by Horowitz (1983). With $L(0)$ = -489,77; model 1 = the model with lowest $\bar{\rho}^2$; model 2 = the model

[60] It should be noted that the WTP's are additive with regards to the characteristics of the respondents.
[61] McCullough (2002) discourages the use of aggregate models due to such failures to capture heterogeneity.

with highest $\bar{\rho}^2$ and z set to 0,0123 (i.e. the difference in $\bar{\rho}^2$'s between the two models) we get that:

$$\Pr\left(\bar{\rho}_2^2 _ \bar{\rho}_1^2 > 0{,}0123\right) \leq \Phi\left(-3{,}88\right) = 0{,}00005$$

Thus, the test shows that the probability that the observed difference in $\bar{\rho}^2$'s would be exceeded is less than 1 out of 20.000. Consequently, the observed difference is unlikely to be accidental, and therefore it can be concluded that the model with the highest $\bar{\rho}^2$ - i.e. the model including respondent characteristics – has a better fit than the model with a lower $\bar{\rho}^2$ [62].

9.4.6 Step 6: Cutting Boards

In addition to eliminating alternatives A, B, C and D, the cutting board choice sets have been reduced by eliminating on account of the different wood types. The results of the tests for violations of IIA for the model of cutting board choice are presented in table 9.12. The results show, that in two of the cases considered, namely when the beech and the teak alternatives are excluded, the hypothesis of no violation of IIA is rejected. This is of course somewhat problematic since it indicates that the estimated model – and thereby also the estimated WTP for CERT – is contingent upon the chosen wood types. Ideally, another mode specification should be applied, but, due to practical limitations, this is not possible.

Table 9.12 Hausman & McFadden's (1984) tests for violation of IIA in the model for cutting boards.

Alternative removed	**A**	**B**	**C**	**D**	**Hevea**	**Beech**	**Teak**	**Rose-wood**
Test Statistic	9,84	8,87	4,56	13,6	13,46	20,96	45,62	0,27
$\chi^2_{0,95}$	15,51	15,51	15,51	15,51	14,07	14,07	15,51	15,51

9.4.7 Step 7: Cutting Boards

Estimating the model derived in step 4 on a data set where respondents with a self-reported level of confidence in their ranking below 7 are excluded, and insignificant variables subsequently are removed, the following model results:

$$U_{CB} = 0{,}8467 \cdot C_INC2^{(**)} - 0{,}5846 \cdot C_SEX^{(*)} + 0{,}8451 \cdot C_ENV2^{(*)} + 0{,}805 \cdot C_INFO^{(**)} - 0{,}0027 \cdot PRICE^{(***)} - 0{,}4706 \cdot HEVEA^{(*)} + 0{,}4975 \cdot BEECH^{(**)}$$

with: $L\left(\hat{\beta}\right)$=-284,57; $\rho^2 = 0{,}1621$ and $\bar{\rho}^2 = 0{,}1415$.

[62] Stated another way, the test states that the probability of choosing the incorrect model – when choosing the model with the highest $\bar{\rho}^2$ - is les than 0,00005.

As it appears, the interaction variable C_ENV1 is no longer significant, and the coefficient estimates, and the associated significance levels, for the remaining variables change. In terms of the implicit WTP for certification it is now estimated to be approximately:

- 316 DKK/cutting board for respondents living households with a yearly income >400.000 DKK.
- 300 DKK/cutting board for respondents who receive the high level of information.
- 315 DKK/cutting board for respondents with great interest in environmental issues.
- 218 DKK/cutting board for men.

Thus, following the exclusion of uncertain respondents the WTP associated with different respondent characteristics appear to have increased significantly.

9.5 Analysis of the Table Tops Data Set

In this section the analysis will be performed on the data set obtained from the rankings of table tops, i.e. Rank C and D. As the ranking of table tops was specific to the "low information" scenario, step 1 is not relevant for the table top data. Consequently, the derivation of the final choice model for table tops will begin with step 2; prior to this, however, a few comments will be made on the data set.

9.5.1 The Data Set

The data obtained from the rankings of table tops distinguishes itself from the data pertaining to the two other product types in at least two ways; 1) each respondent was asked to complete two separate rankings, and 2) the respondent was asked to compare the two alternatives ranked 1 in the separate rankings and subsequently to state which of those he/she preferred the most. As a result, the data set contains (up to) 3 observations for each respondent.

9.5.2 Step 2: Table Tops

For the table top dataset, 9 different choice models with product attributes as the sole explanatory variables are estimated. The models are presented in table 9.13. As it appears from the last column of the table, a model has been estimated solely based on the data obtained from Rank E, that is the choice between the two alternatives ranked 1 in Rank C and Rank D. In addition, two rank-order logit models and two 1st Choice MNL models have been estimated; one where data from Rank E is included and one where it is not. These models have been estimated in order to test whether or not data from Rank E can be pooled with the data from Rank C and D.

From table 9.13 it is evident that all included product attributes are important determinants of choice, even though the number of significant variables decline from 1st to 3rd Choice. Also, PRICE and CERT are the only significant variables in the model based on Rank E. With reference to the last row of the table, the implicit WTP for certification varies quite significantly between the different models, indicating that it may not be appropriate to pool the data from all ranks and data sets.

Comparing the LRI's of the models representing explosion of the rankings to depth 1, 2 and 3, it is indicated that explosion should not be taken further than to the depth of 1. This goes for the case where Rank E is included as well as for the case where Rank E is excluded.

Table 9.13 Models of table top choice.

	RL - all ranks	**RL - 1st + 2nd Choice**	**1st Choice MNL**	**2nd Choice MNL**	**3rd Choice MNL**
Parameter	β	β	β	β	β
DESIGN1	0,9*** (0,1055)	0,902*** (0,1089)	1,0082*** (0,1253)	1,0552*** (0,1318)	1,2932*** (0,1765)
DESIGN2	0,5788*** (0,1031)	0,6229*** (0,1058)	0,5405*** (0,1284)	0,6156*** (0,1343)	0,4216* (0,1921)
DESIGN3	0,7033*** (0,1021)	0,7511*** (0,1055)	0,7088*** (0,1262)	0,797*** (0,133)	0,7677*** (0,1793)
VAR	-0,139* (0,0709)	-0,1818* (0,0739)	-0,202* (0,082)	-0,2687** (0,0866)	-0,2842** (0,1084)
CERT	0,6265*** (0,0737)	0,6211*** (0,0762)	0,8254*** (0,0845)	0,8334*** (0,0883)	0,9633*** (0,1194)
PRICE	-0,0006*** (0,0001)	-0,0006*** (0,0001)	-0,0006*** (0,0001)	-0,0006*** (0,0001)	-0,0007*** (0,0001)
Log Likelihood	-1032,059	-929,508	-813,41	-710,46	-485,192
ρ^2	0,1	0,102	0,123	0,129	0,164
$\bar{\rho}^2$	0,094	0,096	0,116	0,122	0,153
N	501	340	501	340	501
WTP for CERT	**1.048 DKK/ l. m**	**1.124 DKK/ l. m**	**1.346 DKK/ l. m**	**1.514 DKK/ l. m**	**1.293 DKK/ l. m**

* significant at the 5% level; ** significant at the 1% level; *** significant at the 0,1% level. Numbers in brackets are the standard errors. The implicit WTP for certification estimate, presented in the last row, is obtained by dividing the coefficient estimate for CERT with that for PRICE; it states WTP in DKK/l. m.

Likelihood ratio tests are used to verify what the LRI's indicate. The results of the tests are presented in table 9.14. The hypothesis of equal coefficient vectors can be rejected in all but one of the cases considered.

Table 9.14 Likelihood ratio tests for pooling different levels of rank data for table tops.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice (incl. Rank E)	64,25	12	21,03
1st Choice vs 2nd Choice vs 3rd Choice (excl. Rank E)	66,94	12	21,03
1st Choice vs 2nd Choice (incl. Rank E)	25,4	6	12,59
2nd Choice vs 3rd Choice	11,08	6	12,59
1st Choice Ranks C + D vs Rank E	18,93	6	12,59

Note: Rank C refers to the first ranking of table tops; Rank D to the second and Rank E to the comparison of the alternatives ranked 1 in Rank C and Rank D.

Based on the results of the LR-test presented in the last row of Table 9.14, along with inspection of the model for Rank E (where price is the only variable that is significant), it is excluded in the further analysis. Thus, the rescaled models to be used for the LR-test for proportionality of beta-coefficient vectors are estimated for the data set pertaining solely to Rank C and D.

The results of the rescaled LR-tests are presented in table 9.15. They show that the hypothesis of equality still can be rejected. This suggests that the determinants of choice are different for different ranks – or stated differently, it suggests that respondents make inconsistent choices. Thus, data should not be pooled across ranks, and consequently the further analysis will be based on the 1st Choice MNL model specification.

Table 9.15 Likelihood ratio tests for pooling different levels of rank data for table tops after correcting for different scales.

Data set comparisons	Test-statistic	DF	$\chi^2_{0,95}$
1st Choice vs 2nd Choice vs 3rd Choice (excl. Rank E)	31,97	12	21,03
1st Choice vs 2nd Choice (excl. Rank E)	13,41	6	12,59

Note: μ_m was found to be 0,55 for the comparison of data set 1 and 2; and to be 0,35 for the comparison of data set 1 and 3 (see appendix VII). These values conform to the expectation that the variance associated with lower ranks is higher than that of higher ranks.

9.5.3 Step 3: Table Tops

The results of the univariable regressions for table tops (sub-sample C and D) are presented in the last column of table 9.2, where 12 out of 29 hypotheses regarding interaction terms are potentially significant factors in explaining choice of table tops.

9.5.4 Step 4: Table Tops

Based on the results of the univariable analysis, the preliminary multivariable model for table top choice is given by:

$$
\begin{aligned}
U_{TT} = {} & \beta_{P_I1} \cdot P_INC1 + \beta_{P_I2} \cdot P_INC2 + \beta_{C_I1} \cdot C_INC1 + \beta_{C_I2} \cdot C_INC2 \\
& + \beta_{C_A2} \cdot C_AGE2 + \beta_{C_E2} \cdot C_EDU2 + \beta_{C_R} \cdot C_REG + \beta_{C_R1} \cdot C_REG1 \\
& + \beta_{C_T2} \cdot C_TRUST2 + \beta_{C_T3} \cdot C_TRUST3 + \beta_{C_V1} \cdot C_VFOR1 + \beta_{C_A1} \cdot C_ACCO1 \\
& + \beta_{C} \cdot CERT + \beta_{P} \cdot PRICE + \beta_{D1} \cdot DESIGN1 + \beta_{D2} \cdot DESIGN2 + \beta_{D3} \cdot DESIGN3 + \beta_{V} \cdot VAR
\end{aligned}
$$

Subjecting this model to the model selection procedures available in SAS, the following model results:

$$
\begin{aligned}
U_{TT} = {} & 0{,}0009 \cdot P_INC1^{(*)} + 0{,}001 \cdot P_INC2^{(*)} + 0{,}6968 \cdot C_EDU2^{(*)} + 0{,}8854 \cdot C_TRUST3^{(**)} \\
& + 0{,}9698 \cdot C_ACCO1^{(***)} - 0{,}0015 \cdot PRICE^{(***)} + 1{,}4838 \cdot DESIGN1^{(***)} + 0{,}6047 \cdot DESIGN2^{(**)} \\
& + 1{,}033 \cdot DESIGN3^{(***)} - 0{,}4717 * VAR^{(***)}
\end{aligned}
$$

with: $L(\hat{\beta}) = -\,368{,}3715$, $\rho^2 = 0{,}2138$ and $\bar{\rho}^2 = 0{,}1925$.

The model contains 10 explanatory variables, which seems to be quite a lot. The two variables entered last are P_INC1 and P_INC2 (in stepwise and forward selection). It may be relevant to test, if the addition of these two variables, do in fact improve the fit of the model significantly. A likelihood ratio test is used to test the derived model (i.e. the unrestricted model) against the model, where the coefficients on P_INC1 and P_INC2 are restricted to be zero (i.e. the restricted model). With $K_R = 8$, and $L(\hat{\beta})_R = -\,371{,}9445$ the test statistics, testing the null hypothesis that the unrestricted model is no better than the restricted model, is given by:

$$-2\,(L_R - L_U) = -2\,(-371{,}9445 - (-368{,}3715)) = 7{,}146$$

and is χ^2 distributed with $K_U - K_R = 10 - 8 = 2$ degrees of freedom. Compared with $\chi^2_{2,0.95} = 5{,}99$, the hypothesis can be rejected. This implies that including the two additional variables in the model indeed constitutes a significant improvement of the fit of the model. Consequently, the model presented above is kept as the final model.

As in the cutting board model, CERT as a main effect has disappeared from the model. Instead the utility derived from certification depends on the level of education, trust in eco-labels on products produced in LDC's, type of accommodation, and indirectly, it also varies over income groups. More specifically, it is found that the implicit WTP for certification for a certified table top as opposed to an otherwise identical, but not certified, table top varies across individuals as described in table 9.16.

Table 9.16 Implicit WTP for CERT for table tops (DKK/l.m.) as a function of respondent characteristics.

Respondent characteristics	**Income group (DKK)**		
	< 200.000	200-400.000	> 400.000
C_EDU2	477	1.156	1.408
C_TRUST3	606	1.469	1.790
C_ACCO1	664	1.609	1.960

An implication of the model specification is, that people fitting the base case have a zero WTP. At the other extreme, it is implied that people with 5-6 years of education, who trust eco-labels on products produced in LDC's, who are homeowners and have a household income > 400.000 DKK/year are willing to pay a premium of: (1.408+1.790+1.960) = 5.158 DKK/running meter to get a certified as opposed to a non-certified table top. This estimate appears unrealistically high and will be discussed in chapter 11.

In terms of the variables comprising the final model, it is considered to be an intuitively appealing model, which to a great extent conform to the expectations. Thus, the income effect, expressed through the parameters P_INC1 and P_INC2, increase quite beautifully with increased income. The utility derived from CERT depends on level of education – a proxy for knowledge, e.g. about environmental issues. Utility also depends on whether or not one has trust in environmental labels on products produced in LDC's – a relevant parameter in determining whether or not one gets value for one's money. And lastly, utility depends on whether or not one owns one's home – an indication of the relevance of making more long-term investments such as a high quality wooden table top. In terms of the other attributes, the coefficient on VAR is also as expected. That is, the utility derived from a table top decreases as the variation in the wood – i.e. a proxy for quality – increases. For the design attributes, the DESIGN1 attribute, which is associated with the highest level of utility, is also the most expensive and high quality design. Similarly the base case design, which is the one associated with the lowest level of utility, is also the least expensive design. All in all – the final model constitutes an intuitively appealing model for explaining table top choice.

