Choice modelling, non-use values and benefit transfer

Author:
Morrison, Mark D

Publication Date:
1998

DOI:
https://doi.org/10.26190/unsworks/6909

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/60783 in https://unsworks.unsw.edu.au on 2023-09-15
Various researchers have recommended the use of conjoint analysis for estimating non-market values, given perceived limitations with the contingent valuation method. The use of conjoint analysis for benefit transfer has also been proposed because of the capacity to value different environmental improvements using the results from a single survey. This thesis focuses on one conjoint technique, known as choice modelling (CM). The purpose of this thesis is to investigate the validity of using CM to estimate non-use values, and the validity of extrapolating these values across different contexts.

This thesis has three parts. The literature relevant to using CM is reviewed in the first part. The economic theory underlying the estimation of non-use values is explored. Five techniques that can be used for estimating non-use values are evaluated to determine which is most appropriate for use in this thesis. Methodological issues associated with applying CM are reviewed.

The results from the CM applications are presented in the second part of the thesis. The two case study sites - the Macquarie Marshes and the Gwydir Wetlands - are described. The results from the process of questionnaire design are detailed. This involved the use of eight focus groups and a pretest. The data are analysed using discrete choice regression techniques. The estimated models are generally robust, although specification problems were identified with rudimentary models. Two hypotheses are tested using these results. The first is that CM can produce valid estimates of non-use values. Relatively strong support is found for this hypothesis, providing support for the use of CM to estimate non-use values. The second hypothesis is that non-use values exist for non-environmental outcomes. Support for this hypothesis is mixed: non-use values appear to exist for non-environmental outcomes only in certain cases.

The third part of the thesis focuses on benefit transfer. The literature pertaining to the use of benefit transfer is examined, and two hypotheses concerning the validity of benefit transfer across either sites or populations are tested. The evidence suggests that neither type of transfer is valid, however the former appears to be the least problematic.

Declaration relating to disposition of project report/thesis

I am fully aware of the policy of the University relating to the retention and use of higher degree project reports and theses, namely that the University retains the copies submitted for examination and is free to allow them to be consulted or borrowed. Subject to the provisions of the Copyright Act 1968, the University may issue a project report or thesis in whole or in part, in photostate or microfilm or other copying medium.

I also authorise the publication by University Microfilms of a 350 word abstract in Dissertation Abstracts International (applicable to doctorates only).

Signature: ________________________________
Witness: ________________________________
Date: ________________________________
CHOICE MODELLING, NON-USE VALUES AND BENEFIT TRANSFER

By

Mark D Morrison

September 1998

A thesis submitted for the degree of Doctor of Philosophy at the University of New South Wales
Disclaimer

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgment is made in the thesis. Any contribution made to the research by colleagues, with whom I have worked at UNSW or elsewhere, during my candidature, is fully acknowledged.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged.

14/9/98

Mark Morrison
Acknowledgments

I would like to sincerely thank my supervisor, Associate-Professor Jeff Bennett, for his help in completing this thesis. Jeff is an excellent teacher as well as a friend. He is always available for discussion on new ideas and critical issues. He comments quickly and constructively on written material. I believe that I have learned much from Jeff over the last few years and for this I am grateful.

I would like to thank my co-supervisor Dr Richard Kingsford from the NSW National Parks and Wildlife Service. I would like to thank Richard for believing in the research that I was wanting to do and providing the necessary seed money. I would also like to thank Richard for his assistance in gathering ecological data for use in the questionnaires.

I would like to thank Dr Russell Blarney for his assistance in designing the questionnaires used in this thesis, and in conducting several of the focus groups. I have learnt a lot about taking time to carefully consider problems from Russell. Russell, Jeff and I have worked together as colleagues designing and refining the experiments presented in this thesis and elsewhere. The discourse between the three of us has been very valuable for developing the choice modelling technique.

Thanks must also go to Professor Jordan Louviere. Jordan first cultivated my interest in the area of choice modelling, and has provided valuable assistance over the last few years. In particular, I would like to thank him for teaching
me experimental design, and for commenting on draft questionnaires and providing advice on modelling issues.

I would like to thank the Land and Water Resources Research and Development Corporation for providing the bulk of the money that funded this research. I would also like to thank Mr Drew Collins from the NSW Environment Protection Authority for providing additional funding half way through the project when it was looking as if it was going to be completed part time.

I would also to thank the faculty at the School of Economics and Management for their support. The past few years have been very enjoyable and I will be sad to leave the School.

I would also like to thank my wife Jenny for her patience whilst being a ‘PhD widow’, and her constant love and support.
Dedication

To my dear wife Jenny and my daughter Jessica

Psalm 103:2
Abstract

Various researchers have recommended the use of conjoint analysis for estimating non-market values, given perceived limitations with the contingent valuation method. The use of conjoint analysis for benefit transfer has also been proposed because of the capacity to value different environmental improvements using the results from a single survey. This thesis focuses on one conjoint analysis technique, known as choice modelling (CM). The purpose of this thesis is to investigate the validity of using CM to estimate non-use values, and the validity of extrapolating these values across different contexts.

This thesis has three parts. The literature relevant to using CM is reviewed in the first part. The economic theory underlying the estimation of non-use values is explored. Five techniques that can be used for estimating non-use values are evaluated to determine which is most appropriate for use in this thesis. Methodological issues associated with applying CM are reviewed.

The results from the CM applications are presented in the second part of the thesis. The two case study sites—the Macquarie Marshes and the Gwydir Wetlands—are described. The results from the process of questionnaire design are detailed. This involved the use of eight focus groups and a pretest. Next the results from three surveys are presented. The data are analysed using discrete choice regression techniques. The estimated models are generally robust, although specification problems were identified with rudimentary models. Two hypotheses are tested using these results. The first is that CM can produce valid estimates of non-use values. Relatively strong support is found for this hypothesis, providing support for the use of CM to estimate non-use values. The second hypothesis is that non-use values exist for non-
environmental outcomes. Support for this hypothesis is mixed: non-use values appear to exist for non-environmental outcomes only in certain cases.

The third part of the thesis focuses on benefit transfer. The literature pertaining to the use of benefit transfer is examined, and two hypotheses concerning the validity of benefit transfer across either sites or populations are tested. The evidence suggests that neither type of transfer is valid, however the former appears to be least problematic.
# Table of Contents

## Chapter 1 Introduction

1.1 The Policy Context  
1.2 Hypotheses  
1.3 Thesis Structure  

## Part 1 Literature Review

## Chapter 2 Non-use Values and Economic Theory

2.1 Introduction  
2.2 Option Value  
  2.2.1 Demand side option value  
  2.2.2 Supply-side option value  
2.3 Quasi-Option Value  
2.4 Existence Value  
  2.4.1 Existence value and motives  
2.5 Total Economic Value  
2.6 Measures of Consumer Surplus  
2.7 Aggregation  
2.8 Conclusions  

## Chapter 3 Techniques for Estimating Non-use Values

3.1 Introduction  
3.2 Contingent Valuation  
  3.2.1 Description  
  3.2.2 Welfare estimation  
  3.2.3 Other issues  
3.3 Choice Modelling  
  3.3.1 Description
3.3.2 Welfare estimation 52
3.3.3 Other issues 54

3.4 Contingent Ranking 57
3.4.1 Description 57
3.4.2 Welfare estimation 59
3.4.3 Other issues 60

3.5 Contingent Rating 62
3.5.1 Description 63
3.5.2 Welfare estimation 66
3.5.3 Other issues 69

3.6 Paired Comparison 70
3.6.1 Description 70
3.6.2 Welfare estimation 73
3.6.3 Other issues 73

3.7 Conclusion 74

Chapter 4 Methodological Issues in Choice Modelling 77
4.1 Introduction 77

4.2 Choice Set Design 78
4.2.1 Framing 79
4.2.2 Labelled versus generic alternatives 81
4.2.3 Inclusion of a constant base alternative 82
4.2.4 Selection of attributes 83
4.2.5 Selection of attribute levels 84

4.3 Experimental Design 86
4.3.1 Combining resource use alternatives to form choice sets 90
4.3.2 Choosing between different experimental designs 96
4.3.3 Dealing with dominated and implausible alternatives 100

4.4 Model Selection 102
6.2 The Macquarie Marshes Focus Groups 135
  6.2.1 Macquarie Marshes focus group logistics 135
  6.2.2 Wetland attributes 137
  6.2.3 Reactions to the draft questionnaire 138
6.3 The Gwydir Wetlands Focus Groups 147
  6.3.1 Gwydir Wetlands focus group logistics 148
  6.3.2 Differences between the Macquarie Marshes and Gwydir Wetlands questionnaires 149
  6.3.3 Reactions to the Gwydir Wetlands questionnaire 152
6.4 The Macquarie Marshes and Gwydir Wetlands Questionnaires 161
6.5 Conclusion 167

Chapter 7 Survey Results 169
  7.1 Introduction 169
  7.2 Survey Logistics 170
  7.3 Survey Statistics 172
  7.4 Multinomial Logit Models 173
    7.4.1 Multinomial logit models with choice set attributes 175
    7.4.2 MNL models with socioeconomic and attitudinal variables 178
    7.4.3 MNL models with socioeconomic, attitudinal and questionnaire evaluation variables 184
  7.5 Hypothesis Tests 187
  7.6 Conclusion 195

Chapter 8 Aggregation 197
  8.1 Introduction 197
  8.2 Dealing with Divergences between Sample and Population Characteristics 200
    8.2.1 Adjustment of mean sample values 201
8.2.2 Dealing with non-responses 203

8.3 A Comparison of Approaches for Aggregating Sample Estimates 207

8.3.1 Use of independent variables 208
8.3.2 Weighted regression analysis 209
8.3.3 Weighted averaging 210
8.3.4 Assuming non-respondents have zero willingness to pay 211
8.3.5 Classifying non-respondents 211

8.4 Conclusion 213

Part 3 Benefit Transfer 216

Chapter 9 Benefit Transfer 217

9.1 Introduction 217
9.2 The Use and Development of Benefit Transfer 218
9.3 The Determinants of Value Estimates 220
9.4 Tests of Benefit Transfer 225
  9.4.1 Reliability over time 225
  9.4.2 Transfers across sites 227
  9.4.3 Transfers across populations 229
9.5 Conclusion 230

Chapter 10 Benefit Transfer Tests 232

10.1 Introduction 232
10.2 MNL Models Used for Benefit Transfer Tests 233
10.3 Transferability Across Sites 235
10.4 Transferability Across Populations 245
10.5 Conclusion 251
Chapter 11 Conclusion

10.1 Summary
10.2 Main findings and their Implications for Policy and Research
10.3 Caveats

Bibliography

Appendix 1 The Macquarie Marshes Draft Questionnaire
Appendix 2 The Gwydir Wetlands Draft Questionnaire
Appendix 3 Questionnaires used in the Macquarie Marshes and Gwydir Wetlands Surveys
Appendix 4 Alternative Choice Set Formats
   A4.1 Alternative Choice Set Formats Trialed in the Macquarie Marshes Focus Groups
   A4.2 Alternative Choice Set Formats Trialed in the Gwydir Wetlands Focus Groups
Appendix 5 Further Modelling Results
   A5.1 Tests for IIA Violations
      A5.1.1 Multinomial logit models with choice set attributes
      A5.1.2 MNL models with socioeconomic and attitudinal Variables
      A5.1.3 MNL models with socioeconomic, attitudinal and questionnaire evaluation variables
   A5.2 The Effect of Experimental Design on the Macquarie Marshes Model
   A5.3 Other Models
List of Tables and Figures

Table 1.1:  Summary of hypotheses 7
Table 1.2:  The structure of the thesis 10

Figure 2.1:  Existence value and relative scarcity 23
Figure 2.2:  Compensating and equivalent surplus 33

Table 3.1:  The dichotomous choice CVM format 46
Table 3.2:  Example of one choice set in a choice modelling questionnaire 49
Table 3.3:  Example from a contingent ranking questionnaire 58
Table 3.4:  Example from a contingent rating questionnaire 63
Table 3.5:  Example from a paired comparison survey 71

Table 4.1:  A simple experimental design 87
Table 4.2:  A 2\textsuperscript{j} experimental design 92
Table 4.3:  Available balanced incomplete block designs 93
Table 4.4:  An L\textsuperscript{MN} design 95
Table 4.5:  Attribute differences in a foldover and an L\textsuperscript{MN} design 98
Table 4.6:  A choice set with a dominated alternative 100
Table 4.7:  A choice set with an implausible alternative 100
Figure 4.1:  A tree structure for transport choice 104

Figure 5.1:  The Macquarie Valley 121
Figure 5.2:  The Gwydir Valley 127

Table 6.1:  Age and sex statistics for the Sydney and Dubbo focus groups 136
Table 6.2: Draft Macquarie Marshes questionnaire choice set format 139
Table 6.3: Age and sex statistics for the Sydney and Moree focus groups 148
Table 6.4: The non-balanced choice set format 151
Table 6.5: Example of a choice set from the Macquarie Marshes questionnaire 166

Table 7.1: Sampling statistics 172
Table 7.2: Reasons for non-response 172
Table 7.3: Socio-demographics of the survey respondents 173
Table 7.4: Definitions of variables 174
Table 7.5: Multinomial logit models with choice set attributes 178
Table 7.6: Multinomial logit models with socioeconomic and attitudinal variables 183
Table 7.7: Multinomial logit models with socioeconomic, attitudinal and questionnaire evaluation variables 186
Table 7.8: T-statistics for tests of content validity 189

Table 8.1: Multinomial logit models estimated using weights 210
Table 8.2: Reasons for non-response 212
Table 8.3: Mean household willingness to pay 212
Table 8.4: Aggregate willingness to pay 213

Table 10.1: MNL models used for benefit transfer tests 234
Figure 10.1: Plot of Gwydir Sydney and Macquarie Marshes parameter vectors 236
Table 10.2: MNL models estimated using the combined Gwydir Sydney and Macquarie Marshes data sets 237
| Table 10.3: | Implicit prices and confidence intervals for the Gwydir Sydney and Macquarie Marshes models | 240 |
| Table 10.4: | Estimates of compensating surplus | 244 |
| Figure 10.2: | Plot of Gwydir Moree and Gwydir Sydney parameter vectors | 246 |
| Table 10.5: | MNL models estimated using the combined Gwydir Sydney and Gwydir Moree data sets | 248 |
| Table 10.6: | Implicit prices and confidence intervals for the Gwydir Sydney and Gwydir Moree models | 250 |

| Table A4.1: | Draft Macquarie Marshes questionnaire choice set format | 320 |
| Table A4.2: | A choice set with fewer attributes | 321 |
| Table A4.3: | A choice set with the current situation specified | 322 |
| Table A4.4: | Dissonance-minimising choice set format | 323 |
| Table A4.5: | The non-balanced choice set format | 324 |
| Table A4.6: | The balanced choice set format | 324 |
| Table A4.7: | Combined balanced and non-balanced format | 327 |
| Table A4.8: | Alternative 1 | 330 |
| Table A4.9: | Alternative 2 | 331 |
| Table A4.10: | Alternative 3 | 331 |
| Table A4.11: | Alternative 4 | 332 |

<p>| Table A5.1: | Mother logit models with choice set attributes | 336 |
| Table A5.2: | Mother logit models with socioeconomic and attitudinal variables | 338 |
| Table A5.3: | Mother logit models with socioeconomic, attitudinal and questionnaire variables | 340 |
| Table A5.4: | Macquarie Marshes MNL models with choice set attributes only and one block of the experimental design deleted from each model | 342 |</p>
<table>
<thead>
<tr>
<th>Table A5.5:</th>
<th>Heteroscedastic extreme value models</th>
<th>344</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table A5.6:</td>
<td>Nested logit models</td>
<td>346</td>
</tr>
</tbody>
</table>
List of Acronyms

CM – choice modelling

CVM – contingent valuation method

IIA – independence of irrelevant alternatives

IID – independently and identically distributed (error terms)

OLS – ordinary least squares

SP – stated preference
Chapter 1  Introduction

1.1  The Policy Context

In many situations there are conflicting uses for natural resources. Natural resources can be left unused and preserved. Alternatively, they can, in part or in total, be used consumptively in various developmental activities. There are many well known examples of such conflicts in Australia and elsewhere. Examples in Australia that have received attention in the popular media during the early eighties and early nineties include whether to dam the Franklin River in Tasmania to enable the production of hydro-electricity, and whether to allow uranium mining at Coronation Hill in the Northern Territory. In the case of the former, the construction of the dam would have resulted in the loss of wilderness areas due to submersion, while in the latter an open-cut mine was to be constructed in an area adjacent to the Kakadu National Park, which contains a major wetland (Saddler, Bennett, Reynolds and Smith 1980; Imber, Stevenson and Wilks 1991). Such conflicts require decision makers to make trade-offs, either explicitly or implicitly, about the preservation and developmental values of different resource uses.

Given that trade-offs between consumptive and preservation uses of natural resources are frequently faced by resource managers, how can they be made on a rational basis? Economics offers one solution to the problem. This broadly involves estimating, in dollars, the community wide benefits and costs of different resource use alternatives to determine the alternative that has the greatest net benefits. The framework for comparing different alternatives is known as cost-benefit analysis.
In cost-benefit analyses it is necessary to quantify, in dollar terms, the value of all of the costs and benefits of different resource use options. This includes both use and non-use values. Use values result from the in situ use of the resource. For example, for a proposed forestry project, use values would result from the sale of timber as well as recreation in the area. In contrast, non-use values occur when the community values a resource apart from any in situ use. For example, the community may value the continued existence of the forest if it is a unique area (Krutilla 1967). It has also been suggested that the community may value the preservation of forestry employment, especially in areas of high unemployment (Portney 1994). These non-use values are not generally traded in conventional or related markets.