According to Lovell's RoT, the appropriate significance level to impose on coefficient estimates is approximated by: $\hat{\alpha} = (10/18)\cdot 0{,}05 = 0{,}0278$. Applying this more stringent significance level all coefficients except P_INC1 remain significant. Using the RoT the deflated significance level for the P_INC1 is estimated to be: $\alpha = (18/10)\cdot 0{,}0311 = 0{,}056$. This is so close to the desired level of significance that it is not considered relevant to remove the variable from the model.

9.5.5 Step 5: Table Tops

To see if the extension of the model to include interactions between product attributes and respondent characteristics has improved the fit of the model, the model derived in step 4 can be compared with the 1st Choice MNL model (excl. Rang E) presented in table 9.13. Since the CERT attribute has disappeared from the extended model, the models are no longer nested, implying that the Horowitz (1983) test has to be used.

With $L(0)$ = - 468,5675; model 1 = the model with lowest $\bar{\rho}^2$; model 2 = the model with highest $\bar{\rho}^2$ and z set to (0,1925 - 0,173) = 0,0195 (i.e. the difference in $\bar{\rho}^2$'s between the two models) the test is given by:

$$\Pr\left(\bar{\rho}_2^2 - \bar{\rho}_1^2 > 0{,}0195\right) \leqslant \Phi(-4{,}72) \approx 0{,}000\,001$$

Thus, the test show that the probability that the observed difference in $\bar{\rho}^2$'s would be exceeded is less than 1 out of 1million. Consequently, the observed difference is unlikely to be accidental, and therefore it can be concluded that the model with the highest $\bar{\rho}^2$ - i.e. the model including respondent characteristics – has a better fit than the model with a lower $\bar{\rho}^2$.

9.5.6 Step 6: Table Tops

In addition to eliminating alternatives A, B, C and D, the table top choice sets have been reduced by eliminating on account of the different designs. The results of the tests for violations of IIA for the model of table top choice are presented in table 9.17. The results show, that the hypothesis IIA only can be accepted in three of the cases considered. The violation that is observed when Design3 and Design4, respectively, are eliminated, may be explained by the fact that the only difference between the two designs is that in design 3, the sticks are assembled symmetrically whereas they are assembled at random in the other. Thus, they are quite similar, implying that they might be too close substitutes.

Once again, rejection of IIA is of course somewhat problematic; but, once again, due to practical limitations it is not possible to try other model specifications.

Table 9.17 Hausman & McFadden's (1984) tests for violation of IIA in the model for cutting boards.

Alternative removed	A	B	C	D	Design1	Design2	Design3	Design4
Test Statistic	49,6	32,66	13,75	25,95	15,99	16,64	54,89	115,5
$\chi^2_{0,95}$	18,31	18,31	18,31	18,31	16,92	16,92	16,92	16,92

9.5.7 Step 7: Table Tops

Excluding respondents with a self-reported level of confidence in their ranking below 7, re-estimating the model derived in section step 4 and removing variables that turn insignificant, the following model results:

$$U_{TT} = 0{,}7562 \cdot C_EDU2^{(*)} + 1{,}0524 \cdot C_TRUST3^{(*)} + 0{,}8503 \cdot C_ACCO1^{(***)} - 0{,}0006 \cdot PRICE^{(***)} + 1{,}151 \cdot DESIGN1^{(***)} + 0{,}6854 \cdot DESIGN3^{(***)} - 0{,}3416 \cdot VAR^{(*)}$$

with: $L(\hat{\beta}) = -254{,}03$; $\rho^2 = 0{,}1631$ and $\bar{\rho}^2 = 0{,}1856$.

Comparing this model to the model in section 9.5.4, it is seen that P_INC1, P_INC2 and DESIGN2 are no longer significant (their respective significance levels actually drop to 0,134; 0,987 and 0,636). In terms of the implicit WTP's for certification, they no longer depend on the size of household income. More specifically, they are now estimated to be:

- 1260 DKK/l.m for respondents with 5-6 years of secondary education.
- 1754 DKK/l.m for respondent who have confidence in eco-labels on products produced in LDC's.
- 1417 DKK/l.m for respondents who are homeowners.

9.6 Summary

Prior to estimating the final models, the data sets are tested for choice consistency across ranks. For toilet paper and cutting boards, the lower ranks are found to be associated with greater variances than higher ranks, but there are found no indications of actual inconsistency. For table tops, however, the hypothesis of no inconsistency has to be rejected. Despite the fact that, the RL-all ranks model specification could be applied to the toilet paper and cutting board data sets, the 1st Choice MNL model specification for all three product types is nevertheless chosen.

Subsequently, choice of all three types of products is found to be explained most accurately by a combination of product attributes and respondent characteristics. In terms of WTP for certification, the number of respondent characteristics influencing WTP increase from toilet paper, over cutting boards to table tops. In terms of the effect of information on WTP, it is only in relation to cutting boards that the effect is significant. Tests for violations of IIA show that the assumption is violated for two of the products, but no alternative model specifications are explored. Exclusion of uncertain respondents is shown to have an effect on model composition. Thus, for all three models the number of significant variables decline. The three final models (step 4) are presented in box 9.3

Final model for toilet paper

$$U_{TP} = 1{,}2189 \cdot CERT^{(***)} - 0{,}09827 \cdot PRICE^{(***)} + 0{,}5579 \cdot RECYC^{(***)} + 0{,}4092 \cdot LAY^{(***)} - 0{,}5951 \cdot C_VFOR1^{(*)}$$

with: $L(\hat{\beta}) = -398{,}895$, $\rho^2 = 0{,}2036$ and $\bar{\rho}^2 = 0{,}1936$.

Final model for cutting boards

$$U_{CB} = 0{,}6764 \cdot C_INC2^{(**)} - 0{,}6808 \cdot C_SEX^{(**)} + 0{,}4399 \cdot C_ENV1^{(*)} + 0{,}8462 \cdot C_ENV2^{(**)} + 0{,}4889 \cdot C_INFO^{(*)} - 0{,}0032 \cdot PRICE^{(***)} - 0{,}5818 \cdot HEVEA^{(***)} + 0{,}4474 \cdot BEECH^{(***)}$$

with: $L(\hat{\beta}) = -411{,}1975$, $\rho^2 = 0{,}1604$ and $\bar{\rho}^2 = 0{,}1441$

Final model for table tops

$$U_{TT} = 0{,}0009 \cdot P_INC1^{(*)} + 0{,}001 \cdot P_INC2^{(*)} + 0{,}6968 \cdot C_EDU2^{(*)} + 0{,}8854 \cdot C_TRUST3^{(**)} + 0{,}9698 \cdot C_ACCO1^{(***)} - 0{,}0015 \cdot PRICE^{(***)} + 1{,}4838 \cdot DESIGN1^{(***)} + 0{,}6047 \cdot DESIGN2^{(**)} + 1{,}033 \cdot DESIGN3^{(***)} - 0{,}4717 * VAR^{(***)}$$

with: $L(\hat{\beta}) = -368{,}3715$, $\rho^2 = 0{,}2138$ and $\bar{\rho}^2 = 0{,}1925$.

Box 9.3 Final models (step 4)

10 Estimation of Willingness to Pay

In the present chapter, estimates of mean and median Willingness to Pay (WTP) for certification, derived from the models estimated in the previous chapter will be presented. Subsequently, the relative importance of certification compared to other product attributes will be discussed in relation to the estimated WTP's for the other attributes. Lastly, the WTP's estimated in the study will be adjusted to arrive at estimates of the mean WTP for the Danish population with respect to the three product types. For documentation of the estimates presented in sections 10.2, 10.3 and 10.4, reference is made to appendix VIII.

10.1 Mean vs. Median WTP

Mean WTP specifies the maximum amount that people (i.e. in the relevant sample or population) on average is willing to pay to get a certified, as opposed to a conventional, product. In contrast, median WTP specifies the maximum amount that at least half of the sample or population will be willing to pay to get a certified product.

An advantage of the median WTP measure is that it is unaffected by the presence of aberrant WTP's, which may have a significant impact on mean WTP estimates. According to Garrod & Willis (1999), mean WTP is the most correct measure in relation to assessment of welfare economic effects, since it correctly represent the utility that individuals derive from the good or service being valued.

10.2 WTP for Certification, Toilet Paper

In figure 10.1 the distribution of respondents' WTP for certified toilet paper based on the estimation of the final model in section 9.3.4 is depicted. The distribution is not particularly differentiated across the sample; respondents are either willing to pay a premium of 12,40 DKK/pack or 6,34 DKK/pack. Since the largest proportion of respondents is willing to pay the high premium, median WTP is found to be greater than mean WTP.

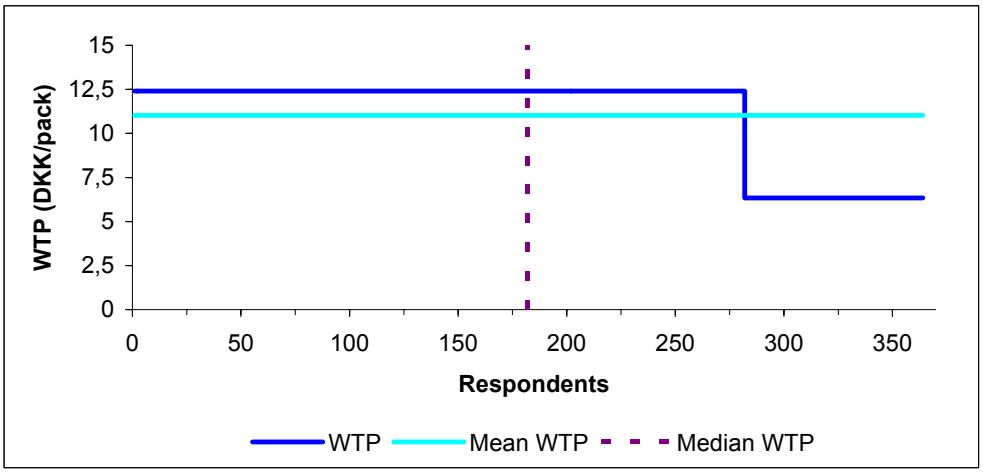


Figure 10.1 Distribution of WTP for certified toilet paper in sample. Based on the model derived in section 9.3.4.

In table 10.1 the mean and median WTP for certified toilet paper resulting from the models estimated in sections 9.3.4, 9.3.5 and 9.3.7, respectively, are presented. Both mean and median WTP are seen to be highest for the final model, and lowest for the certain choices model. However, the differences in WTP across the models are quite small. For the attribute only model and the certain choices model mean and median WTP coincide. This is caused by the fact that no respondent specific variables are included in the model; thus, all respondents are estimated to have the same WTP.

Table 10.1 Mean and median WTP for certification for different toilet paper models (DKK/pack).

	Final model (Step 4)	**Attribute only model (Step 5)**	**Certain choices model (Step 7)**
No. of respondents	364	364	288
Mean WTP	11,03	10,91 (-1,1%)	10,51 (-4,7%)
Median WTP	12,40	10,91 (-12%)	10,51 (-15,2%)

Note: Numbers in brackets denote the percentage change in WTP compared to the final model.

10.3 WTP for Certification, Cutting Boards

Figure 10.2 depicts the distribution of WTP for certified cutting boards resulting from the estimation of the final model in section 10.4.4. Compared to the distribution derived for toilet paper, it is once again seen that median WTP is greater than mean WTP. However, the distribution for cutting boards is much more differentiated across the sample: over 24 different categories, the WTP varies from 628 DKK/cutting board to –213 DKK/cutting board. This reflects that there are actually some respondents for whom certification is associated with a significant disutility. An explanation for this may be that some people

consider environmental labels to be a hoax, which with reference to section 5.4. would be an example of a probability of provision bias.

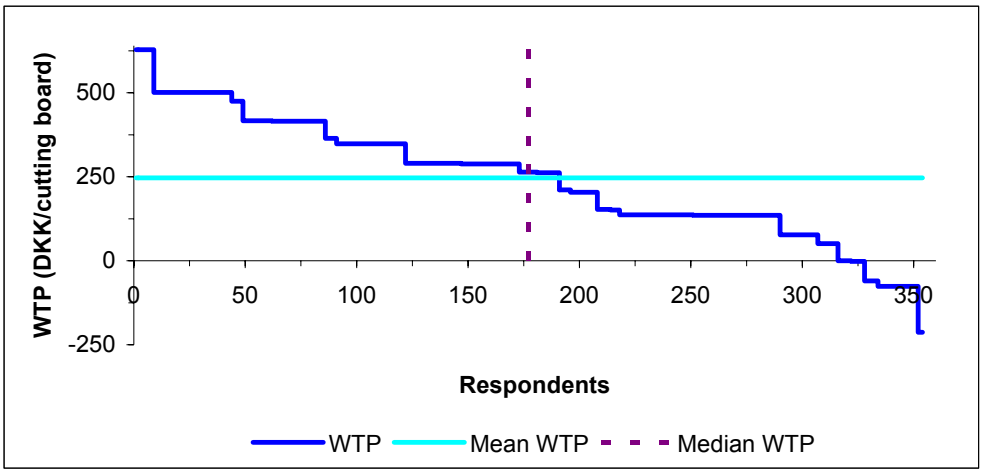


Figure 10.2 Distribution of WTP for certified cutting boards in the sample. Based on the final model derived in section 10.4.4.

The mean and median WTP for the models estimated in section 10.4.4, 10.4.5 and 10.4.7, respectively, are presented in table 10.2. It is seen that the different models result in quite different WTP estimates.

Table 10.2 Mean and median WTP for certification for different cutting board models (DKK/cutting board).

	Final model (Step 4)	**Attribute only model (Step 5)**	**Certain choices model (Step 7)**
No. of respondents	354	354	246
Mean WTP	247	226 (-8,5%)	318 (28,7%)
Median WTP	264	308 (16,7%)	316 (19,7%)

Note: Numbers in brackets denote the percentage change in WTP compared to the final model.

Comparing the final model with the certain choices model, it is a bit surprising that WTP increases when respondents with a low level of confidence in their answers are excluded. This indicates that the uncertain respondents have been most conservative in their valuation of certification. Intuitively, this does not correspond with our prior expectations. Considering that:

a) Certification was the focus of the survey.

b) People were asked to express preferences for a concept that might be completely new to them.

It was expected, that people experiencing a high level of uncertainty would be inclined to put too much emphasis on products being certified – i.e. "sustainable forestry - sounds like a good thing – I think I'll take it". With reference to section 5.3, the above-mentioned points might therefore induce an importance bias.