The value of goods that are not traded in conventional markets is often difficult to estimate. In some cases it is possible to use revealed preference techniques to value 'non-market' goods. Revealed preference techniques, such as the travel cost and hedonic price methods, use information from related markets to impute a value for non-market goods (Måler 1974). However, revealed preference techniques can only be used in limited situations where there are existing related market data. Hence they can only be used to estimate use values and they are retrospective.

A second class of techniques that can be used to estimate non-market values are those based on the stated preferences of individuals. Stated preference techniques involve the use of surveys from which estimates are derived of the non-market benefits of different resource use alternatives. Because they rely on the use of surveys, stated preference techniques can be used in more applications than revealed preference techniques. They can be prospective and used where no related market data are available for estimating use values. They can also be used to estimate non-use values.
The most widely used stated preference technique for estimating non-market values is the contingent valuation method (CVM). The dichotomous choice version of the technique is similar to a referendum. The CVM, therefore, has the advantage of being recognised by respondents as a standard public choice instrument. However, despite its wide usage, the CVM has several limitations. It is relatively costly to use, provides limited information about people's preferences and is arguably prone to various biases (Kahneman and Knetsch 1992; Diamond and Hausman 1994; McFadden 1994). In Australia, it has become particularly controversial since its use by the Resource Assessment Commission to estimate the environmental costs of mining at Coronation Hill (RAC 1991). Similar controversy was experienced in the USA where contingent valuation was used in the Exxon-Valdez oil spill case (Bennett 1996).

Because of these limitations, there is a rationale for developing alternative stated preference techniques to estimate non-market values. Some researchers have proposed the use of conjoint analysis to estimate non-market values (eg Mackenzie 1992, 1993; Gan and Luzar 1993; Johnson and Desvousges 1997). Several different techniques fall under the heading of conjoint analysis. These include choice modelling, contingent ranking, contingent rating and paired comparison. Only limited research, however, has been undertaken on the validity of estimates generated by these techniques.

Decision makers not only require valid estimates of non-market values. They also require that such estimates can be generated cheaply and quickly. Because of their survey base, all stated preference techniques are relatively expensive to use and have a lengthy lead time. Some decision makers have responded to this problem by making use of 'benefit transfer'. Benefit transfer is the extrapolation of results from existing studies to new policy options. The existence of databases of non-market valuation studies has simplified the process of benefit transfer (eg Morrison, Groenhout and Moore 1995; De Civita, Frehs and Filion 1998), however, the question
remains whether it is reasonable to extrapolate non-market valuation estimates in this way. Non-market valuation estimates are known to be sensitive to a number of factors, and the extrapolation of results could increase existing errors.

Boyle and Bergstrom (1992) have suggested a research agenda to establish the validity of benefit transfer. One proposed line of investigation is to determine whether benefit transfer is statistically valid, what biases may be expected, and whether they can be corrected. The validity of benefit transfer of non-use values has been tested using the CVM (Bergland, Magnussen and Navrud 1995). A limitation, however, of using results from the CVM for benefit transfer is that estimates are derived only for fixed changes in the quality or quantity of a natural resource. The change in quality or quantity originally valued may be different from the change expected at a second site. Hence benefit transfer using the CVM may be subject to substantial errors. In contrast, with conjoint techniques it is possible to allow for a range of different changes when producing estimates of non-market values. Therefore it would appear that conjoint techniques may be more suited to benefit transfer than CVM. Yet only a couple of tests have been conducted of the transferability of value estimates generated using conjoint techniques, and none have focused on non-use values.

1.2 Hypotheses

Two areas are thus indicated as worthy of research. First is the validity of estimates of non-market values generated using conjoint techniques, and second is the use of conjoint techniques for benefit transfer. This thesis focuses on these two areas. A specific conjoint technique, which is known as choice modelling (CM), is used to address these areas. This thesis is structured around the testing of four hypotheses, using two case studies. The case studies involve two wetlands in New South Wales: the Macquarie
Marshes and the Gwydir Wetlands. Both of these wetlands have been degraded because of reduced, and changed timing of, instream flows resulting from the development of irrigated agriculture. Estimates of the value of changes in wetland quality were derived from three surveys. Two surveys were for the Gwydir Wetlands. The first was conducted in Moree, a rural centre close to the Gwydir Wetlands, and the second was conducted in Sydney. The third survey, for the Macquarie Marshes, was also conducted in Sydney. The four hypotheses tested in this thesis are as follows.

**Hypothesis 1: choice modelling produces valid estimates of non-use values**

The first hypothesis of this thesis is that CM produces valid estimates of non-use values. While this hypothesis has been tested for the CVM (Bennett 1981), it has not previously been tested for CM. This hypothesis is important for determining whether welfare estimates derived using CM can be used in cost-benefit analysis.

**Hypothesis 2: non-use values exist for non-environmental outcomes**

The second hypothesis in this thesis, referred to hereafter as the Portney hypothesis, is that non-use values exist for non-environmental outcomes. When the notion of non-use value was originally conceived, it was argued that it would apply to grand scenic wonders or unique ecosystems (Krutilla 1967). Most studies of non-use values have focused on estimating environmental values. However, later researchers have suggested that non-use value may be relevant for a larger range of goods. Portney (1994) argued that there may be non-market costs associated with the preservation of resource if it results in job losses which impose costs on other people in the community. This hypothesis is of importance as the exclusion of these values in many cost-benefit analyses may result in sub-optimal outcomes being recommended to decision makers.
Hypothesis 3: benefit transfer across sites is valid

The third hypothesis focuses on whether benefit transfer across sites is valid. Transfers across sites involve the extrapolation of value estimates to a different site, but for the same population. This is a common form of benefit transfer. The context for this test is the transferability of value estimates between the Macquarie Marshes survey and the Gwydir Wetlands Sydney survey.

Hypothesis 4: benefit transfer across populations is valid

The focus of the fourth hypothesis is the validity of benefit transfer across populations. This type of transfer involves extrapolating the value estimates generated from one population for a single natural resource to a second, geographically separate, population. There are several possible permutations for this test. It could involve transferring value estimates between different urban centres, or from an urban centre to a rural centre, or between rural centres. In this thesis, the two Gwydir surveys are used to test this hypothesis. Hence this is a test of the transferability of value estimates between an urban centre and a rural centre.

A test of the validity of benefit transfer across both sites and populations (i.e., Macquarie Marshes versus Gwydir Wetlands Moree) will be conducted if transfers across both sites and populations prove to be valid.

The four hypotheses that are tested in this thesis are listed in Table 1.1.
Table 1.1: Summary of hypotheses

<table>
<thead>
<tr>
<th>Hypothesis 1:</th>
<th>choice modelling produces valid estimates of non-use values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 2:</td>
<td>non-use values exist for non-environmental outcomes</td>
</tr>
<tr>
<td>Hypothesis 3:</td>
<td>benefit transfer across sites is valid</td>
</tr>
<tr>
<td>Hypothesis 4:</td>
<td>benefit transfer across populations is valid</td>
</tr>
</tbody>
</table>

1.3 Thesis Structure

The structure of the thesis is shown in Table 1.2.

Part 1 of the thesis provides a review of the literature. It includes a review of the relevant part of consumer theory, a comparison of techniques that could be used to estimate non-use values, and a review of the literature pertaining to the design of CM questionnaires and the analysis of results.

In Chapter 2 the theoretical basis for analysing the effect of resource use changes on community welfare is examined. The chapter starts by reviewing the different components of non-use value, including option value, quasi-option value and existence value. The concept of total economic value is then introduced and various measures of total economic value are considered. Issues involved in the aggregation of individual preferences are then discussed.

In Chapter 3 five alternative stated preference techniques are evaluated in terms of their theoretical capacity to produce valid estimates of non-use values. These are the CVM, CM, contingent ranking, contingent rating and paired comparison. The techniques are reviewed and compared using several criteria including the comparability of response measures across individuals, the consistency of welfare estimates with consumer theory, and the cognitive demands of the different response tasks.
In Chapter 4 methodological issues in conducting CM studies are reviewed. The purpose of this chapter is to provide a detailed review of the CM literature to gather relevant information for use later in the thesis in designing questionnaires and analysing results. Four main issues are explored. These are how to design choice sets, how to construct experimental designs for CM surveys, model selection, and model evaluation and specification tests.

The focus of Part 2 of the thesis is the CM surveys.

In Chapter 5 the two case study sites—the Macquarie Marshes and the Gwydir Wetlands—are overviewed. The irrigation industry in the Macquarie and Gwydir Valleys is described to provide a perspective on the contribution of irrigation to the economy in both valleys, and their water requirements. The impacts of consumptive water use on the two wetlands are then detailed.

In Chapter 6 the process of questionnaire development is described in order to show how the CM surveys used in the thesis were developed, and what alternative designs were considered. The strategies used in the focus groups, the problems that were identified with the draft questionnaires, and the solutions devised, are discussed. The design of the final questionnaires used in the surveys is also described in this chapter. Copies of the draft questionnaires tested in the focus groups and the final questionnaires are contained in Appendices 1-3. The results from the testing of alternative ways of designing elicitation questions for CM are reported in Appendix 4.