Despite the fact that no respondent characteristics are included in the attribute only model, it is seen that mean WTP and median WTP are different. This is caused by the inclusion of the information dummy. The high median WTP is caused by the fact that there are more respondents in the sub-sample having received the pamphlet about FSC than in the sub-sample provided with no pamphlet.

10.4 WTP for Certification, Table Tops

In table 10.3 the distribution of WTP for certified table tops derived from the final model estimated in section 10.5.4 is depicted. Even though there are no negative WTP's it appears that the utility derived from buying a certified table top varies significantly across individuals. More specifically, the estimated WTP's vary from 5.158 to 0 DKK/l.m.. Once again, median WTP is greater than mean WTP; the difference, however, is very small.

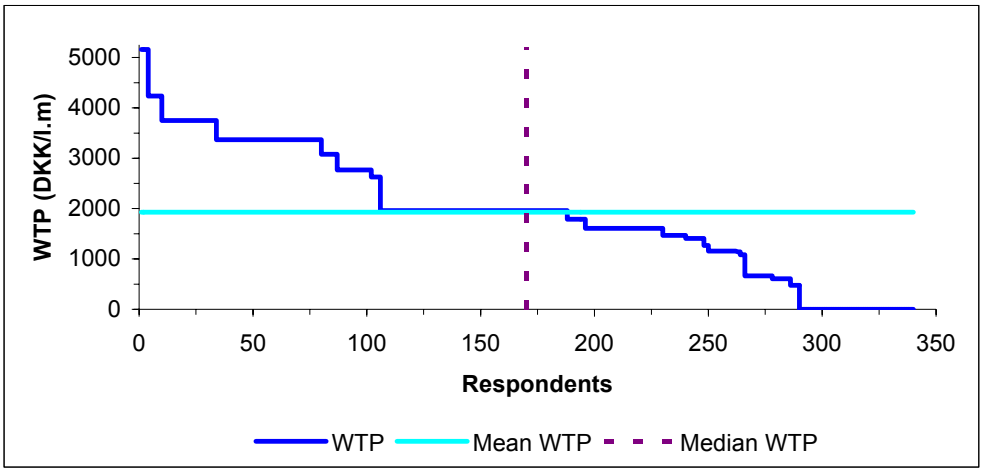


Figure 10.3 Distribution of WTP for certified table tops in the sample. Based on the final model derived in step 10.5.4.

Table 10.3 contains the mean and median WTP derived from the models presented in sections 10.5.4, 10.5.5 and 10.5.7.

Table 10.3 Mean and median WTP for certification for different table top models (DKK/l.m).

	Final model (Step 4)	Attribute only model (Step 5)	Certain choices model (Step 7)
No. of respondents	340	340	226
Mean WTP	1929	1573 (-18,5%)	1820 (-5,7%)
Median WTP	1960	1573 (-19,7%)	1482 (-24,4%)

Note: Numbers in brackets denote the percentage change in WTP compared to the final model.

Comparing the final model with the attribute only model, it is seen that both mean and median WTP increase quite significantly when respondent characteristics are included as explanatory variables. Comparing the final model with the certain choices model, it is seen that the exclusion of uncertain respondents in this case has the expected effect. Thus, both mean and median WTP drops; median WTP even drops to the extent that it now is significantly lower than mean WTP. This indicates that uncertain respondents have had a tendency to state higher than average WTP's. However, the fact that mean WTP > median WTP for the certain choices model also suggests that respondents with very high WTP's generally have a high level of self-reported confidence in their ranking.

10.5 Comparison of WTP across product types

With reference to the estimated mean and median WTP's that have been presented in tables 10.1, 10.2 and 10.3, two trends in the changes in WTP across the different models may be identified. Thus, mean WTP increases – though with very varying percentages – when including respondent characteristics in the model[63]; and for the final models, the estimated median WTP is greater than the mean WTP – though not by much relative to the absolute sizes of the WTP's.

However, figure 10.4 shows that for all products the estimated WTP's appear to be quite high relative to the magnitude of the total prices of the products. In this context it is interesting to note, that there is no definite trend in the change in price premiums across different price categories that the different products present. This is somewhat surprising since the results of other studies show, that the price premium generally tends to decline as the price of the product increases (Jensen *et al.*, 2002; Ozanne & Vlosky, 1997; Spinazze & Kant, 1999).

[63] This may at least in part be due to the linear additive specification of the utility function.

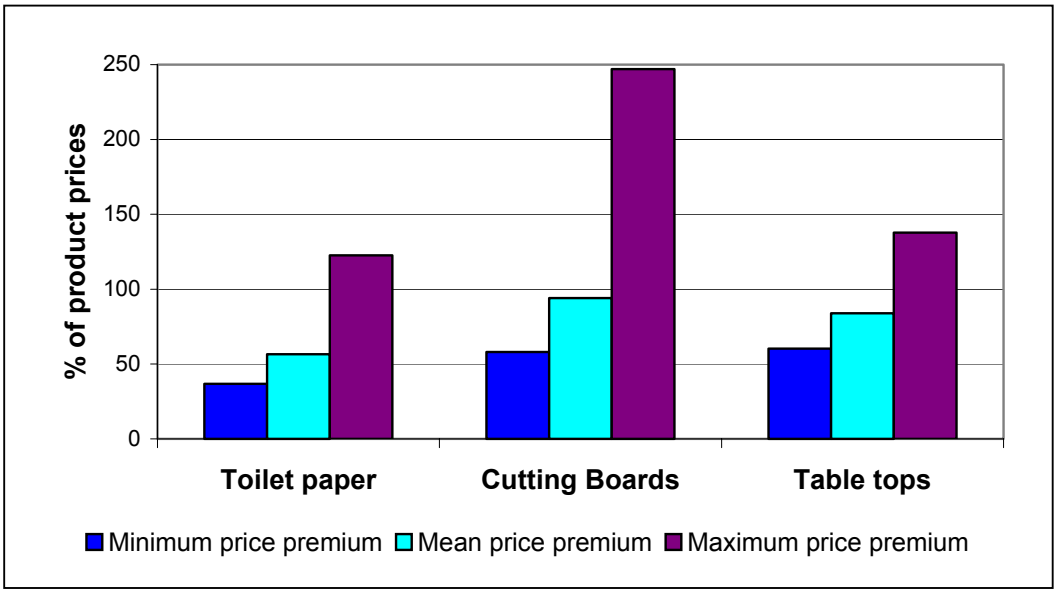


Figure 10.4 WTP for certification relative to product prices.

For all product types it also appears to be the case that WTP varies significantly among individuals. In relation to the choice between products, it should be noted that differences in the WTP's across respondents reflect the fact that the weight attached to the certification attribute, relative to the other attributes of a given product, also varies significantly across respondents.

10.6 Willingness to pay for other product attributes

In the table 10.4 the WTP's for the different product attributes are presented for all three product types. The presented WTP's have been derived from the attribute only models derived in step 5 of the analysis, since comparisons between the WTP for certification and that for the other attributes are more straightforward in this model, than in the final models. However, in addition to becoming more straightforward, the comparisons also become quite crude. Thus, the relative weight associated with different attributes revealed by the attribute only model represents an average across the whole sample. That is, it will not reflect the fact that the relative weight attached to certification varies across the sample.

Table 10.4 Willingness to pay for different product attributes of all three products.

Product	Attribute	Willingness to Pay
Toilet Paper	CERT	10,90 DKK/pack
	RECYC	5,61 DKK/pack
	LAY	3,95 DKK/pack
Cutting Boards	CERT (mean)	226 DKK/cutting board

Product	Attribute	Willingness to Pay
Cutting Boards	BEECH	135 DKK/cutting board
	HEVEA	- 181 DKK/cutting board
Table Top	CERT	1.573 DKK/l.m.
	DESIGN1	2.281 DKK/l.m.
	DESIGN2	907 DKK/l.m.
	DESIGN3	1.543 DKK/l.m.
	VAR	- 672 DKK/l.m.

Note: The WTP estimates have been obtained from the models derived in step 5 of the analysis.

With reference to table 10.4, certification is found to be the attribute that is associated with the highest WTP for toilet paper, whereas it is surpassed by the WTP for wood species for cutting boards and DESIGN1 for table tops[64]. Thus, independent of the type of product, the relative weight attributed to certification compared to other non-monetary attributes is rather high. Thus, based on other studies (Roberts, 1996; Rametsteiner, 1999; Srinam & Forman, 1993), it was not expected that certification would be more important than e.g. quality and design related attributes. Combined with the fact that the estimated WTP's all appear to be quite high relative to the magnitude of the total prices of the products, this suggests that the results may be biased. This, however, pertains to the validity of the results, and will not be discussed until the next chapter.

10.7 Sample WTP vs. Population WTP

In chapter 8, it was made clear that it is not on all accounts that the effective sample can be considered representative of the Danish population. In the following three sub-sections, the extent to which inferences about WTP for certification can be made from this study to the Danish population will be discussed separately for each product type.

10.7.1 Toilet paper

With reference to the final model for toilet paper derived in section 9.3.4 it is seen that the only determinant respondent characteristic was frequency of forest visits. In chapter 8, the sample was found to be representative of the population in terms of forest visit frequency. Thus, provided that the estimates are valid there should be no problems associated with extending the WTP estimated from the sample to the population. The mean WTP found in this study of 11,03 DKK/pack should therefore be representative for the Danish populations.

[64] If the base in the dummy variable for wood species had been Hevea compared to Rosewood, the WTP for Beech would have been 135+181 =316 DKK/board. Thus it would have surpassed the WTP for certification.

10.7.2 Cutting boards

In the final model for cutting boards, which was derived in section 9.4.4, gender, income, level of interest in environmental issues and information were found to have a significant effect on the utility derived from certification – and thereby also on the WTP.

Gender

In terms of gender the sample was found to be representative of the population.

Interest in environmental issues

It has not been possible to assess whether or not the sample is representative in terms of the variables related to respondent's level of interest in environmental issues. In the following it is assumed that the distributions for ENV1 and ENV2 in the sample are representative of the Danish population.

Information

In connection with information (INFO) it is a bit complicated to assess its impact on the general applicability of the WTP estimated from the sample.

Compared to the present information level on the market, the sample is not representative – i.e. half of the population has received the pamphlet about FSC. Consequently, there is an overrepresentation of people with a high level of information in the sample. The estimates will therefore overstate the present WTP of the Danish population.

However, if brochures are distributed to all households or made available in stores selling certified products, then people with little information would potentially be over represented in the sample. This means that the estimated WTP represents an underestimate of true future WTP. In a broader perspective, it may be seen as an indication of the WTP that can be realized, contingent upon the provision of information.

In the following a WTP for the Danish population will is estimated for both the low and high level of information scenarios, respectively.

Income

Individuals living in high-income households are over represented in the sample. Since individuals living in high-income households have a significantly higher WTP for certification, than other individuals, the WTP estimated from the sample will overstate the true WTP of the population.

Assuming that the distributions for ENV1 and ENV2 in the sample are representative of the Danish population, and correction for the overrepresentation of high-income

households in the sample an adjusted WTP for the population can be estimated. The adjusted WTP estimates are presented in Table 10.5 for both the low and high level of information scenarios.

Table 10.5 Adjusted WTP for Cutting Boards (DKK/cutting board)*.

Variable	Proportion of the Danish population	Coefficient	WTP
C_INC2	0,376	0,6764	79,48
C_SEX	0,503	0,6808	-107,10
C_ENV1	0,697	0,4399	87,71
C_ENV2	0,191	0,8462	66,37
C_INFO	0	0,4889	0,00
C_INFO	1	0,4889	152,78
WTP low info			**118,70**
WTP high info			**271,48**

* Danmarks Statistik (2003b) and sample, see appendix II.

In table 10.5 the adjusted average WTP for a Danish consumer is 118,70 DKK/cutting board if the level of information is low and 271,48 DKK/cutting board for consumers presented with a high level of information. Assuming a distribution of information as in this study, the average WTP becomes 192,03 DKK/cutting board. This adjusted WTP is more than 50 DKK lower than the average WTP found in the sample.

10.7.3 Table tops

In the final model for table tops, which was derived in section 9.5.4, education, income, trust in eco-labels on products produced in LDC's and type of residence were found to have a significant effect on the utility derived from certification – and thereby also on the WTP.

Education

Individuals with a high level of education are over represented in the sample. Since individuals with a high level of education have a significantly higher WTP for certification, than other individuals, the WTP estimated from the sample will overstate the true WTP of the population.

Trust in Eco-Labels

It has not been possible to assess whether or not the sample is representative in terms of the variables related to respondents' trust in eco-labels on products produced in LDC's. In the following it is assumed that the distribution in the sample is representative of the Danish population.

Type of Residence
Individuals owning their own homes are over represented in the sample. Since they are found to have a significantly higher WTP for certification, than other individuals, the WTP estimated from the sample will overstate the true WTP of the population.

Income
Individuals living in both middle- and high-income households are over represented in the sample. Since they have a significantly higher WTP for certification, than individuals living in low-income households, the WTP estimated from the sample will overstate the true WTP of the population.

In order to correct for the overrepresentation of middle- and high-income households a weighted average price coefficient based on the true distribution of income in Denmark is estimated, as shown in Table 10.6. It should be noted, than when using the estimated weighted price coefficient, it is assumed that the income is uncorrelated with; the level of education, type of residence and trust in eco-labels on products produced in LDC's.

Table 10.6 Estimation of a weighted average price coefficient for the table top model*.

Income group	Proportion of the Danish population	Price coefficient	Income adjustment	Income adjusted price coefficient	Weighted price coefficient
<200.000	0,325	-0,0015	0	-0,0015	-0,000488
P_INC1	0,299	-0,0015	0,0009	-0,0006	-0,000179
P_INC2	0,376	-0,0015	0,001	-0,0005	-0,000188
Weighted price coefficient					-0,000855

* Danmarks Statistik (2003b).

Assuming that the distribution for TRUST3 in the sample is representative of the Danish population, and correcting for the overrepresentation of homeowners and individuals with a high level of education, two different adjusted WTP's can be estimated for the population. Using the weighted average price coefficient to correct for the overrepresentation of middle- and high-income households the WTP_{wpc} is estimated. In contrast, if it is assumed, that correlation between income and education, trust and homeownership observed in this study is representative of the Danish population the WTP_{cor} is estimated. Both these adjusted WTP estimates are presented in Table 10.7.

Table 10.7 Adjusted WTP for table tops (DKK/l.m.)*

Variable	Proportion of the Danish population	Variable coefficient	WTP_{wpc}	WTP_{cor}
C_EDU2	0,057	0,6969	46,46	64,74
C_TRUST3	0,221	0,8854	228,46	280,62
C_ACCO1	0,615	0,9698	697,66	989,85
WTP			**972,58**	**1335,20**

*Danmarks Statistik (2002c); Danmarks Statistik (2003d) and sample, see appendix II

Table 10.7 shows that the adjusted average WTP for the Danish population lies between 973 and 1335 DKK/l.m., depending on the assumptions on correlation. These WTP's are approximately 30-50 % lower than the WTP of 1929 DKK/l.m. representing the sample.