In Chapter 7 the results generated from the three surveys are described. The survey logistics and statistics are initially described to show how the surveys were conducted and the characteristics of the survey samples. Specification tests are not detailed in Chapter 7, because of their length, but are reported in Appendix 5. The tests of the first two hypotheses are
presented at the end of Chapter 7.

In Chapter 8 empirical issues associated with the aggregation of individual level estimates are considered. Several different approaches for adjusting mean estimates of compensating surplus to allow for divergences between sample and population characteristics are overviewed and trialed. The purpose in undertaking this chapter is to show how policy relevant results can be derived, and which factors are likely to effect results when aggregating individual estimates.

Part 3 of the thesis contains the next two chapters. This part of the thesis is centred on the tests of benefit transfer.

In Chapter 9 benefit transfer is overviewed in order to provide the context for the benefit transfer tests that are conducted in the next chapter. The chapter starts by examining the use and development of benefit transfer. Part of this development was the suggestion of a research agenda for benefit transfer by Boyle and Bergstrom (1992). The research agenda has two main lines of investigations. The first is finding the determinants of value estimates, and the second is the conducting of tests to determine when benefit transfer is valid. The final sections of this chapter are concerned with reviewing the literature on both of these topics.

The focus of Chapter 10 is the benefit transfer tests. It, therefore, contains the results of the third and fourth hypothesis tests.

Finally, in Chapter 11 conclusions are offered about the results of this thesis.
Table 1.2: The structure of the thesis

<table>
<thead>
<tr>
<th>Chapter 1: Introduction</th>
</tr>
</thead>
</table>

**Part 1 Literature Review**

- Chapter 2: Non-use Values and Economic Theory
- Chapter 3: Techniques for Estimating Non-use Values
- Chapter 4: Methodological Issues in Choice Modelling

**Part 2 The Choice Modelling Surveys**

- Chapter 5: Case Study Description
- Chapter 6: Questionnaire Design and Development
- Chapter 7: Survey Results
- Chapter 8: Aggregation

**Part 3 Benefit Transfer**

- Chapter 9: A Review of Benefit Transfer
- Chapter 10: Benefit Transfer Tests
- Chapter 11: Conclusions
Part 1 Literature Review

In the first part of this thesis the literature pertaining to the theory and estimation of non-use values is reviewed. In Chapter 2, theoretical issues relating to non-use values are examined. In Chapter 3, five stated preference techniques that can potentially be used to estimate non-use values are evaluated. Then, in Chapter 4, methodological issues associated with applying one of these techniques—choice modelling—are explored.
Chapter 2  Non-use Values and Economic Theory

2.1  Introduction

This thesis is principally concerned with the estimation of non-use values derived from changes in natural resources. As described in Chapter 1, non-use values refer to any benefits that accrue to individuals apart from the direct use of goods. The overriding goal of this thesis is to further the development of techniques that are capable of estimating non-use values, and to test whether estimates of non-use values are transferable. Therefore, it is relevant to this thesis to understand what the existing body of literature says about the nature and components of non-use value. This is the first goal of this chapter, to identify the values associated with the preservation of a natural resource. The second goal of this chapter is to define the appropriate measure for the estimation of non-use values. The third and final goal is to show how individual level estimates can be aggregated so that they can be used for policy purposes.

Various types or components of non-use values have been identified in the welfare economics literature. These include option value, quasi option value and existence value. Option value was first described by Weisbrod (1964) and reflects the value that individuals place on uncertainty. For instance, people may be willing to pay an amount in addition to their expected consumer surplus if it means that any uncertainty attached to the provision of the good is removed. A related concept, first suggested by
Krutilla (1967) and defined in detail by Arrow and Fisher (1974) and Henry (1974), is quasi-option value. Quasi-option value reflects the value that people place on obtaining further information about a good. For example, quasi-option value might exist if it is believed that more information about the benefits of preserving a natural resource will be available in the future. A third component of non-use value is existence value. Existence value, which was also defined by Krutilla (1967), reflects individuals' willingness to pay to preserve a resource apart from any in situ use. Existence values typically result from bequest, altruistic, stewardship and self-seeking motives. Existence value is generally conceived of being independent of either option or quasi-option value. These three components of non-use values are explored in detail in Sections 2.2, 2.3 and 2.4. What theory says about the nature of these values, and implications for the estimation of non-use values, are noted in these sections. Of particular interest is the type of goods for which there might be existence value.

While it may be possible to estimate separately use, option, quasi-option and existence values through careful wording in a questionnaire, it is more often the case that individuals make holistic judgements about the value of preserving a natural resource. As Mitchell and Carson (1987, p.83) have suggested:

Our view of respondent behaviour in the CV setting is that when people are asked to value an amenity...they do so by making a holistic judgement. Instead of going through a mental process where they separately value each of the relevant benefit categories (such as use and existence values) before combining them in their minds to arrive at total value, respondents arrive at a global judgement about what the amenity is worth to them...

Consistent with this view of respondent behaviour is the concept of total economic value. Total economic value is the overall value that an individual has for preserving a natural resource (Randall and Stoll 1983).
It includes use and all non-use values. This concept is described in Section 2.5, and is considered to be the most relevant measure for this thesis given how respondents determine value.

Total economic value can be measured by determining ‘net willingness to pay’. There are, however, several different measures of net willingness to pay, including Marshallian consumer surplus and several Hicksian surpluses. The theoretically correct measure of total economic value is shown, in Section 2.6, to be either compensating or equivalent (Hicksian) surplus, depending on the context of the valuation exercise.

Estimates of compensating (or equivalent) surplus are calculated at the individual level. After they have been estimated for individuals, these estimates then have to be aggregated to determine the net benefits of different resource use alternatives for the relevant community. Once aggregate estimates have been derived, information is available to decision makers about which alternative will provide the greatest net benefit to the community. However, aggregation is complicated because of the difficulties associated with interpersonal comparisons of utility. Methods for aggregating individual benefits and costs, and their implications for model estimation, are described in Section 2.7.

Conclusions about what economic theory says about the estimation of non-use values are presented in Section 2.8. Thus, with the values associated with preservation identified, the appropriate unit of measurement defined, and a process for incorporating the results into policy established, the context of the thesis will be set in place.

2.2 Option Value

A frequently cited component of non-use value is option value (eg Pearce and Markandya 1989). Option value reflects the effect of uncertainty
about the supply of, or demand for, a natural resource on people’s willingness to pay for a natural resource. It is therefore expected to be a part of the value estimates derived in Part 2 of this thesis.

The concept of option value has stimulated considerable debate in the economics literature, with various theoretical (eg Long 1967; Cichetti and Freeman 1971; Schmalensee 1972; Bishop 1982; Freeman 1984, 1985; Smith 1984; Plummer 1985, 1986; Plummer and Hartman 1987) and applied works (Greenley, Walsh and Young 1982; Brookshire, Eubanks and Randall 1983; Smith, Desvousges and Fisher 1983) providing evidence about the existence of option value and its sign. Option value is an important concept, as its existence implies that consumer surplus may be an inaccurate measure of value unless the effects of uncertainty are considered. However, the debate about option value is complex, and neither the sign or magnitude of option value is clear from theory, except in a few cases.

### 2.2.1 Demand side option value

Option value was introduced by Weisbrod (1964), who was responding to the argument that a free-market allocation would provide an optimal outcome for forestry resources (Friedman 1962). Weisbrod (1964) argued that people may be willing to purchase an option to consume a good in the future where demand is uncertain, and where there are no close substitutes for the good. He used as an example visits to a national park such as Sequoia in the USA, which is irreplaceable and has no close substitutes. If the present value of future operating costs is not exceeded by the present value of future revenue from entrance fees, Weisbrod (1964) suggested that a profit-maximising entrepreneur would cease operating and use the park for extractive purposes. However, from a social perspective this could be inefficient, because people who anticipate visiting the park in the future, but will never actually do so, may be willing to pay something to
maintain this option. In this context, option value is seen as a risk aversion premium. It is the amount in addition to consumer surplus that people would be willing to pay to ensure future access to a good.

Cicchetti and Freeman (1971) and Schmalensee (1972) further developed Weisbrod’s (1964) notion of option value. Cicchetti and Freeman (1971) attempted to determine whether it was possible \textit{a priori} to know the sign of option value. Cicchetti and Freeman (1971) showed that option value resulting from demand uncertainty will be positive. Schmalensee (1972), however, reached a different conclusion: that the sign of option value resulting from demand uncertainty was indeterminate (see also Graham 1981). Anderson (1981) showed that these contradictory conclusions resulted from two assumptions about the structure of consumers’ preferences made by Cicchetti and Freeman (1971). Bishop (1982) argued that these two assumptions violate consumer theory, and the correct result was reached by Schmalensee (1972): that is, the sign of option value resulting from demand side uncertainty is indeterminate\(^1\).