All in all the WTP adjusted to reflect the true distribution of income, education and accommodations is significantly lower, than the WTP of the sample. The WTP's are however still relatively high, and must be interpreted with caution.

10.8 Summary

The mean WTP for all product types varies significantly across respondents. In terms of the WTP for certification relative to the price of the product, the average price premium is found to be between 57-97%, which is considerably higher than expected. In relation to certification it is also found to be if not the most important then one of the most important attributes of the three products. The estimated WTP's are adjusted for distributional differences of determinant variables between the sample and population, and thereby approximating the WTP for certification of the Danish population.

11 Validation of the Results

In the previous chapter the WTP for certification was presented. Common for all three product types is that the estimated WTP was relatively high, which is of some concern. In this chapter the validity of the results are discussed in the context of the validity criteria and potential biases presented in chapter 5.

11.1 Content Validity

Referring to chapter 5 and 6, the content validity of the study was established through pre-testing of the questionnaire. It was furthermore assessed by Mr. Dubgaard and Mrs. Thorkildsen, whom both have extensive experience within the field of questionnaire and survey design.

In chapter 5, response rates were mentioned as a potential post-survey indicator of content validity. As mentioned the experienced response rate is considered good. Thus, it does not suggest that there is any reason to seriously question the content validity of the study. However, the overrepresentation in the effective sample of highly educated respondents suggests that the perceived complexity of the questionnaire may have discouraged some respondents from participating in the survey.

11.2 Construct Validity

Referring to chapter 5, construct validity refers to the extent to which the results are consistent with economic theory, intuition and prior expectations.

11.2.1 General Issues

The variables in the three models are in accordance with economic theory. The parameter for price is negative in all three models, demonstrating a disutility of price, i.e. the probability of choosing a product decreases as the price increases. In the next subsection the price sensitivity displayed by respondents is investigated in more detail. The subsequent three subsections contain a theoretical validation of the specific qualitative and socio-economic variables included in the models for each of the products.

11.2.1.1 Price Sensitivity

According to economic theory, the slope of the Price Consumption Curve is positive for a normal good. This means that as the price of a good increases the demand decreases and vice versa (Gravelle & Rees, 1992). As mentioned above, the observed choices exhibit this property, which is illustrated in figure 11.1. The probability of choice as a function of price is partitioned into two: 1) the probability of choosing a certified product, and 2) the probability of choosing a conventional product.

Figure 11.1 shows that certified products in general have a higher probability of being chosen compared to conventional products. This is also illustrated by the odds ratios, which specify the ratio between the probability of choosing a certified product to a conventional product. In general, the odds ratios are above 2, which mean that the respondents in the sample have chosen a certified product twice as often as a conventional product. Strikingly, the odds ratios for toilet paper and table tops seem to increase quite drastically as the price increases. This suggests that the respondents' sensitivity to price is significantly smaller for certified products than for conventional products.

It should be noted, that the probabilities presented in figure 11.1 are based on the characteristics of the products ranked 1 in the questionnaire. They are therefore relatively rough estimates of the odds ratio for choosing a certified to a conventional product. The exact odds ratios across all prices, based on the attribute only models (step 5 models) are presented in table 11.1

Table 11.1 Odds ratios in the attribute only models.*

Product	Odds Ratio
Toilet Paper	1.569
Cutting Boards	2.938
Table Tops	2.766

*See appendix IV, V and VI.

The odds ratios in table 11.1 imply that on average, and across prices, the respondents have chosen a certified product 1.5-2.9 times as often as a conventional product. These high odds ratios suggest, that the respondents to a large extent have based their choices on whether or not the product was certified. This is especially evident in the case of cutting boards and table tops.

Based on the odds ratios in table 11.1, and the high level of price insensitivity for certified products illustrated in figure 11.1, it appears reasonable to conclude that, despite the fact that the respondents have behaved in accordance with economic theory, they have to some extent neglected the price of the product, when choosing an eco-labelled product.

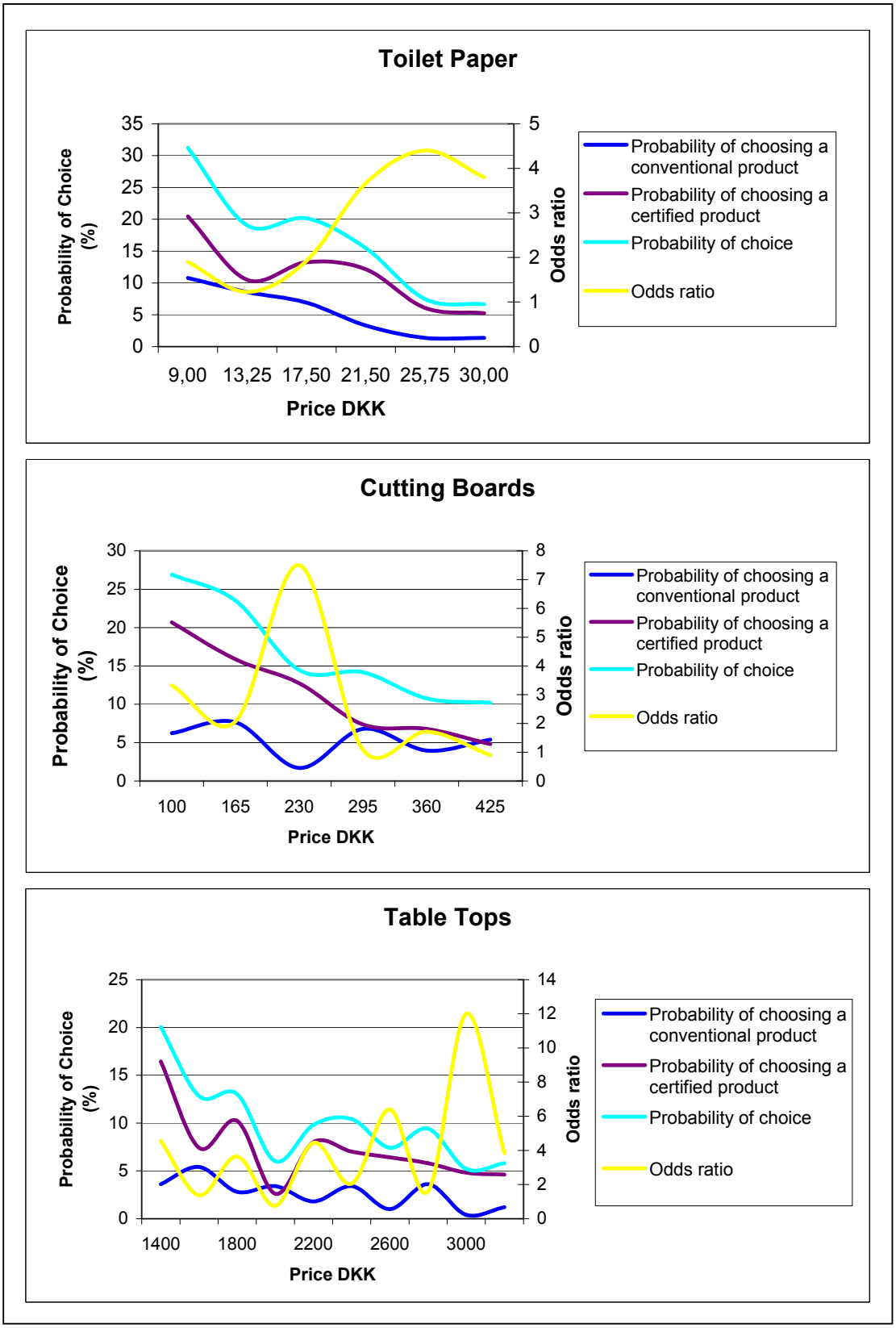


Figure 11.1 Probability of choice as a function of price.

11.2.2 Toilet Paper

In the final model presented in section 9.3.4, the utility derived from toilet paper is determined by the variables: the number of layers, whether the paper is made from new or recycled fibres, the price, whether the paper is certified or not and the frequency of forest visits of the respondent.

As expected (see table 9.2) an increase in the number of layers is found to have a positive influence on utility. Likewise, respondents derive more utility from paper made of recycled fibres, as opposed to paper made of new fibres. These findings correspond with the results in Blamey *et al.* (1999). Intuitively it also makes sense, since a change to a higher attribute level (more layers and more environmentally friendly) must be seen as an increase in the quality of the toilet paper.

Respondents who visit forests often are found to be more reluctant to pay for a certified product. This result is in opposition to the results in Jensen *et al.* (2002), where forest users were found to have a higher WTP for certified wood products than non-users. An explanation could be that certification of forests mainly represents public goods from non-identified forests, as discussed in chapter 2. Given that these goods primarily hold non-use values, a positive relation between frequent visitors (users) and the WTP for certification is therefore a priori not expected. In this study, people whom visit forests often seem to view the Danish forest as managed more sustainable than less frequent users. Subsequently, if present management is perceived as sustainable, there is no or less reason to pay.

The level of information turned out to be insignificant. This may be explained by the fact that the brochure used in the high-level information scenario primarily focuses on forest management on a global scale. The respondents have most likely associated toilet paper with national forest management, why the information may not have been perceived particularly relevant in relation to toilet paper. Another explanation could be that respondents find the level of information less important, when they purchase a relatively cheap product.

Finally, the softness of the toilet paper turned out not to be a significant determinant of choice. This is quite surprising since softness is believed to be an important characteristic of toilet paper. A possible explanation could be the verbal and thereby subjective description of the softness attribute. If so, this contradicts the result from the pre-test of the questionnaire, where the respondents expressed that a verbal description of the softness attribute was sufficient.

11.2.3 Cutting Boards

In the final model presented in section 9.4.4, the utility derived from cutting boards is determined by the variables: wood species, the level of information, level of interest in environmental issues, gender and level of household income.

The variable for Teak is not present in the final model. This is probably due to the similar properties of Rosewood. (the same colour), which is the base variable in the dummy variable for Teak. In that sense, the two wood species may be perceived as close substitutes. In terms of the other wood species, Hevea is the least preferred of all the species and Beech the most preferred. Since Beech is a well-known species, whereas Hevea is less known, this makes sense.

In terms of the different WTP's for men and women, the results of the present study correspond to the expected. Several international and national studies have also found that males generally have a lower WTP for certified/eco-labelled products than females (Jensen *et al.*, 2002; Veisten, 2002; Blend & Ravenswaay, 1999; Bjørner *et al.*, 2002).

People with great interest in environmental issues gain more utility from certification of forests and therefore have a higher WTP than people with moderate interest, who in turn have a higher WTP than people with low interest. This conforms to the expectations displayed in table 9.2.

Respondents living in high-income households have a higher WTP than others. This indicates that the income elasticity of the public goods ensured by certification is positive[65], which is as expected according to Garrod & Willis (1999).

Finally, the level of information turned out to be a significant determinant of WTP. This demonstrates that the content of the brochure in the high-level information scenario has evoked stronger preferences for certification compared to the low-level information provided in the other scenario. Additionally, it demonstrates that people have used the supplied information. With reference to section 5.4, this may be taken to indicate that respondents have been sensitive to scope.

11.2.4 Table tops

In the final model presented in section 9.5.4, the utility derived from table tops is determined by the variables: design, variation of wood, price, level of household income and the respondents' level of education, type of residence and trust in eco-labels on products produced in LDC's.

[65] A positive income elasticity of WTP is defined as: $\varepsilon = (\partial WTP/WTP)/(\partial income/income)$, and illustrates the percentage increase in WTP as a function of a percentage increase in income.

As expected the base design (DESIGN4) was the least preferred and DESIGN1 was the most preferred. Similarly, the parameters for variation in the wood turned out as expected. The negative sign associated with increased variation (lower quality), illustrates that people react to a quality change as expected, and as observed on the real market.

The variables P_INC1 and P_INC2, which are the price sensitivity in relation to income, are positive. Consequently, as predicted by theory, respondents living in households with higher income levels have a smaller marginal disutility of price.

Supported by Blend & Ravenswaay (1999) education has a positive impact on the choice of certified products in the model. A similar effect is found in Bjørner *et al.* (2002) for eco-labelling of detergents.

Respondents who trust eco-labels on products produced in LDC's are found to derive more utility from certification than people who do not. With reference to section 5.4.1 this makes sense, since respondents who have doubted the authenticity of certification should have a lower WTP for certification.

Finally, in terms of type of residence, homeowners are willing to pay more for a certified table top than others. With reference to section 6.2, this may be due to the fact that home-owners have had more well-defined preferences for table tops. They have therefore been more concerned about the characteristics of the chosen products, there among certification.

11.2.5 Summing Up Construct Validity

All in all, the three models performed well in connection with economic theory, and the models as such pass a theoretical validation. Both the qualitative attributes of the products and the socio-economic variables have either the expected coefficients or coefficients that intuitively make sense.

However, it is interesting to note, that the socio-economic variables in the models are different. As the products become more expensive, sophisticated and less frequently purchased, the variables similarly become more economically orientated and complex. Intuitively this is appealing, since it conforms to the expectation that choosing a table top, is more complicated, and requires substantially more attention, than choosing a pack of toilet paper.

However, the extent to which respondents exhibit a lower sensitivity to prices, when a certified product is chosen than when a conventional product is chosen is considered critical. Possible explanations for the observed behaviour will be discussed in section 11.5.

11.3 Criterion Validity

As mentioned, the Contingent Ranking method sets up a hypothetical market for the environmental good, why the answers can be biased (hypothetical). The hypothetical estimates can be validated by comparing them to market prices or to simulated market experiments (Garrod & Willis, 1999). In relation to this study however, it has not been possible to test for the criterion validity.

11.4 Convergent Validity

The convergent validity of the present study will be assessed on two levels. Firstly, the WTP's for certification obtained in the present study will be compared to estimates of WTP's for certification from international studies. Secondly, the WTP's from this study will be compared to the result of Danish studies on consumers' WTP for various eco-labelled products.

11.4.1 Other Studies of WTP for Certification

A few other studies have previously investigated people's WTP for certified wood products. In table 11.2 the WTP's (price premiums) obtained in this study and the studies used for comparison are presented. In the last column the price premiums (i.e. WTP) as a percentage of the price of a conventional product are presented.

Table 11.2 Comparison of WTP for certification across studies.

Study	Method	Product	WTP	Price premium (%)
Jensen *et al.* (2002)	CVM/Open-ended	Shelf	$ 5,20 - 11,72	18 - 41
do	*do*	Chair	$ 24,50 - 55,18	12 - 28
Veisten (2002)	CVM/ Dichotomous choice	Chair	$ 2,79 - 5,16	1 - 1,6
Ozanne & Vlosky (1997)	CVM/Payment card	Chair	$ 14.4	14.4
do	*do*	Dining room set	$ 142	14.2
This study[66] (final model)	Contingent Ranking	Toilet paper, Sample/DK	11,03 DKK	56
do	*do*	Cutting boards, Sample	247 DKK	94
do	*do*	Cutting boards, DK	119 - 271 DKK	45 - 103

[66] The percentage price premium presented in the table is the average price premium, see figure 10.1, table 10.5 and table 10.7.

Study	Method	Product	WTP	Price premium (%)
This study (final model)	Contingent Ranking	Table tops, Sample	1929 DKK	83
do	*do*	Table tops, DK	973 - 1335 DKK	42 - 57

As it can be seen in table 11.2 the percentage price premium for both the sample and the Danish population (DK) are well above the premiums found the other studies. However, comparing the WTP_{DK} for toilet paper, cutting boards (low level of information) and table tops with the most optimistic WTP for a shelf in Jensen *et al.* (2002) they do not deviate considerably.

The studies used for comparison in table 11.2 have all used CVM to estimate the WTP for certified wood products. According to Hanley *et al.* (1998), Foster & Mourato, (2003) and Pickering *et al.* (2002), the difference in methods may explain some of the deviance between WTP_{CVM} found in other studies and WTP_{CRM} in this study[67].

- Hanley *et al.* (1998) found that the WTP_{CE} was almost 6 times higher than WTP_{CVM}. It should though be mentioned that the elicitation method in the CVM, was the open-ended method, which gives relatively low estimates (List & Gallet, 2001).

- In Foster & Mourato (2003) the ratio between WTP_{CE}/WTP_{CVM} was found to be 5:1. Foster & Mourato estimated the value of services provided by the charitable sector in the UK using CVM (double-bounded dichotomous choice) and CE.

- Pickering *et al.* (2002) investigated the WTP for eco-labelled fish products using CVM (open-ended) and CE. The ratio WTP_{CE}/WTP_{CVM} was approximately 1,5:1.

Based on these ratios, the WTP's found in this study are adjusted with the factors 1,5 and 5. In table 11.3 the new adjusted WTP's are compared to the WTP's found in the CVM studies in table 11.2.

[67] CE and CRM are both discrete choice modelling methods, see chapter 5. It is therefore believed that the deviance between CE and CVM also can explain the deviance between CVM and CRM in table 11.2.

Table 11.3 Comparison of WTP for certification across studies when adjusting for different methods.

Study	**Method**	**Product**	**Price premium**		**Price premium (%)**
Jensen *et al.* (2002)	CVM/open-ended	Shelf	$ 5,20 - 11,72		18 - 41
do	*do*	Chair	$ 24,50 - 55,18		12 - 28
Veisten (2002)	CVM/Dichotomous Choice	Chair	$ 2,79 - 5,16		1 - 1,6
Ozanne & Vlosky (1997)	CVM/Payment card	Chair	$ 14.4		14.4
do	*do*	Dining room set	$ 142		14.2
This study	Contingent Ranking		Adj. Factor 5	Adj. Factor 1,5	Price premium (%)
do	*do*	Toilet paper sample/DK	2,21	7,35	11 - 37
do	*do*	Cutting boards, Sample	49	165	19 - 63
do	*do*	Cutting boards, DK	24-79	54-181	9 - 30/20 - 69
do	*do*	Table tops, Sample	386	1286	17 - 55
do	*do*	Table tops, DK	195-267	650-890	8 - 11/28 - 38

The adjustment of the WTP's in table 11.3 increases the convergent validity of this study significantly. Thus, using an adjustment factor of 5 the WTP's estimated for the Danish population are within the range of WTP's found in the other studies. However using an adjustment factor of 1.5 the WTP's are still relatively large, especially for cutting boards and table tops.

11.4.2 WTP for Eco-labelled Non-wood Products

In Denmark several studies on consumers' WTP for different types of eco-labels and products have been carried out the last couple of years. All of the studies are based on choice modelling of the same market data set consisting of 2000 households' daily purchases from 1997-2001. The studies and the estimated WTP's are presented in table 11.4.

Table 11.4 Comparison of WTP for certification with WTP for other eco-labelled products.

Study	Method	Product	Price premium (%)
Bjørner *et al.* (2002)	Discrete choice modelling of panel data	Toilet paper and detergents	10 - 17
Andersen (2002b)	*do*	Organic Eggs	57 - 111
do	*do*	*do*	43 - 50
Wier *et al.* (2002)	*do*	Broad range of organic products	0,4 - 12,84

Bjørner *et al.* (2002) found that Danish consumers are willing to pay a price premium of 10-17 % for the Swan label. Andersen (2002b) estimated a price premium for eco-labelled eggs between 43-111 % depending on the econometric model and the type of eco-label. In Wier *et al.* (2002) the percentage price premiums for a broad range of organic products are between 0,4 and 12,84. These price premiums in Andersen (2002b) are much higher than in Bjørner *et al.* (2002) and Wier *et al.* (2002). At least to some extent, this is probably explained by the fact that the estimates are derived from purchases made in SuperBrugsen, which is a relatively expensive shop.

11.4.3 Summing Up on Convergent Validity

The above mentioned revealed and stated preference studies find a positive WTP for eco-labelled products. To some extent these studies thus support the findings of positive WTP's this study. However, it is important to emphasize that the WTP's for certification found in this study, by far exceed the WTP found in other studies on WTP for certification. Part of this difference may be caused by the application of different methods. Compared to Danish consumers' revealed WTP's for other eco-labels, the estimates in this study are also found to be remarkably high. All in all, this indicates that some elements of the survey instrument have upward biased the WTP. .

11.5 Biases

In the previous section, the WTP's found in this study proved to be much higher than the WTP's found in other studies in relation to eco-labels. Based on this, it is most relevant to investigate the potential biases, which can have caused the high WTP for certification. The discussions in the following subsections refer to those of the biases discussed in chapter 4, which are considered likely to have had a potential impact on the results obtained in the study.

11.5.1 Yea-saying

In chapter 5 it was mentioned, that Choice Experiments (CE), and thereby also the Contingent Ranking Method (CRM) should be less susceptible to yea-saying than to CVM

(Hanley *et al.*, 2002). However, the higher probability of choosing a certified product as opposed to a conventional product, which was illustrated in figure 11.1 indicates that yea-saying might have been a problem in this survey. This means that the respondents in the survey to some extent have chosen a certified product more often than they would have done on the real market.

Based on random utility theory the probability that an individual (n) chooses a certified product ($P_{n|cert}$) is equal to the probability that the utility derived from a certified product ($U_{n|cert}$) is greater than the utility derived from a conventional product ($U_{n|conv}$). That is:

$$(P_{n|cert}) = \Pr(U_{n|cert} > U_{n|conv}) \tag{11.1}$$

If the probability ($P_{n|cert}$) increases due to yea-saying bias, this is equivalent to observing an increase in $U_{n|cert}$ compared to $U_{n|conv}$ (Train, 2003).

Given that the optimal consumption bundle is characterised by (Gravelle & Rees, 1992):

$$\frac{U_{n|conv}}{U_{n|cert}} = \frac{P_{n|conv}}{P_{n|cert}} = MRS_{n|cert,conv} \tag{11.2}$$

an increase in $U_{n|cert}$ is therefore equivalent to a proportional increase in the optimal price for a certified product $P_{n|cert}$. The above-mentioned argumentation thus confirms that yea-saying might have contributed to the estimated high WTP's. However, the observed price insensitivity for certified wood products could also have been initiated by other types of biases or even have been explained by specific motives of the respondents, as will be discussed.

Even though it would have been interesting, it has not been possible to test for yea-saying within the scope of this project. Such a test would require a more specific setting of the questionnaire, see Blamey & Bennett (2001).

11.5.2 Importance Bias

From respondents' answers to general questions in the questionnaire, it is seen that only 19,7 % of the respondents were familiar with the meaning of the FSC eco-label prior to the survey (see appendix II). The remaining 80.3 % of the respondents thus had no prior knowledge on the FSC label, implying that they have had to construct their preferences for certification while completing the questionnaire. In this context the results of the study suggest that respondents in the process of filling out the questionnaire have come to view the FSC -label as an essentially important characteristic of a wood product (see section 10.6). This does to an extent not seem completely plausible. This would explain the high

odds ratios in table 11.1. In accordance with the argumentation presented in the theoretical discussion of yea-saying bias, importance bias can have caused the respondents to overstate the probability of choosing a certified product. Importance bias would consequently result in an overstatement of the WTP for certified products.

However, it can be argued that the importance bias should have a constant effect across products. If so, the observed differences in price insensitivity across products indicate that other biases also have influenced the choice of the respondents.

11.5.3 Probability of Provision Bias

If the respondents in this study have:

- viewed the probability of the outcome of certification as smaller than the true probability,
- directly distrusted the certification scheme or
- distrusted certified products originating from other countries.

this may have biased the estimates of WTP for certification. In the final model for table tops the level of trust in eco-labels on products produced in LDC's was found to have a significant influence on the WTP for certification. This points towards, that less trusting respondents view the outcome of certification in LDC's less reliable/certain, illustrated by their lower WTP. This suggests that some kind of probability of provision bias may be present. However, since neither the model for table tops, nor the setting of the ranking, permit an identification of which country the chosen product originates from this is merely speculation.

11.5.4 Hypothetical Bias

The high WTP could also be explained by the fact, that the choice situation is hypothetical. Hypothetical bias is believed to increase, as the product in focus becomes more complex or less familiar to the respondent (Carson *et al.*, 2001). In the present context, if the respondents have not been familiar with prices on cutting boards and table tops, they might have perceived the presented prices as merely symbolic rather than real[68]. If the respondents thus have viewed the choice scenarios as hypothetical, their choices would be less influenced by the price of the products; choices are for free. In such a situation, the respondents may have been inclined to maximise utility with regards to the other product attributes independent of, or with only a small weight, on the price. This would result in a

[68] Apart from increasing the risk of hypothetical bias, such perceptions may also increase the risk of starting point/range bias, which was discussed in chapter 5.

flat MRS curve between the attributes and price (savings) implying that; a small increase in the product attributes can be substituted by a large reduction in price savings, and thus a relatively high WTP for all of the attributes, including certification. This way the presence of hypothetical bias could serve to explain some of the price insensitivity illustrated in figure 11.1.

11.5.5 Ordering Bias

The price attribute in this study is presented last in all the choice sets. According to Hart-Hansen & Kjær (2003) this could give rise to an ordering bias, which would downward bias the WTP. Looking at the relatively higher probability of choosing a certified product, the associated odds ratios (table 11.1) and the magnitude of the estimated WTP's, an ordering biased is hardly evident. As mentioned the high WTP's found, indicate that the respondents have not paid much attention to the price attribute. If the price had been presented as the first attribute, the respondent might have focused even less on it and result in higher WTP's. As such the WTP's would be upward biased even further.

11.5.6 Framing/Information

As mentioned in chapter 4, CRM is believed to frame the choice situation more realistically than CVM. Nevertheless, the framing of the good in this study might have biased the choice of the respondents. In the scenarios the respondents were not informed on the wood's origin or the exact impact of certification. However, as discussed previously, providing very specific and detailed information on the character and magnitude of the benefits associated with a given product is practically impossible. Consequently, it may be considered likely that the information consumers will be provided with on the actual market will be equally general in nature as the information used in this study.

However, the lack of specific information might have caused some respondents to believe, that by buying certified products would save the Panda, Forest Rhinos etc. rather than just representing a more general contribution to the promotion of sustainable management. If this is the case WTP might have been upward biased. The extent to which such bias may be relevant could have been investigated by including more specific debriefing questions.

11.5.7 Disregard of Budget Constraint

If the respondents have disregarded their budget constraint, their choices become highly hypothetical. It is difficult to verify, if a disregard of the budget constraint has taken place in this study. The relatively high WTP indicates that it might have occurred, but the inclusion of income as a significant determinant of choice in the models for cutting boards and table tops suggests that this is most likely not the case.

11.6 Reliability of WTP estimates

In general, it is very difficult – or at least very costly – to assess the reliability of studies. Basically, it requires that the study is repeated, either by "re-testing" the original sample or by conducting an identical study on a new sample (Hanley & Spash, 1993; Andersen, 2002a). As mentioned in section 5.8, the result obtained from the two sub-samples represented by the two different information scenarios may be compared and used as an indicator of the reliability of the study. In sections 9.3.1 and 9.4.1 it was found that the coefficient estimates pertaining to the two sub-samples were sufficiently identical that data could be pooled across the two scenarios. Thus, indicating that the estimates obtained in the study can be considered reliable.

11.7 Validity, Biases and CRM

As discussed, a number of biases may have contributed to the high WTP's found in this study. Especially yea-saying-, importance-, framing- and hypothetical bias might have increased the WTP for certification. Nevertheless, it should be remembered that the models defining the choice of the respondent have well-behaved economic properties. Moreover, it should be noted that the Likelihood Ration Indexes (LRI's) of the final models, which range from 0.16 for cutting boards, over 0.20 for toilet paper to 0.21 for table tops, may actually be perceived to be quite good (see section 9.2.1). Thus, despite the fact that estimated WTP's are established to be too high, a large proportion of the observed WTP is nevertheless believed to be true. The latter is supported by the observations on market data in Andersen (2002b) where similar high price premiums were estimated.

The magnitude of the hypothetical bias and the associated price insensitivity has been quite surprising. Thus, with reference to chapter 4, the primary motives for choosing CRM opposed to CVM was the framing advantages of the choice situation and its capacity to highlight the other characteristics of the product rather than just focus on certification. Consequently, it was expected that certification would not have attained a quite as prominent role as a determinant of choice as it apparently has. However, in terms of the ability of the CRM to elicit consistent choices from respondents, the method has performed quite well in this study. Thus, it was only in relation to table tops that choice consistency across ranks could not be established. This may be taken to suggest that the potential problems associated with CRM primarily are relevant when dealing with more unfamiliar, less frequently purchased and relatively expensive products.