\section{Supply-side option value}

Bishop (1982) explored the effect of supply-side uncertainty on option value, given that there is no demand side uncertainty. Supply-side uncertainty occurs when the future provision of a good is uncertain. Previously, the discussion had centred on demand side uncertainty, where consumers’ demands were uncertain. Bishop (1982) showed that in the

\footnote{Freeman (1984) and Plummer and Hartman (1987) explored several other issues related to demand uncertainty. Freeman (1984) examined uncertainty about income and prices of substitute or complementary goods, and found that uncertainty about income will be negative for risk averse individuals. The sign for uncertainty about the prices of substitutes or complements was found to be ambiguous. Plummer and Hartman (1987) explored the effect of uncertainty due to income and quality. Quality is defined as any non-monetary attribute that affects the valuation of an activity (e.g., amount of congestion, rainfall). Similar to Freeman (1984), Plummer and Hartman (1987) showed that option value is negative for income uncertainty, but for uncertainty about quality, Plummer and Hartman (1987) showed that option value is positive.}
case where demand is certain and supply-side uncertainty could be completely eradicated through a new project, then option value will be positive for risk averse households.

Freeman (1985) and Plummer (1986), however, showed that Bishop (1982) examined only one of the four possible scenarios for supply-side uncertainty, as shown below:

Case A:  No project - no supply  Case B:  No project - possible supply
         With project - sure supply       With project - sure supply

Case C:  No project - no supply  Case D:  No project - possible supply
         With project - possible supply  With project - possible supply

Note that the case analysed by Bishop (1982) is represented by Case B. Freeman (1985) confirmed Bishop’s (1982) result for Case B, but in addition showed that in Case A option value is zero, in Case C option value is indeterminate, and in Case D option value is negative. Plummer (1986) reached similar conclusions for Cases A and B, but found that it is possible to determine the sign of option value in Case C, and Case D only in some circumstances. Hence the sign of option value can be determined in three of the four cases involving supply uncertainty. However, if demand uncertainty occurs in addition to supply uncertainty, then the sign of option is indeterminate (Freeman 1985).

A question that may reasonably be asked is whether option value is simply a theoretical curiosity given the limited number of cases where it is possible to know \textit{a priori} the sign of option value. Moreover, it may not be necessary to derive separate estimates of option value. An estimate of option price, which is the sum of consumer surplus (in the absence of uncertainty) and option value, would be adequate in most cases. This point is made by Bishop (1982, p.14):
On an empirical level, the primary contribution of literature reviewed here may not be option value, but option price. Few would object to the assertion that if it could be measured, option price rather than projected consumer surplus is the correct measure of consumer welfare in cases involving considerable uncertainty. Furthermore option price would appear to be much more amenable to measurement than option value taken alone.

Freeman (1993) also argues that attempts to separately estimate option value and add it to consumer surplus are inappropriate. He contends that option value should not be considered a separate component of non-use value, because it simply reflects the difference between an *ex ante* perspective and an *ex post* perspective when the effects of uncertainty are eliminated (see also Smith 1987).

In conclusion, the main implication from the option value literature, given the ambiguity of theory regarding the *a priori* sign of option value, and questions about the relevance of deriving separate estimates of option value, is that option price should be estimated when valuing natural resources. The main practical implication of this is that if there is any uncertainty regarding the provision of the good of interest, this should be detailed in the questionnaires. This is the approach followed in this thesis.

### 2.3 Quasi-Option Value

While the notion of option value was being developed, a similar notion termed quasi-option value was suggested\(^2\). This concept was first identified by Krutilla (1967) who suggested, in the context of preserving certain natural occurring plant species for agricultural production or medicine, that there is also value in:

\(^2\) There has been some confusion about the relationship between option and quasi-option value. Henry (1974) suggested that quasi-option value is the same as option value. However other researchers, while suggesting that supply-side option value and quasi-option value could be related, believe that the two concepts are fundamentally different (eg Bishop 1982).
preserving the opportunity to examine all species among the natural biota for this purpose (p.779)

This concept of value was more fully examined by Arrow and Fisher (1974). Arrow and Fisher (1974) showed that positive quasi-option value occurs because net benefits are lower when there is uncertainty about future benefits. This is because of the loss of options due to irreversible development. Arrow and Fisher (1994) questioned whether uncertainty about the costs or benefits of a project that would have an irreversible effect on the natural environment would effect the magnitude of benefits and costs beyond their expectations. They concluded that:

If the development involves some irreversible transformation of the environment...and if information about the costs and benefits of both alternatives realised in one period results in a change in their expected values for the next, the answer is yes—net benefits from developing the area are reduced and, broadly speaking, less of the area should be developed (p.313).

This reduction in net benefits due to uncertainty about future benefits of preservation is equal to quasi-option value. Arrow and Fisher (1974) suggested that the implication of quasi-option value is that in situations involving both uncertainty and irreversibility, decision makers should err on the side of under investment in development. Further information can then be gained about the benefits and costs of development.

Conrad (1980) further clarified the concept of quasi-option value. He showed, using an intergenerational growth model, that it is equivalent to the expected value of information. In contrast to Arrow and Fisher (1974),

3 A simplified and easily understandable model of quasi-option value is provided in Kerr and Sharp (1985).
4 It should be recognised that any decision not to develop to gain further information implies an opportunity cost in terms of foregone development. This doesn’t affect the sign of quasi-option value, which is always positive (or equal to zero) as further information will always have a non-negative value. However, it may be the case that the net benefits associated with waiting could be negative if deferring a development decision implies a loss of production and it turns out that extra information has limited value.
he showed that 'the value of lost options is an expected value based on what one might learn' (p.818). In other words, quasi-option values result because in later periods more information is gained about the benefits of preservation, and hence more efficient outcomes can be achieved. Conrad (1980) suggests that one of the main uses of quasi-option value is not as part of a cost-benefit analysis, but rather as a complement to traditional evaluation techniques. For example, the concept of quasi-option value could be used to assess how future developments in information or technology could alter the desired level of development. In this sense it is similar to conducting a sensitivity analysis where information or technology is the critical variable.

Quasi option value is a concept that is not particularly amenable to a priori estimation and hence inclusion in cost-benefit analyses. The main implication of the literature appears to be that in situations involving irreversibility, analysts should recognise that a level of development less than what is suggested by a cost-benefit analysis may be appropriate. The extent that the level of development should be reduced, though, cannot be exactly calculated a priori, but can be conceived of depending on expectations about the possibilities of gathering further information. Kerr and Sharp (1985, p.95) summed up the case for the use of quasi-option value as follows:

While quasi-option value is definable, it is useless for decision making. Individual values are not identifiable because individuals possessing quasi-option values are unaware of them...The implication of the concept of quasi-option value is that the conventional cost-benefit analysis approach to decision making of using expected values of welfare in future periods may be inappropriate in cases of irreversibility where there is prospect of new information becoming available...This seems to imply that...present development should be reduced from the optimal levels calculated using the traditional approach.
While the usefulness of quasi-option value is certainly limited, it is not clear that it is of no use for decision making. While respondents cannot accurately know the magnitude of quasi-option value without knowing what the future holds, it is possible that respondents may realise that further information about the benefits of preserving a natural resource may become available in the future. Respondents may make a subjective assessment of the magnitude of quasi-option value and be willing to pay an amount in addition to expected consumer surplus to preserve a natural resource. Hence it may be reasonable to expect that part of respondents’ willingness to pay to preserve a natural resource is a subjective assessment of quasi-option value, even if their assessment proves to be incorrect in the future.

2.4 Existence value

The third main component of non-use value is existence value. Existence value occurs when people value the preservation of a natural resource where this does not involve any in situ use of the good. Krutilla (1967, p.781) first noted existence values, suggesting that:

When the existence of a grand scenic wonder or a unique and fragile ecosystem is involved, its preservation and continued availability is a significant part of the real income of many individuals.

Put another way, Krutilla (1967, p.781) wrote:

There are many persons who obtain satisfaction from mere knowledge that part of wilderness North America remains even though they would be appalled by the prospect of being exposed to it.

McConnell (1983) provided a more formal definition of when existence value occurs. He defined existence value using a utility function \( U(x, R) \), where \( R \) is the resource for which there is existence demand and \( x \) is a
vector of goods. Existence demand occurs when $U(x, R)$ is weakly separable in $x$ and $R$ and $\partial U/\partial R > 0$. In other words, existence demand occurs when an increase in the resource results in an increase in utility apart from any use of the resource.

Following McConnell (1983), the compensating surplus\(^5\) associated with existence value can be calculated using expenditure functions as shown in equation (2.1). The price is set at a level that ensures there is no use of a resource (the choke price\(^6\)). Hence, any remaining value can be interpreted as being existence value\(^7\).

\[
CS = e(p^*(R'), R', U) - e(p^*(R^0), R^0, U^0)
\]

where $CS$ is compensating surplus, $p^*$ is a choke price that sets the use of $R$ to zero, $U^0$ represents the initial level of utility, $R^0$ is the initial level of the resource and $R'$ is the subsequent level of the resource.

In the definition of existence value given earlier, Krutilla (1967) suggested that uniqueness and irreversibility are essential to existence value. Randall and Stoll (1983), however, suggested that existence value is relevant to a large range of goods. Randall and Stoll (1983, p.268) contended that:

Thus, even commonplace artefacts of human civilisation (eg drink cans) may have existence value, although the circumstances which would make it large are unlikely. Empirically significant existence values are not confined to natural objects; we believe they occur for human artefacts and cultural manifestations, from historic buildings to grand opera. Nor are existence values confined to 'the last few ___s on earth.' We would expect to find positive existence values for local amenities, local subpopulations of flora and fauna, and for local cultural amenities.

---

\(^5\) Compensating surplus is defined in Section 2.6.  
\(^6\) This assumes that it is possible to put a price on the use of the resource, which in many cases is a reasonable assumption.  
\(^7\) Freeman (1993) summarises five different ways of defining existence value given different levels of the resource, and different market prices and hence use values. The example provided here is Freeman's (1993) Case 1 involving 'simple non-use values'. 

22
Randall and Stoll (1983) utilised demand and supply analysis to demonstrate the wider relevance of existence values. They illustrated how existence values are affected by supply and demand using Figure 2.1. Shown in the figure are demand and supply curves for cattle and condors. The marginal existence value for condors is much higher than for cattle because of restricted supply. This figure shows that: (1) marginal existence value will be smaller if goods have many close substitutes; (2) for resources in large supply marginal existence value will be small; and (3) if a project will only cause a small change in supply the change in existence value will also be small (depending on the elasticities of the two curves).

**Figure 2.1: Existence value and relative scarcity**

![Figure 2.1: Existence value and relative scarcity](image)

*Source: Randall and Stoll (1983)*

More recently, Portney (1994, p.13)\(^8\) has suggested that it would also be possible to estimate existence values for non-market costs resulting from

\(^8\) Earlier references to the estimation of existence values for non-environmental costs in the 'grey' literature include Brunton (1991) and Carson (1991).
environmental regulations$^9$:

If I derive some utility from the mere existence of certain natural environments I never intend to see (which I do), might I not also derive some satisfaction from knowing that refineries provide well-paying jobs for hard-working people, even though neither I nor anyone I know will ever have such a job? I believe I do. Thus, any policy change that “destroys” those jobs imposes a cost on me...Since regulatory programs will always impose costs on someone—taking the form of higher prices, job losses, or reduced shareholder earnings—lost existence values may figure every bit as prominently on the cost side of the analytic ledger as the benefit side.

Boyle and Bishop (1987) also made the point that existence value need not be positive (ie $\partial U/\partial R < 0$). This would occur if people preferred less of a natural resource$^{10}$.

2.4.1 Existence value and motives

A relevant question is what are the motives behind existence value, and do these motives affect existence value? McConnell (1983) concluded that ‘the motive is immaterial’ (p.257), that because it is not possible to differentiate between, say, pure existence demand and bequest demand (ie for the sake of one’s children), one should not be concerned with motives.

---

$^9$ Some evidence for the existence of non-market values for production has been provided by Lockwood, Loomis and De Lacy (1994). They found, using separate CVM surveys, that willingness to pay for production was substantially smaller than willingness to pay to preserve a natural resource. Adamowicz, Boxall, Williams and Louviere (1998) also attempted to determine whether non-use values exist for employment by including an employment attribute in a choice model focusing on willingness to pay for caribou habitat enhancement. However, the coefficient for the employment attribute was insignificant in each model specification trialed, indicating that non-use values for employment were equal to zero. Adamowicz et al (1998) suggested that this finding may have resulted because (1) the survey was conducted a long way from where the job impacts were to occur, (2) the employment impacts were minor, or (3) respondents were told that those who lost their jobs would be retrained.

$^{10}$ There is some empirical evidence of this phenomenon. Swallow, Weaver, Opaluch and Michelmans (1994) found that certain people preferred less wetland area to more, while Stevens, Echeverria, Glass, Hager and More (1991) found that people were willing to pay for a smaller coyote population (due to danger).
However, Madariaga and McConnell (1987, p.938) concluded that motives can matter:

Understanding the motives that underlie contingent value responses can help design and interpret contingent valuation experiments. Further, understanding motives may alter the role of existence value in benefit-cost analysis.

Brookshire and Smith (1987, p.932) also commented that:

the reason for concern over motives for nonuse values...[is that it] provide[s] a link explaining why these values arise outside the conventional concepts of consumption.

As Brookshire and Smith (1987) have suggested, motives are relevant because they explain why existence values occur for natural resources, even if they are not traded in conventional markets. They may also be relevant for predicting the magnitude of existence value, and whether existence value will conform with utility theory, as shown in the following discussion.

Randall and Stoll (1983) suggested that there are three main motives for existence value. These are interpersonal altruism (also known as vicarious value) where existence value results from knowing that a resource is available for others to use; intergenerational altruism (also known as bequest value) which results from knowing a resource will be available for future generations; and Q-altruism (also known as intrinsic value) which results from knowing that a resource is undisturbed, and is essentially a self focused motivation.

Madariaga and McConnell (1987) examined the effect of different altruistic motives on existence values. They defined two types of altruism: individualistic altruism where people gain from the well being of others, without regard to the manner in which the gains were achieved, and paternalistic altruism where people gain value when others get to use a
specific resource. To distinguish between these two types of altruism consider the following utility functions, the first of which shows how the utility of person A is affected by individualistic altruism, and the second shows the effect of paternalistic altruism:

\[ U_A = U_A(Y_A, U_B(Y_B, R)) \] (2.2)

\[ U_A = U_A(Y_A, U_B(R)) \] (2.3)

where \( U_A \) and \( U_B \) are the utility of persons A and B respectively, \( Y_A \) and \( Y_B \) are the income of persons A and B, and \( R \) is the amount of the natural resource.

Madariaga and McConnell (1987) showed that these two different motives can imply quite different existence values. The effect of paternalistic motives is straightforward: an increase in the quantity of the resource implies positive existence values. The effect of individualistic altruism is, however, potentially different because it is based on the welfare of all of the arguments in individual B’s utility function. If an increase in \( R \) also implies a decrease in \( Y_B \) because of, say, the introduction of a levy, then there are potentially two offsetting effects. The increase in the resource implies an increase in welfare, but the decrease in \( Y_B \) implies a decrease in welfare. Given these opposing effects, existence value resulting from individualistic altruism could potentially be positive, zero or negative\(^{11}\).

Loomis (1987a) argued that if the motive for existence value is self interest, such as with Q-altruism, existence value could be much larger than if it is based on altruistic motives ‘owing to the dominant role that self-interest plays in a person’s valuation of goods he consumers versus goods he provides for others to consume’ (p.24). In other words, existence value would be larger if a person’s utility function could be described as \( U^A = U^A(Y_A, R) \), rather than equation 2.3.

\(^{11}\) Madariaga and McConnell (1987) and Johannesson, Johansson and O’Connor (1996) provided empirical evidence that shows that individualistic as well as paternalistic motives exist.
A fourth motive cited by Brookshire, Eubanks and Sorg (1986) is stewardship. Brookshire et al (1986) commented that stewardship refers to the management of another’s property, finances etc. Hence the person with stewardship motives could have a utility function similar to equations (2.2) or (2.3). However, it is possible that stewardship motives may result if a person feels that by preserving a resource they are preserving something for which they have been entrusted in some collective sense. In this case, other peoples’ utility functions need not figure in the utility function of a person with stewardship motives.

If a person has stewardship motives as just described, they could be considered to have motives similar to what was described by Sen (1977) as commitment. Commitment, according to Sen (1977), is most important with public goods. Sen (1977) made a distinction between sympathy and commitment. Sympathy is similar to the present definition of altruism: it ‘corresponds to the case in which concern for others directly affects one’s own welfare’ (p.326). Commitment, however, is not necessarily altruistic, in that person B’s utility function need not figure as an argument in person A’s utility function. Further, if a person is motivated by commitment, an increase in the size of a resource need not result in increased utility, even though a person may be willing to pay something for the increase. As stated by Sen (1977, p.327):

One way of defining commitment is in terms of a person choosing an act that he believes will yield a lower level of personal welfare to him than an alternative that is also available to him.

It is, however, difficult to conceive of a situation where a person’s commitment to a cause does not provide any quid pro quo benefit. It may be that an increase in a good does not provide any extra utility. But, because of their commitment, a person may suffer disutility or a sense of guilt if the size of a resource declines. Alternatively, a person with
commitment motivation may derive utility from knowing that what they have done is right.

A possible consequence of commitment motivations is that it may cause lexicographic or non-compensatory preferences, whereby a person is unable to make trade-offs and Hicksian welfare measures are undefined (Edwards 1986, 1992; Opaluch and Segerson 1989, Lockwood 1998). Given situations where commitment motivations are pervasive, people may appear not to make trade-offs if the increase in utility from knowing that they are doing what is right exceeds the disutility from lost income. This may be a significant problem when valuing the existence value attached to wildlife if ethical or moral principles are invoked (Stevens et al 1991). The extent that lexicographic preferences are genuinely evident is, however, yet to be clarified\textsuperscript{12}.