11.8 Summary

It is found that there is no reason to seriously question the content validity of the study. Assessment of the construct validity shows that estimated models in general are in accordance with economic theory and intuition, however, it is found that respondents

appear to have been somewhat insensitive to price. Due to unavailability of data, the criterion validity of the study cannot be assessed. The convergent validity of the results is to some extent established. Other studies support the findings of positive WTP's, but do not validate the high WTP's found in the present study. Subsequently, it is discussed the extent to which; yea-saying, importance bias, probability of provision bias, hypothetical bias, ordering bias, framing bias and disregard of budget constraints, may have caused the high WTP's. The reliability of the study is supported by the fact that data could be pooled across the two information scenarios. Finally, it is concluded that despite the nice properties of the obtained models, the contingent ranking method has not performed as well as expected.

12 Discussion

In this chapter, the results of the study will be discussed in the context of some of the points made in chapter 3. More specifically, it will be discussed how the results conform to the theoretical expectations in relation to the implications of the weak complementarity and the potential effect of altruistic preferences.

12.1 Estimated vs. Expected Compensating Surplus

As concluded in section 3.8, the method used in this study should theoretically result in an underestimate of the true Compensating Surplus (CS) associated with the transition from Conventional Forest Management (CFM) to Sustainable Forest Management (SFM). However, as discussed in the previous chapter the obtained WTP's for certification are considered to represent an overestimate of respondents' actual WTP, and thereby also that fraction of their CS, which they associate with the purchase of a given wood product. As such, the results do not conform to the theoretical expectations. There are several possible explanations for this, which will be discussed in the following sections.

12.2 Implications of Weak Complementarity

Two important implications of the weak complementarity assumption are; the existence of a choke price and the fact that individuals cannot express their preferences for certification without purchasing a wood product, see section 3.5.

12.2.1 Choke Price

To derive welfare estimates based on weak complementarity between an environmental good and a market good, it is a requirement that there exists a choke price. In the present case, this choke price is represented by the combination of the price and the attributes, excl. certification, where demand for the product falls to zero. In this connection it seems that the respondents in this study have overestimated their choke price. That is, they appear to have made consistent and rational choices between the presented products, but it is suspected that they, when asked if they would actually consider buying the given product, might have disregarded the relative price of the product. Consequently, we may have estimated a positive WTP for respondents whose true choke price has been exceeded. The extent to which such behavior has taken place could be verified by comparing respondents' WTP for the other attributes with the price of these attributes on the market.

12.2.2 Constraints on Preference Expression

With reference to section 3.5 the fact that respondents have been asked to state their preferences for SFM through the purchase of a market good, should theoretically lead to an underestimate of the true CS for SFM. However, due to the hypothetical nature of the present study it may in fact have lead to an overestimate of the WTP for certification, at

least for the specific products. More specifically, with reference to (3.38) it is suspected that some respondents have been inclined to express their total CS, rather than the fraction of their total CS that they associate with the purchase of a specific certified product, namely WTP_{ic}. It is not necessarily the case that they have done so deliberately. As such, it is reasonable if respondents have found it difficult to determine exactly which fraction of their total CS that should be allocated to a given product in order to satisfy (3.38). However, the high WTP's estimated in the study indicate that the preferences expressed by a significant proportion of the respondents reflect the value of a more encompassing good than intended. Seen from this perspective, the estimated WTP's should probably be seen as an estimate of respondents' total WTP for SFM rather than as an estimate of their WTP for the specific certified products.

12.3 Potential Effect of Altruistic Preferences

As discussed in chapter 3, impure altruistic motives are expected to play an important role in relation to WTP for certified products. The high WTP for certification suggests that in general impure altruistic motives are more important than free-rider motives. More specifically, impure altruistic motives are believed to explain at least some of the observed WTP for certification. That is, two aspects indicate that respondents' stated WTP might be influenced by the warm glow of giving:

1. Due to the media attention on deforestation of tropical forests, the warm glow of giving is expected to be higher for saving the rain forest compared to saving the European forests, which all in all must be considered less threatened. As such the warm glow of giving could explain the relatively large WTP's for cutting boards and table tops, compared to toilet paper, since some of these products are made of exotic species.

2. The observed difference in the percentage of the price premiums across product types could be explained by the observation that the marginal utility of warm glow is likely to decrease as a function of the number of voluntary contributions. This means, as the frequency of purchases increases, that part of WTP, which is associated with a warm glow decreases.

Apart from proving a rational explanation for positive WTP's, warm glow may also provide part of the explanation for the high WTP's. Just as there might be a warm glow associated with contributing to an environmental good, there may also be a warm glow connected with *stating* a willingness to contribute to the good (Kahneman & Knetsch, 1992). If it is the latter form of warm glow, that has influenced respondents' statement of their WTP, then the estimates will be biased upwards.

12.4 Summary

The theoretical implications of weak complementarity are found to provide a part of the explanation for the high WTP found in this study. More specifically, it is suspected that respondents might have neglected their choke price when stating if they would actually buy the most preferred product, causing an upward bias of the WTP. Likewise the fact, that the respondents only have been able to express their preferences for sustainable forest management trough their WTP for certification may also have upward-biased the WTP. Finally, the WTP's might have been biased upwards, if the respondents have received a warm glow from the mere act of stating a positive WTP.

13 Conclusion

The focus of the present study has been on estimating Danish consumers' Willingness to Pay (WTP) for certified wood products, and to explore the link between this WTP and the welfare economic benefits associated with the transition from Conventional Forest Management (CFM) to Sustainable Forest Management (SFM). In the present chapter, the main findings of the study will be presented in the context of the objectives of the study.

Based on a survey of 376 Danish consumers using the Contingent Ranking Method (CRM), Danish consumers' WTP for certification for three different types of wood products, toilet paper, cutting boards and table tops, is estimated. The results of the survey demonstrate that on average the respondents are willing to pay a premium of 11 DKK for a pack of certified toilet paper as opposed to conventional toilet paper. For cutting boards the premium is found to be 247 DKK per cutting board, and for table tops the premium is 1.929 DKK per linear meter. Adjusting the estimates from the study to represent the Danish population, the WTP for toilet paper is unchanged, whereas the WTP's for cutting boards and table tops are reduced to 192 DKK per cutting board and 1.335 DKK per linear meter, respectively. Compared to the results of other studies even the adjusted estimates of consumers' WTP for certification obtained in this study are found to be very high. This suggests that the obtained estimates are significant overestimates of the actual WTP.

Comparing the WTP for certification across wood product types, the price premiums as a percentage of the price of the product remain quite stable. Thus there seem to be no definite relation between the prices of the products and WTP. Common for all product types, it was also found that certification was perceived to be, if not the, then one of the most important attributes of the products.

The influence of information on WTP is investigated for both toilet paper and cutting boards, but is only found to have an effect on the WTP for cutting boards. More specifically, providing the respondents with more detailed information about certification significantly augments the WTP for certification. This suggests that the role of information is greater the more expensive the product is.

The CRM has in general performed well in terms of the ability to elicit consistent choices across the ranks. Preferences are though found to be less stable for unfamiliar, less frequently purchased and relatively expensive products, such as table tops. Judging from the high WTP's obtained, the CRM has not performed as well as expected in this study. The inclusion of other attributes has apparently not induced respondents to consider their own true marginal rate of substitution between product attributes. At least respondents appear to have favoured certification to an extent that seems unreasonable.

In terms of the link between the WTP for certification and the welfare economic value of the benefits associated with SFM, it is found that estimates of WTP for certification, due to several factors, only are likely to provide a lower bound estimate of the true Compensating Surplus (CS) associated with the transition from conventional to sustainable forest management. Furthermore, it is found that even the derivation of such a lower bound estimate is quite complicated, since it requires aggregation of WTP across all purchases of wood products.

Despite the fact, that the WTP's obtained in this study for several reasons cannot be used to derive an exact estimate of the CS associated with SFM. The fact that the respondents have stated a positive WTP may nevertheless be taken to indicate that Danish consumers have quite strong preferences for the environmental benefits associated with certification. In this connection, it is suspected that part of the explanation for the high WTP estimates obtained in this study may be that respondents have expressed their WTP for the benefits associated with SFM rather than their WTP for a given certified product. If this is in fact the case, the results of the present study suggest that the welfare economic gain associated with SFM is of significant proportions. As a concluding comment, it should however be emphasised that the costs associated with the transition from CFM to SFM need to be known, before jumping to any conclusions regarding the overall welfare economic implications of SFM.

References

Aakerlund, N. F. (1998): *Contingent Ranking Studie af Danskernes Præferencer for Skovkarakteristika,* Master Thesis, IØSL, KVL, Copenhagen. pp. 148.

Aalborg Træindustri A/S (2002): On the internet http://www.aalborg-massiv.dk.

Abt, K. and E. Sills (2003): *Forests in a Market Economy*, Kluwer Academic Publishers, (forthcoming).

Adamovicz, W. L.; Boxall, P.; Luckert, M. K.;Phillips; W. E. and W. A. White (1996): *Forestry, Economics and the Environment*, CAB International.

Adamovicz, W. L.; Boxall, P.; Williams, M. and J. Louviere (1998a): Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation, *American Journal of Agricultural Economics*, Vol. 80, pp. 64-75.

Adamowicz, W. L, Louviere, J. J. and J. Swait (1998b): *Introduction to Attribute-Based Stated Choice Methods. Final Report to Resource Valuation Branch*, Damage Assessment Center, US National Oceanic and Atmospheric Administration, Department of Commerce.

Adamovicz, W. L.; Boxall, P. C.; Louviere, J. J.; Swait, J. and M. Williams (1999): Stated-Preference Methods for Valuing Environmental Amenities. In Bateman, I. J. and K. G. Willis (eds.): *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries,* Oxford University Press, New York, pp. 460-479.

Adger, W. N.; Brown, K.; Cervigni, R. and D. Moran (1995): Total Economic Value of Forests in Mexico, *Ambio*, Vol. 24 (5), pp. 286-295.

Allison, P. D. (2001): *Logistic Regression Using the SAS System: Theory and Application*, SAS Institute Inc., Cary, NC, pp. 287.

Alpizar, F.; Carlsson, F. and P. Martinsson (2001): Using Choice Experiment for Non-Market Valuation, *Working Paper in Economics*, no. 52, Department of Economics, Göteborg University, Sweden.

Andersen, I. (2002a): *Den Skinbarlige Virkelighed – Vidensproduktion inden for Samfundsvidenskaberne*, Samfundslitteratur, Frederiksberg, pp. 353.

Andersen, L. M (2002b): Consumer Evaluation of Environmental and Animal Welfare Labelling: An Econometric Analysis on Panel Data Using Mixed Multinomial Logit, *Working* Paper, no. 6, Consumer Demand for Organic Foods, AKF.

Anderson, L. R. and C. R. Holt (2002): Experimental Economics and Public Choice. Prepared for: Rowley, C. and F. Schneider (eds) *The Encyclopedia of Public Choice,* forthcoming.

Andreoni, J. (1989): Giving with Impure Altruism: Applications to Charity and Richardian Equivalence, *Journal of Political Economy*, Vol. 97 (6), pp.1447-1459.

Andreoni, J. (1990): Impure Altruism and Donations to Public Goods: A theory of Warm-Glow Giving, *The Economic Journal*, Vol. 100, pp. 464-477.

Andreoni, J and J. Miller (2002): Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, *Econometrica*, Vol. 70, pp. 737-753.

Arrow, K. J. and A. C. Fischer (1974): Environmental Preservation, Uncertainty and Irreversibility, *The Quarterly Journal of Economics*, Vol. 88 (2), pp. 312-319.

Arrow, K. J.; Solow, R.; Portney, P. R.; Leamer, E. E.; Radner, R. and H. Schuman (1993): Report of the National Oceanic and Atmospheric Administration (NOAA) Panel on Contingent Valuation, *Federal Register*, Vol. 58 (10), pp. 4016-614.

Banzhaf, M. R., Johnson, F. R. and K. E. Mathews (2001): Opt-out Alternatives and Anglers' Stated Preferences, In Bennett, J. and R. Blamey (eds): *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar, pp. 157-177.

Bateman, I. J. and K. G. Willis (1999): Introduction and Overview. In Batemann, I. J. and K. G. Willis (eds): *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, Oxford University Press, pp. 1-14.

Batsell, R. R. and J. J. Louviere (1991): Experimental Analysis of Choice, *Marketing Letters*, Vol. 2, pp. 199-214.

Baumol, W. and W. Oates (1988): *The Theory of Environmental Policy*, 2nd Edition, Cambridge University Press.

Becker, G. (1981): *A Treatise of the Family*, Harvard University Press, Cambridge and London.

Beggs, S.; Cardell, S. and J. Hausman (1981): Assessing the Potential Demand for Electric Cars, *Journal of Econometrics*, Vol. 17(1), pp. 1-20.

Ben-Akiva, M. and S. R. Lerman (1985): *Discrete Choice Analysis*, The Massachusetts Institute of Technology.

Ben- Akiva, M.; Morikawa, T. and F. Shirioshi (1991): Analysis of the Reliability of Preference Ranking Data, *Journal of Business Research,* Vol.23 (3), pp. 253-268.

Bennett, J. and W. Adamowicz (2001): Some Fundamentals of Environmental Choice Modelling. In Bennett, J. and R. Blamey (eds): *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar, pp. 37-69.

Bennett, J. and R. Blamey (2001): Environmental Choices and Values. In Bennett, J. and R. Blamey (eds): *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar, Cheltenham, UK, pp. 1-10.

Bjørner, T. B.; Hansen, L. G.; Russel, C. S. and T. Olsen (2002): *The Effect of the Nordic Swan Label on Consumers' Choice*, AKF Forlaget, Copenhagen, pp. 79.

Blamey, R. K. (1996): Citizens, Consumers and Contingent Valuation: Clarification and the Expression of Citizen Values and Issue-Options. In Adamovicz, W. L.; Boxall, P.; Luckert, M. K.;Phillips; W. E. and W. A. White (eds): *Forestry, Economics and the Environment*, CAB International, pp. 103-133.

Blamey, R.; Common, M. and J. Quiggin (1995): Respondents to Contingent Valuation Surveys: Consumer or Citizens, *Australian Journal of Agricultural Economics*, Vol. 39, pp. 263-88.

Blamey, R. K.; Bennett, J. W.; M. D.; Louviere, J. J. and M. D. Morrison (1999): Validation of a Choice Model Involving Green Product Choice, *Choice Modelling Report no. 10*, May 1999. The University of New South Wales, Australia, pp. 28.

Blamey, R. K. and J. W. Bennett (2001): Yea-saying and Validation of a Choice Model of Green Product Choice. In Bennett, J. and R. Blamey (eds): *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar, pp. 178-201.

Blend, J. R. and E. O. van Ravenswaay (1999): Measuring Consumer Demand for Ecolabeled Apples, *American Journal of Agricultural Economics*, Vol. 81(5), pp. 1072-1077.