In summary, this discussion of the motives behind existence value has shown that motives can affect the magnitude of existence value. These motives explain why respondents may have existence values, and that there is reason to believe that existence value does occur and should be estimated as part of non-use value. The other main implication from this discussion of motives is that, like option value, there may be situations where existence value appears to be low or is difficult to estimate.

2.5 \textit{Total Economic Value}

In previous sections of this chapter, the different components of non-use value have been considered, including option value, quasi-option value and existence value. The question of what is the aggregate or total economic

value associated with preserving a natural resource is now addressed.

Total economic value was first defined by Randall and Stoll (1983) (see also Pearce and Markandya 1989). Broadly there are two main types of benefits that are part of total economic value. The first reflects user benefits. User benefits include use value, which reflects both consumptive and non-consumptive use of a natural resource. As discussed under option value in Section 2.2, the effects of uncertainty should be admitted when estimating user benefits, if they are relevant. And, as discussed in Section 2.3, it is also possible that quasi-option value could be a part of user benefits if respondents make a subjective assessment of the benefits of obtaining information in the future. The sum of use value, option value and quasi-option value is equal to user benefits.

The second main component of total economic value is existence value, which reflects peoples' willingness to pay to preserve a natural resource apart from any in situ use of the resource. The sum of user benefits and existence value is equal to total economic value. Therefore, total economic value can be expressed as follows:

\[
\text{Total economic value} = \text{user value} + \text{existence value}
\]

As discussed in the introduction, it is likely that respondents form an assessment of the overall value that they have for a change in the quality or quantity of a natural resource (Mitchell and Carson 1987). It is unlikely that respondents explicitly calculate their use, option, quasi-option and existence values and add them to determine their aggregate willingness to pay. Hence it would appear to be appropriate in stated preference studies to attempt to estimate total economic value, rather than the magnitude of its components. This is likely to have the advantage of being simpler for respondents as well as avoiding the possibility of double counting of
values\textsuperscript{13}.

This concept of economic value is used in this thesis when deriving estimates of the benefits associated with preserving or improving the quality or quantity of a natural resource\textsuperscript{14}. In the next section, different measures of total economic value are considered.

\textbf{2.6 Measures of Consumer Surplus}

In welfare economics, the total economic value of a good is indicated by its consumer surplus. Consumer surplus, which was first described by Dupuit (1844), is equal to the difference between the price actually paid and the price that a consumer would have been willing to pay\textsuperscript{15}. Marshall (1920) developed Dupuit’s concept of consumer surplus for the case where consumer goods are perfectly divisible, and it has since become known as Marshallian consumer surplus. Marshallian consumer surplus is equal to the area under the demand curve but above price.

The Marshallian measure of consumer surplus, however, has several limitations. One limitation is that it is not always possible to obtain unique estimates of consumer surplus. When there are multiple price and income changes, the ordering of these changes can affect the magnitude and sign

\textsuperscript{13} It may still be possible to disaggregate total value estimates into component parts after respondents have provided an aggregate value estimate (eg Walsh, Loomis and Gillman 1984). However, the warnings of Freeman (1993) regarding the problems with this approach should be noted.

\textsuperscript{14} While most attention is placed on deriving estimates of total economic value, in some cases it is possible to get indications of the magnitude of some of the components of total economic value. This does not require respondents to separately estimate the different components of total economic value, as was criticised by Mitchell and Carson (1987). Rather, it involves categorising respondents given that they are the only group likely to have certain values. For instance, the higher value that people who are planning one day to visit a site have provides an indirect indication of the extent of use and option value (see Chapter 7).

\textsuperscript{15} Note that there is a similar surplus for producers (known as producer surplus) described by Marshall (1920). It is equal to the difference between the price paid and marginal cost.
of consumer surplus (Silberberg 1972; Just, Hueth and Schmitz 1982). A second limitation results from the effect of price changes on real income. Unless the marginal utility of income is constant with respect to prices and income, changes in consumer surplus may not correspond to changes in utility (Just et al 1982, Hanley and Spash 1993). Thirdly, it is generally not possible to estimate Marshallian demand curves, and hence Marshallian consumer surplus, using the results from a stated preference application.

Given the deficiencies in the Marshallian concept of consumer surplus, other welfare measures have been proposed. One alternative set of measures, in which the marginal utility of income is held constant, was proposed by Hicks (1940/41, 1943). Unlike Marshallian consumer surplus, the Hicksian measures are path independent by construction (Just et al 1982).

Hicks proposed two measures known as compensating and equivalent surplus for use when quantity is constrained. This is generally the case when valuing changes in natural resources which are public goods as the changes are usually discrete and consumers are not free to vary their level of consumption. Broadly, the two measures differ according to whether welfare is held at the initial or subsequent level. Welfare is held at the initial level for the compensating measure. This is appropriate when consumers have implicit or explicit property rights to the initial situation.

---

16 There are, however, *path independency* conditions that ensure the existence of a unique estimate of consumer surplus. A money measure of consumer surplus is path independent if utility functions are homothetic (Silberberg 1972). Homothetic utility functions imply that income effects for all goods are zero (ie $\partial x_i/\partial y = 0$ for all i, j where $x_i$ represents the quantity demanded of good i and y is income). Silberberg (1972) showed that this property results if the cross-price elasticity of demand for all goods whose price changes are equal.

17 Note that the cross-price elasticity of demand for all goods whose price changes is equal for all curvilinear indifference curves; hence Hicksian measures are path independent.

18 Hicks (1943) actually used the terms 'quantity-compensating variation' and 'quantity-equivalent variation'.

19 Hicks also proposed two measures for when quantity is not constrained. These are known as compensating and equivalent variation. For parsimony these are not reviewed here. See, for example, Hanley and Spash (1993).
Compensating surplus is the change in income that will make an individual indifferent between the original situation and the subsequent situation given the new quantity. Compensating surplus is equal to willingness to pay for an improvement in environmental quality if the individual does not have property rights for the improved level of quality. Compensating surplus is shown in Figure 2.2. On the horizontal axis is the good in question (X) and on the vertical axis is income (Y). There are two indifference curves (IC₁ and IC₂) that show combinations of X and Y between which there is indifference. Higher indifference curves signify higher levels of welfare. There is a change in the quantity of the good from Q₁ to Q₂, such that the individual moves from point A on IC₁ to point B on IC₂. However, as the individual does not have the right to be at the subsequent level of welfare, income is taken away from them so that he/she moves to point C. The vertical distance BC is the compensating surplus.

Equivalent surplus is the change in income that is required at the initial quantity to make an individual as well of under the new quantity. The measure is equal to the consumer’s willingness to accept for the improvement not to occur, and is relevant from a policy perspective when the consumer has the property rights for the improvement. It is shown by distance AD in Figure 2.2. Budget constraints are not shown in Figure 2.2 as, in most cases, natural resources are public goods and, hence, price is irrelevant.
Compensating and equivalent surplus can also be defined using expenditure functions. Compensation surplus is the difference in expenditure (or income) required to maintain utility at the initial level \( (U_1 - \text{associated with } IC_1) \) given the new quantity (note that prices for marketed goods are assumed not to have changed):

\[
CS = E(P_1, Q_1, U_1) - E(P_1, Q_2, U_1)
\]  \hspace{1cm} (2.4)

where \( CS \) is compensating surplus, \( P_1 \) is the initial price level, \( Q_1 \) is the initial quantity, and \( U_1 \) is the initial level of utility.

Equivalent surplus is the difference in expenditure (or income) required to maintain utility at the subsequent level \( (U_2) \) given the new quantity:

\[
ES = E(P_1, Q_1, U_2) - E(P_1, Q_2, U_2)
\]  \hspace{1cm} (2.5)

where \( ES \) is equivalent surplus, \( P_1 \) is the initial price level, \( Q_1 \) is the initial quantity, and \( U_2 \) is the subsequent level of utility.

Randall and Stoll (1980)\textsuperscript{20} examined the relationship between

---

compensating, equivalent and Marshallian consumer surplus. They found that, for indivisible or lumpy goods with a normal income elasticity of demand, and where only a small proportion of the total budget is spent on the good, the Hicksian surplus measures are closely approximated by Marshallian consumer surplus. However, in evaluations of lumpy goods the measures are not likely to be equal if the change is large, the good is highly valued, and the income elasticity is high. Hanemann (1991) further developed Randall and Stoll’s (1980) argument, by demonstrating the importance of a variable that previously been taken to represent the income elasticity. He showed that the parameter epsilon (ξ), which represented income elasticity, was actually equal to the ratio of the income elasticity and the elasticity of substitution between the public good being valued and private goods. For public goods with few substitutes and where income elasticity is high, Hanemann (1991) showed that compensating and equivalent surplus could be substantially disparate\textsuperscript{21}. Hence estimates of willingness to pay for an increase in a natural resource and willingness to accept compensation to allow reductions will potentially be quite different\textsuperscript{22}. The implication of this result is that practitioners need to be concerned about property rights when estimating the value of natural resources, and willingness to pay should not be considered a reasonable approximation of willingness to accept, despite recommendations made in the literature (Brookshire, Cummings and Schulze 1986; Arrow, Solow, Portney, Leamer, Radner and Schuman 1993).