Bockstael, N. E. and K. E. McConnell (1999): The Behavioural Basis of Non-market Valuation. In Herriges, J. A. and C. L. Kling (eds): *Valuing Recreation and the Environment*, Edward Elgar, 1999.

Bockstael, N. E. (2003): Personal Comment, Professor at the College of Agricultural & Resource Economics (AREC), Maryland University.

Boyle, K. J.; Holmes, T. P.; Teisl, M. F. and B. Roe (1999): A Comparison of Conjoint Analysis Response Formats, *American Journal of Agricultural Economics*, Vol. 82 (2), pp. 441-454.

Bunch, D. S., Louviere, J. J. and D. Anderson (1996): A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes, *Working Paper*, Graduate School of Management, University of California, Davis.

Börkey, P.; Glachant, M. and F. Lévêgue (1999): *Voluntary Approaches in Environmental Policy. An Assessment*, Study prepared for the OECD, Cerna, Ecole Nationale Supérieure des Mines de Paris, OECD, Paris.

Carlsson, F and P. Martinsson (2001): Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to Valuation of the Environment, *Journal of Environmental Economics and* Management, Vol. 41 (2), pp. 179-192.

Carson, R. T.; Louviere, J. J.; Anderson, D. A.; Arabie, P.; Bunch, D. S.; Hensher, D. A.; Johnson R. M.; Kuhfeld; W. F.; Steinberg, D.; Swait J.; Timmermans H. and J. B. Wiley (1994): Experimental Analysis of Choice, *Marketing Letters*, Vol. 5, pp. 351-368.

Carson, R. T; Flores, N. E.; Martin, K. and J. Wright (1996): Contingent Valuation and Revealed Preference Methodologies, Comparing Estimates for Quasi-Public Goods; *Land Economics*, Vol. 72, pp. 80-99.

Carson, R. T.; Flores, N. E. and N. F. Meade (2001): Contingent Valuation: Controversies and Evidence, *Environmental & Resource Economics*, Vol. 19, pp. 173-210.

Caswell, J. A. (1997): *Uses of Food Labelling Regulations.* OECD, Paris, pp. 33.

CBD (2001): *The Value of Forrest Ecosystems*, CBD Technical Series No. 4, Secretariat of the Convention on Biological Diversity, Montreal, pp. 67.

Cerda, A. and V. Lira (2002): The Economics of Sustainable Forest Management Certification, Paper presented at *2002 World Congress of Environmental and Resource Economists*, Monterey California, June 24-27 2002.

Champ, P. A.; Bishop, R. C.; Brown, T. C. and D. W. McCollum (1997): Using Donation Mechanism to Value Nonuse Benefits from Public Goods, *Journal of Environmental Economics and Management*, Vol. 33, pp. 151-162.

Chapman, R. G. and R Staelin (1982): Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model, *Journal of Marketing Research*, Vol. 19 (August), pp. 288-301.

Ciriacy-Wantrup, S. V (1947): Capital Returns from Soil-Conservation Practices, *Journal of Farm Economics*, Vol. 29, pp. 1181-1196.

Cochran, W.D and G.M Cox (1992*): Experimental Designs*, Second Edition, Wiley Classics Library Edition.

Cornes, R. and T. Sandler (1996): *The Theory of Externalities, Public Goods, and Club Goods,* 2nd Edition, Cambridge University Press, New York.

Cummings, R.; Harrison, G. and E. E. Rutström (1995): Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?, *American Economic Review*, Vol. 85, pp. 260-266.

Curtis, J. A. and K. E. McConnell (2002): The Citizen Versus Consumer Hypothesis: Evidence From a Contingent Valuation Survey, *The Australian Journal of Agricultural and Resource Economics*, Vol. 46, pp. 69-84.

Danmarks Statistik (2002a): *Statistisk Tiårsoversigt 2002- Tema om Skatter og Skattetryk* , Danmark Statistik, pp. 208.

Danmarks Statistik (2002b): *Nøgletal på Postnumre- Befolkning, Bolig, Indkomst, Statistisk*, Danmark Statistik

Danmarks Statistik (2002c): *Statistik Årsbog2002*, Danmark Statistik

Danmarks Statistik (2003a): BOL4, *Statistikbanken.*

Danmarks Statistik (2003b): BIL4, *Statistikbanken.*

Danmarks Statistik (2003c): PRIS9, *Statistikbanken.*

Danmarks Statistik (2003d): FU4, *Statistikbanken.*

Deaton, A. and J. Muelbauer (1980): *Economics and Consumer Behaviour*, Cambridge University Press, New York.

Dellaert, B. G. C; Brazell, J. D. and J. Louviere (1999): The Effect of Attribute Variation on Consumer Choice Consistency, *Marketing Letters*, Vol. 10 (2), pp. 139-147.

Deshazo, J. R. and G. Fermo (2002): Designing Choice Sets for Stated Preference Methods: The Effect of Complexity on Choice Consistency, *Journal of Environmental Economics and Management*, Vol. 44, pp. 123-143.

Dhar, R. and I. Simonson (2003): The Effect of Forced Choice on Choice, *Journal of Marketing Research*, Vol. 11, pp. 146-160.

Domencich, T and D. McFadden (1975): *Urban Travel Demand. A Behavioural Analysis*, North Holland, Amsterdam.

Dubgaard, A.; Sandøe, P.; Gamborg, P. and A. Larsen (1999): Bæredygtighed – Økonomi, Etik og Energi, *National Økonomisk Tidsskrift*, Vol. 137 (3), pp. 256-283.

Edward, G. C. and R. A. Zeller (1994): Reliability and Validity Assessment. In Lewis-Beck, M. S (eds): *Basic Measurement. International Handbook of Quantitative Applications in the Social Science*, Vol. 4, Sage Publications, London, pp. 1-58.

FAO (2002): Forrest Certification Update for the UNECE Region, Summer 2002, *Geneva Timber and Forest Discussion Papers,* FAO Timber Section, Geneva.

FAO (1997): *State of the Worlds Forests*, FAO, Rome.

Foster, V. and S. Mourato (1999): Elicitation Format and Part-whole Bias: Do Contingent Valuation and Contingent Ranking Give the Same Result; *CSERGE Working Paper*, GEC 99-17.

Foster, V. and S. Mourato (2002): Testing for Consistency in Contingent Ranking Experiments, *Journal of environmental Economics and Management*, Vol. 44, pp. 309-328.

Foster, V. and S. Mourato (2003): Elicitation Format and Sensitivity to Scope, Do Contingent Valuation and Choice Experiments Give the Same Results, *Environmental and Resource Economics*, Vol. 24, pp. 141-160.

Freeman, M. A. (1979): *The Benefits of Environmental Improvement*, The John Hopkins University Press, Baltimore.

Freeman, M. A. (1993): *The Measurement of Environmental and Resource Values, Theory and Methods*, Resource for the Future.

Garrod, G. and K. Willis (1998): Using Contingent Ranking to Estimate the Loss of Amenity Value for Inland Waterways from Public Utility Service Structures, *Environmental and Resource Economics,* Vol. 12, pp. 241-247.

Garrod, G and K. Willis (1999): *Economic Evaluation of the Environment; Methods and Case Studies*, Edward Elgar Publishing Limited, UK.

Georgiou, S.; Whittington, D.; Pearce, D. and D. Moran (1997): *Economic Values and the Environment in the Developing World*, Edward Elgar Publishing Limited, Cheltenham UK.

Gravelle, H. and R. Rees (1992): *Microeconomics,* Second Edition, Longman, New York.

Green, C. and S. Tunstall (1999): A Psychological Perspective. In Bateman, I. J. and K. G. Willis (eds.): *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries,* Oxford University Press, New York, pp.207-257.

Green, P. E and V. R. Rao (1971): Conjoint Measurement for Quantifying Judgemental Data, *Journal of Marketing Research*, Vol. 8 (August), pp. 355-363.

Green, P. E. and V. Srinivasan (1978): Conjoint Analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research,* 5(September), pp. 103-123.

Greene, 2003: *Econometric Analysis*, 5th Edition, Prentice Hall, Upper Saddle River, N.J, pp. 1026.

Gyrd-Hansen, D. (2002): Personal Comment, Professor at Institute of Public Health and Health Economics, University of Southern Denmark.

Gyrd-Hansen, D. (2003): Personal Comment, Professor at Institute of Public Health and Health Economics, University of Southern Denmark.

Gyrd-Hansen, D. and J. Søgaard (2001): Analysing Public Preferences for Cancer Screening Programmes, *Health Economics*, Vol.10, pp. 617-634.

Hanemann, M. W. (1984): Welfare Evaluation in Contingent Valuation Experiments with Discrete Responses, *American Journal of Agricultural Economics,* August, pp.332-341.

Hanemann, M. W. (1994): Valuing the Environment through Contingent Valuation, *Journal of Economic Perspective*, Vol. 8, pp. 19-43.

Hanemann, W. M. (1995): Contingent Valuation and Economics. In: Willis, K. G. and J. T. Corkindale (eds): *Environmental Evaluation – New Perspectives*, CAB International.

Hanley, N. and C. L. Spash (1993): *Cost-Benefit Analysis and the Environment*, Edward Elgar, Chelterham, UK.

Hanley, N.; Shogren, J. F. and B. White (1997): *Environmental Economics in Theory and Practice*, Macmillan Press Ltd. London.

Hanley, N.; MacMillan, D.; Wright, R. E.; Bullock, C.; Simpson, I.; Parsisson, D. and B. Craptree (1998):Contingent Valuation Versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland, *Journal of Agricultural Economics*, Vol. 49, pp.1-15.

Hanley, N. Mourato, S. and R. E. Wright (2001): Choice Modelling Approaches: A Superior Alternative for Environmental Valuation?, *Journal of Economic Surveys*, Vol. 15 (3), pp. 435-462.

Hanley, N, Adamowicz, V. and R.E. Wright (2002): *Do Choice Experiments Pass the ScopeTest? A Test of Scope in a Choice Experiment Examining the Benefits of Water Quality Improvements*, Paper presented at 2002 World Congress of Environmental and Resource Economists, Monterey California, June 24-27 2002

Hart-Hansen, K. and T. Kjær (2003): *Eliciting Preferences for Topical Treatment of Psoriasis – An Application of the Discrete Choice Experiment*, Master Thesis, IØSL, KVL, Copenhagen, pp. 152.

Hausman, J. and D. McFadden (1984): Specification Tests for the Multinomial Logit Model, *Econometrica*, Vol. 52 (5), pp.1219-1240.

Hausman, J. A. and P. A. Ruud (1987): Specifying and Testing Econometric Models for Rank-Ordered Data, *Journal of Econometrics,* Vol. 34, pp. 83-104.

Hensher, D. A (1994): Stated Preference Analysis of Travel Choices: The State of Practice, *Transportation*, Vol. 21, pp.107-133.

Hicks, J. R. (1943): The Four Consumer's Surpluses, *The Review of Economic Studies,* Vol.11 (1), pp.31-41.

Holmes, T. and K. J Boyle (2003): Stated preference methods for valuing forest attributes. In K. Abt and E. Sills (eds): *Forests in a Market Economy*, Kluwer Academic Publishers (forthcoming).

Horowitz, J. L. (1983): Statistical Comparison of Non-nested Probabilistic Discrete Choice Models, *Transportation Science*, Vol. 17(3), pp.319-350.

Hosmer, D.W and S. Lemeshow (2000): *Applied Logistic Regresseion*, 2nd Edition, John Wiley &Sons Inc.

Huber, J. and J. Pinnell (1994): The Impact on set quality and Choice Difficulty on the Decision to Defer Purchase, *Working Paper*, the Fuqua School of Business, Duke University.

Huber, J. and K. Zwerina (1996): The Importance of Utility Balance in Efficient Choice Designs, *Journal of Marketing Research*, Vol. 33 (August), pp. 307-317.

Jaeger, S. R.; Hedderley, D. and H. J. H. MacFie (2001): Methodological Issues in Conjoint Analysis: A Case Study, *European Journal of Marketing*, Vol. 35 (11/12), pp. 1217-1237.

Jensen, K.; Jakus, P.; English, B. and J. Menard (2002): Willingness to Pay for Environmentally Certified Hardwood Products by Tennessee Consumers, *Study Series No. 01-02*, Department of Agricultural Economics, University of Tennessee.

Johansson, P.-O. (1987): *The Economic Theory and Measurement of Environmental Benefits*, Cambridge University Press.

Johansson, P.-O. (1991): *An Introduction to Modern Welfare Economics*, Cambridge University Press.

Johansson, P.-O. (1992): Altruism in Cost-Benefit Analysis, *Environmental and Resource Economics,* Vol. 2 (6), pp. 605-613.

Johansson-Stenman, O. (1997): Optimal Pigouvian Taxes with Regard to Altruism, *Land Economics*, Vol. 73(3), pp.297-308.

Johansson-Stenman, O. and H. Svedsäter (2003): Self Image and Choice Experiments: Hypothetical and Actual Willingness to Pay, *Working Paper in Economics*, no. 94, Department of Economics, Göteborg University.

Jyllands-Posten (2003): Svindel med Miljømærker, *Jylland-Posten*, 10.04.03.

Junckers Industrier A/S (2002): On the internet http://www.junckers.dk.

Kahn, J. R. (2002): *The Development of Markets and Economic Incentives for Sustainable Forestry: Application to the Brazilian Amazon*, Working Party on Global and Structural Policies, Working Group on Economic Aspects of Biodiversity, OECD, Paris.

Kahneman. D. and J. L. Knetsch (1992): Valuing Public Goods: The Purchase of Moral Satisfaction, *Journal of Environmental Economics and Management*, Vol. 22, pp. 57-70.

Karl, H. and C. Orwat (1999): Economic Aspects of Environmental Labelling. In Folmer, H. and T. H. Tietenberg (eds): *The International Yearbook of Environmental and Resource Economics 1999/2000: A Survey of Current Issues*, Edward Elgar Publishing Limited, pp. 107-170.

Keeney, R. L. and H. Raiffa (1993): *Decisions with Multiple Objectives –Preferences and Value Trade-offs*, Cambridge University Press, UK.

Kiker, C. F. and P. E. Putz (1997): Ecological Certification of Forest Products: Economic Challenges, *Ecological Economic*, Vol. 20, pp. 31-57.

Klemperer, D. W (1996): *Forest Resource and Finance*, McGraw-Hill Inc.

Kotchen, M. J. and S. D. Reiling (1999): Do Reminders of Budget Constraints and Substitutes Influence Contingent Valuation Estimates? Another Comment, *Land Economics*, Vol. 75(3), pp. 478-82.

Kristensen, P. (2003): Personal Comment, FSC-sekretariatet, Barrit Langgade 182
DK-7150 Barrit, Denmark.

Krutilla, J. V. (1967): Conservation Reconsidered, *American Economic Review*, Vol. 57 (4), pp. 777-786.

Kuhfeld, W., Tobias, R. D. and M. Garrat (1994): Efficient Experimental Design with Marketing Research Applications, *Journal of Marketing Research*, Vol. 31, pp. 545-557.

Kuhfeld, W. F (2000): *Multinomial Logit, Discrete Choice Modelling, Marketing research Methods in the SAS System*, Report from SAS Institute Inc.

Lampietti, K. J. and J. A. Dixon (1995): *To see the Forest for the Trees*, Environmental Department Papers, The World Bank.

Lancaster, K. J. (1966): A New Approach to Consumer Theory, *Journal of Political Economy,* Vol. 74 (2), pp. 132-157.

Lareau, T. J. and D. A. Rae (1989): Valuing WTP for Diesel Odor Reductions: An Application of the Contingent Ranking Technique, *Southern Economic Journal*, Vol. 55, pp. 728-742.

List, J. A. and C. A. Gallet (2001): What Experimental Protocol Influences Disparities Between Actual and Hypothetical Stated Values, *Environmental and Resource Economics*, Vol. 30 (3), pp. 241-254.

List, J. A. (2002): Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Fields Auctions for Sportscards, *The American Economic Review*, Vol. 91 (5), pp. 1498-1507.

Loomis, J.; Gonzales-Caban, A. and R. Gregory (1994): Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?, *Land Economics*, Vol. 70 (4), pp. 499-506.

Loureiro, M. L.; McClusky, J. J. and R. C. Mittelhammer (2000): *Willingness to Pay for Sustainable Agriculture products*, Report to the Federal State Marketing Improvement program, United States Department of Agriculture.

Louviere, J. J. and G. Woodworth (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach based on Aggregate Data, *Journal of Marketing Research*, Vol. 20 (November), pp. 350-367.

Louviere, J. J. (1988): Analyzing Decision Making. Metric Conjoint Analysis*, Sage University Paper Series on Quantitative Applications in the Social Sciences,* no. 67,Sage Publications, Newbury Park, pp. 95.

Louviere, J. J.; Hensher, D. A. and J. D. Swait (2000): *Stated Choice Methods, Analysis and Application*, Cambridge University Press.

Lovell, M. C. (1983): Data Mining, *The Review of Economics and Statistics,* Vol. 65(1), pp. 1-12.

Luce, R. D. (1959): *Individual Choice Behaviour: A theoretical Analysis*. John Wiley & Sons, New York.

Luce, D. R and P. Suppes (1968): Preference, Utility and Subjective Probability, In Luce, D. R.; Bush R. B. and E. Galanter (eds): *Handbook of Mathematical Psychology*, Vol. 3, John Wiley & Sons Inc.

Luce, R. D. and J. W. Turkey (1964): Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement, *Journal of Mathematical Psychology,* Vol. 1(1), pp. 1-27.

Lusk, J. L. and T. C. Schroeder (2002): Incentive Compatibility of Choice Experiments, *Working Paper,* Kansas State University.

Mackenzie, J. (1993): A Comparison of Contingent Preference Models, *American Journal of Agricultural Economics,* Vol. 75 (August), pp. 593-603.

Mazzota, M and J. Opaluch (1995): Decision Making when Choices are Complex: A tests of Heiners Hypothesis, *Land Economics*, Vol. 71, pp. 500-515.

Marschak, J. (1960): Binary-Choice Constraints and Random Utility Indicators, In Arrow, K. (eds): *Standford Mathematical Studies in the Social Science IV*, Standford University Press, Stanford CA, pp. 312-329.

Marshall, A. (1920): *Principles of Economics. An Introductory Volume*, 8th Edition. London.

McCullough, D. (2002): A User's Guide to Conjoint Analysis, *Marketing Research,* Summer 2002, pp.19-23.

McFadden, D. (1974): Conditional Logit Analysis of Qualitative Choices, in Zarembka. P. (eds): *Frontiers in Econometrics*, Academic Press, pp. 312-329
.

McFadden, D. (1986): The Choice Theory Approach to Market Research, *Marketing Science*, Vol. 5 (4), special Issue on Consumer Choice Models, pp. 275-297.

Mcfadden, D. (2001): Economic Choices, *American Economic Review,* Vol. 90, pp. 351-378.

Milgrom, P. (1993): Is Sympathy an Economic Value? Philosophy, Economics and the Contingent Valuation. In Hausman, J.A. (eds): *Contingent Valuation. A Critical Assessment*, Elsevier Science Publisher B. V.

Mitchell, R. C. and R. T. Carson (1989): *Using Surveys to Value Public Goods: The Contingent Valuation Method,* Resources for The Future, Washington, D. C.

Moran, D. (2001): *Market Creation for Biodiversity: The Role Organic farming in the EU and US*, Draft paper prepared for the OECD Working Group on Economic Aspects of Biological Diversity (WGEAB).

Morrison, M. D.; Blamey, R. K.; Bennett, J. W. and J. J. Louviere (1996): A Comparison of Stated Preference Techniques for Estimating Environmental Values*, Choice Modelling Report no. 1*, The University of New South Wales, Australia. pp. 30.

Morrison, M. D.; Blamey, R. K. and J. W. Bennett (2000): Minimising Payment Vehicles in Contingent Valuations Studies, *Environmental and Resource Economics*, Vol. 16, pp. 407-422.

Morrison, M. and K. Boyle (2001): Comparative Reliability of Rank and Choice Data in Stated Preference Models, *Faculty of Commerce Working Paper Series,* No. 4/01, Charles Stuart university, Bathurst. pp. 25.

Navrud, S (2001): *The Contingent Valuation Method Step-by-step*, paper presented at Ph.D. course on Economic Valuation of Natural Resources and Environmental Goods, 13-17 May, 2002 in Esbjerg Denmark, http://www.sam.sdu.dk/fame/.

Nicosia, F. M. (1974): Toward an Empirical Theory of Consumer Behaviour Based on the Economics of Goods-Characteristics, Book Review, *Journal of Marketing Research,* Vol. 11 (February), pp. 115-120.

Nunes, P. A. L. D. and Y. E. Riyanto (2001): *Certification (Eco-labelling) as a Policy Instrument to Signal the Non-market Values of Bio-diversity: A Critical Review*, Preliminary Draft.

Nunes, P. A. L. D (2003): personal comment, Free University Amsterdam, Ca' Foscari of Venice and Fondazione Eni Enrico Mattei, Italy.

OECD (2001): *Background Paper*, Experts Workshop on Information and Consumer Decision-making for Sustainable Consumption, 16-17 January, OECD Headquarters, Paris.

Ozanne, L. K. and R. P. Vlosky (1997): Willingness to Pay for Environmentally Certified Wood Products: A Consumer Perspective, *Forest products Journal*, Vol. 47(6), pp. 39-48

Pearce, D. W. and R. K. Turner (1990): *Economics of Natural Resources and the Environment*, Harvester Wheatsheaf.

Pearce, D. W. and J. J. Warford (1993): *World without End: Economics, Environment and Sustainable Development*, Oxford University Press.

Pearce, D. W.; Putz, F. and J. K. Vanclay (1999): A Sustainable Forest Future, *Working Paper*, GEC 99-15 CSERGE.

Pearce, D. W.; Özdemiroglu, E.; Bateman, I.; Carson, R. T.; Day, B.; Hanemann, M.; Hanley, N.; Hett, T.; Jones-Lee, M.; Loomes, G.; Mourato, S.; Sudgen, R. and J. Swanson (2002): *Economic Valuation with Stated Preference Techniques. Summary Guide*, Department of Transport, Local Government and the Regions, London, pp. 94.

Pickering, H.; Jaffry, S.; Whitmarsh, D.; Wattage, P.; Ghulam, Y.; Frere, J.; Roth, E.; Nielsen, M. and R. J. Lopes (2001): *Market Driven Incentive Structures for Sustainable Fisheries Management. Final Project Report.* Centre for the Economics and Management of Aquatic Resources, University of Portsmouth, UK. IME, University of Southern Denmark, Campus Esbjerg, in Conjunction with DIAFE, Copenhagen, Denmark. University of Evora, Portugal. pp. 259.

Price, C. (1989): *The Theory and Application of Forest Economics*, Basil Blackwell Limited.

Price, C. (2000): *Exact Values and Vague Products? Contingent Valuation and Passive Use Values*, Discussion Paper, Unpublished.

Punj, G. N and R. Staelin (1978): The Choice Process for Graduate Business Schools, *Journal of Marketing research*, Vol. 15, pp. 588-98.

Rametsteiner, E. (1999): The Attitude of European Consumers Towards Forest and Forestry, *Unasylva*, No. 196, pp. 42-48.

Ratchford, B. T. (1975): The New Economic Theory of Consumer Behaviour: An Interpretive Essay, *Journal of Consumer Research*, Vol. 2 (September), pp. 65-75.

Richards (1999): Internalising the Externalities of Tropical Forestry: A Review of Innovative Financing Mechanisms, *European Union Tropical Forestry Paper*, No. 1, ODI, London.

Roe, B.; Teisl, M. F.; Levy, A. and M. Russell (2001): US Consumers' Willingness to Pay for Green Electricity, *Energy Policy*, Vol. 29, pp. 917-925

Rolfe, J. C. and J. W. Bennett (2000): Testing for Framing Effects in Environmental Choice Modelling, *Choice Modelling report no. 13*, The University of New South Wales, Australia, pp. 19.

Rolfe, J. C. and J. W. Bennett (2001): Framing Effects. In Bennett, J. and R. Blamey (eds): *The Choice Modelling Approach to Environmental Valuation*, Edward Elgar, pp. 202-224.

Rolfe, J.; Bennett, J. and J. J. Louviere (2002):"Stated Values and Reminders of Substitute Goods: Testing for Framing Effects with Choice Modelling", *The Australian Journal of Agricultural and Resource Economics*, Vol. 46, pp. 1-20.

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy*, Vol. 82(1), pp. 34-55.

Ryan, M (1999): Using Conjoint Analysis to take Account of Patients Preferences and go Beyond Health Outcomes: An Application to in Vitro Fertilization, *Social Science and Medicine*, Vol. 48, pp. 535-546.

Sagoff, M. (1988): *Cambridge Studies in Philosophy and Public Policy Series*, Cambridge University Press.

Samuelson, P. A. (1955): Diagrammatic Exposition of a Theory of Public Expenditure, *The Review of Economics and Statistics*, Vol. 37 (4), pp. 350-356.

Samuelson, P. A. (1954): The Pure Theory of Public Expenditure, *The Review of Economics and Statistics*, Vol. 36 (4), pp. 387-389.

Sedjo, R. A.; Sampson, R. N. and J. Wisniewski (1997): *Economics of Carbon Sequestration in Forestry*, Lewis Publishers, London.

Small, K. A. and H. S. Rosen (1981): Applied Welfare Economics with Discrete Choice Models*,* *Econometrica*, Vol. 49(1), pp. 105-130.

Southgate, D. (1998): *Tropical Forest Conservation. An Economic Assessment of the Alternatives in Latin America*, Oxford University Press, New York.

Spekva (2002): On the internet http://www.spekva.dk.

Spinazze, M. C. and S. Kant (1999): Market Potential for Certified Forest (Wood) Products in Ontario, Canada, *The Forestry Chronicle,* Vol. 75(1), pp. 39-47.

Sriram, V. and A. M. Forman (1993): The Relative Importance of Products' Environmental Attributes: A Cross-Cultural Comparison, *International Marketing Review*, Vol. 10 (3), pp. 51-70

Stevens, T.H.; More T. A. and R.J. Glass (1994): Interpretation and Temporal Stability of CV Bids for Wildlife Existence: A Panel Study, *Land Economics*, Vol. 70 (3), pp. 355-64.

Stevens, T. H., DeCoteau, N.E and C. E. Willis (1997): Sensitivity of Contingent Valuation to Alternative Schedule, *Land Economics*, Vol. 73, pp. 140-148.

Sudgen, R. (1999): Alternatives to the Neoclassical Theory of Choice. In Bateman, I. J. and K. G. Willis (eds.): *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries,* Oxford University Press, New York, pp. 152-180.

Swait, J. and W. Adamowicz (1996): The Effect on Choice Environment and Task Demands on Consumer Behaviour: Discriminating Between Contribution and Confusion, *Staff Paper*, Department of Rural Economy, Faculty of Agriculture & Forestry and Home Economics, University of Alberta, Canada.

Teisl, M. F.; Newman, F.; Buono, J. and M. Hermann (2000): Designing Effective Environmental Labels for Forest Products: Results of Focus Group Research, *Miscellaneous Report*, No. 420, Maine Agriculture and Forest Experiment Station, University of Maine.

Teisl, M. F. and B. Roe (2000): Environmental Certification – Informing Consumers about Forest Products, *Journal of Forestry*, February.

Train, K. (2003):"*Discrete Choice Methods with Simulation*", Cambridge University Press.

Thurstone, L. (1927):"A Law of Comparative Judgement", *Psychological Review*, Vol. 34, pp. 273-286.

Upton, C. and S. Bass (1995): *The Forest Certification Handbook*, Earth Scan Limited, London.

Varian, H. R. (1992): *Microeconomic Analysis*, 3rd Edition, W.W. Norton & Company Inc, USA.

Veisten, K. (2002): Potential Demand for Certified Wood Products in the United Kingdom and Norway, *Forest Science*, Vol. 48(4), pp. 776-778.

Vriens, M.; Loosschilder, G. H.; Rosbergen, E. and D. R. Wittink (1998): Verbal Versus Realistic Pictorial Representations in Conjoint Analysis with Design Attributes, *Journal of Product Innovation Management*, Vol. 15, pp. 455-67.

WCED (1987): *Our Common Future*, Oxford University Press, Oxford.

Weisbrod, B. A. (1964): Collective-Consumption Services of individual-Consumption Goods, *The Quarterly Journal of Economics*, Vol. 78 (3), pp. 471-477.

Wier, M.; Hansen L.G.; Andersen L. M. and K. Millock, (2002): *Willingness to Pay for Organic Foods: A Comparison between Survey Data and Panel Data from Denmark*, OECD Workshop on Organic Agriculture 23-26 September 2002, Washington D.C.

Wiser, R. and S. Pickle (1997): *Green Marketing. Renewables and Free-Riders: Increasing Consumer Demand for Public Goods*, Lawrence Berkley National Laboratory, LBNL-40632, Berkley.

Worldwatch Institute (2003): *State of the World 2003*, Washington DC.

Zwerina, K.; Huber, J. and W. F. Kuhfeld (1996): *A General Method for Constructing Efficient Choice Designs*, Report from SAS Institute Inc.