In sum, the Hicksian surpluses are the most appropriate measures of total economic value, and will therefore be utilised in this thesis. Secondly, it is

\textsuperscript{21} Mitchell and Carson (1989) demonstrated the importance of this result. In the case of the valuation of non-market goods, they showed that Randall and Stoll’s equations suggest that there will only be minor differences in compensating and equivalent surplus, whereas Hanemann’s modification means that the differences should be large.
generally not appropriate to use compensating surplus as an approximation of equivalent surplus, or Marshallian consumer surplus as an approximation of either of the Hicksian surpluses in the context of valuing natural resources except in a few cases. Therefore when estimating values in this thesis, it will be necessary to ensure that the measure used to estimate total economic value conforms to the property rights regime in the case studies.

2.7 Aggregation

Thus far this review has focused on individuals. However, individual estimates of total economic value are of limited use to decision makers. Decision makers generally require aggregate estimates. Therefore the consideration of issues associated with aggregation are important from the perspective of how the results of this thesis might be used. Issues associated with aggregation are, however, important for another reason. As will be discussed later in this section, difficulties associated with aggregation imply that individual values need to be estimated using the assumption of constant marginal utility of income. This has implications for the specification of models in Part 2 and 3 of this thesis. In this section theoretical issues associated with aggregation are reviewed; empirical issues associated with aggregation are considered in Chapter 8.

The simple addition of individual value estimates was criticised by Robbins (1937) who argued that interpersonal comparisons of utility were not possible. This was because of the differential effect of changes in environmental quality. Other reasons have been given for this, including that willingness to accept is not bounded by a budget constraint (Randall and Stoll 1980); respondents reject the willingness to accept property right (Mitchell and Carson 1989); respondents are cautious and, because of unfamiliarity, tend to overstate willingness to accept (Coursey, Hovis and Schulze 1987); and that respondents value gains and losses differently (Peterson, Brown, McCollum, Bell, Birjulin and Clarke 1996).

---

22 Many non-market valuation studies have found that willingness to accept (equivalent surplus) tends to exceed willingness to pay (compensating surplus) for improvements in environmental quality. Other reasons have been given for this, including that willingness to accept is not bounded by a budget constraint (Randall and Stoll 1980); respondents reject the willingness to accept property right (Mitchell and Carson 1989); respondents are cautious and, because of unfamiliarity, tend to overstate willingness to accept (Coursey, Hovis and Schulze 1987); and that respondents value gains and losses differently (Peterson, Brown, McCollum, Bell, Birjulin and Clarke 1996).
income on people’s utility depending on their income level. Robbins (1937) contended that the effect of a loss in income on the utility of people negatively affected by a proposal may be greater than the increase in utility from a similar increase in income for beneficiaries of the proposal. Hence whenever losses were being experienced, one could not be sure that that there was not going to be an overall decrease in utility. Following Robbins’ (1937) criticisms, several approaches have been suggested for deriving aggregate estimates of value.

Hicks (1939) and Kaldor (1939) argued that the problems associated with aggregating cardinal value estimates could be overcome if those who gain could compensate those who lose. Kaldor (1939) argued that a policy can be seen to be welfare enhancing if the policy change increases net benefits so that it is possible to compensate the losers and still provide extra benefits to the wider community. Because of the compensation of the losers from a policy change, it is therefore possible to be certain that a policy change is welfare improving despite interpersonal comparisons of utility\textsuperscript{23}. Hicks (1939) argued that in certain cases (eg perfect competition) it would be possible to use this principle not only to determine whether a policy change was welfare enhancing, but which policy change would be ‘optimal’; in other words, providing the greatest benefits for society. However, the Kaldor-Hicks potential compensation principle is not problem free. This principle was criticised by Scitovsky (1941/42) who showed that, in certain circumstances, two different situations can be shown each to be preferable to the other. This may occur where a policy change results in an increase in the production of one good, but a decrease in the production of another good. This result has become known as the ‘reversal paradox’. Little (1957) has also criticised the potential compensation principle because it does not directly allow for the effects that a project can have on the income distribution. Little (1957) suggested that a two fold criterion be used: first projects be assessed to determine
whether the gainers can compensate the losers, and second projects be assessed to determine whether the distribution of income is improved.

Attempts have also been made to aggregate ordinal rather than cardinal preferences. However, this approach has not met with much success. Arrow (1951) proposed four criteria that are desirable for a social choice rule using ordinal preferences. These are that there is an unrestricted domain, that preferences are independent of irrelevant alternatives, that the weak Pareto principle is satisfied, and that there is no dictatorship (Myles 1995). Arrow (1951) showed, in his impossibility theorem, that there is no social welfare function that satisfies these four criteria.

While it is not possible to aggregate ordinal preferences, aggregation is possible when further information is available on individuals’ utilities (Deaton and Muellbauer 1983, Myles 1995). For example, aggregation can be achieved by allowing cardinal unit comparability. Cardinal unit comparability assumes that changes in welfare can be measured in comparable units, but it is not possible to determine the level of each individuals’ utility. It possible to derive a utilitarian social welfare function by assuming cardinal unit comparability and slightly modifying two other assumptions (Deaton and Muellbauer 1983).

Despite the criticisms of Robbins (1937), cardinal full comparability is often assumed in practice when deriving social welfare functions. Cardinal full comparability represents the maximum level of comparability and allows comparisons of both utility levels and changes in welfare (Myers 1995). Johansson (1987) demonstrated the conditions under which a social welfare function has welfare significance given cardinal full comparability. Johansson (1987) utilised a Bergson-Samuelson social welfare function for an \( H \)-individual society, as follows:

\(^{23}\) This is a critical point. Even if projects are 'potentially' welfare improving, they may not result in a welfare improvement unless compensation is paid, depending on the
\[ W = W(U^1(x^1), \ldots, U^H(x^H)) \]  \hspace{1cm} (2.6)

where \( x \) is a vector of market goods, \( w \) is societal welfare, and \( U^h \) represents the utility of individual \( h, h=1, \ldots, H \).

This function relates the welfare of \( H \) individuals to the welfare of society. The following equation results by substituting indirect utility functions into equation (2.6) and fully differentiating:

\[ dw = \sum_{h=1}^{H} \frac{\partial W}{\partial V^h} \text{d}V^h \]  \hspace{1cm} (2.7)

where \( w \) is societal welfare, \( V^h \) is an indirect utility function for individual \( h, h=1, \ldots, H \).

Equation (2.8) is then obtained by totally differentiating the indirect utility function (see Johansson 1987) and substituting it in equation (2.7). This equation (2.8) shows that the contribution of each individual to social welfare depends, inter alia, on their welfare weight\(^{24} \) and on their marginal utility of income.

\[ dw = \sum_{h=1}^{H} \frac{\partial W}{\partial V^h} \lambda^h [-x^h \text{dp} + dy^h] \]  \hspace{1cm} (2.8)

where \( \frac{\partial W}{\partial V^h} \) shows the change in social welfare given a change in the utility of individual \( h \) (also described as the welfare weight of individual \( h \)), \( \lambda^h \) represents the marginal utility of income of individual \( h \), \( x^h \) represents the demand for a vector goods by individual \( h \), \( dp \) is the change in indirect utility given a change in the vector of prices, and \( dy^h \) is the change in indirect utility given a change in income for individual \( h \).

welfare weights and the marginal utility of income of the gainers and losers.
\(^{24} \) The relative weight given to each persons' welfare when determining the welfare of society as a whole. For example, each person would be given an equal welfare weight under utilitarianism.
The sum of the changes in individual consumer surplus is proportional to the change in social welfare if the product of the welfare weight and the marginal utility of income can be assumed to be constant across all individuals. It is then possible to derive a consistent social welfare function. The sum of changes in consumer surplus need not have the same sign as the change in social welfare if this product is different across individuals. Johansson (1987) pointed out that this assumption has the unreasonable implication that the welfare weight of a high income household exceeds that of a low income household. The only other alternative for deriving a consistent social welfare function is to assume that all individuals have the same welfare weights and that the marginal utility of income is constant across individuals. This assumption is also unreasonable. However, this appears to be the only other possibility for consistent aggregation of individual preferences.

In summary, it appears that it is possible to estimate the aggregate value of improved environmental quality. However, aggregate estimates are subject to certain assumptions. Given the necessity of these assumptions, aggregate estimates should be viewed as an allocative tool that represents the simple sum of the individual benefits of a policy change rather than a true social welfare function that provides a meaningful measure of the aggregate benefits to society of a policy change. This highlights the importance of decision makers separately assessing the distributive effect of policy changes.

2.8 Conclusions

The theoretical basis for analysing the effect of resource use changes on community welfare has been reviewed in this chapter. There have been three goals in undertaking this review. The first is to identify the values associated with the preservation of a natural resource. The second is to define the appropriate measure for the estimation of total economic value.
The third is to show how individual level estimates can be aggregated so that they can be used for policy purposes and any implications for model estimation.

Several components of non-use value have been considered in this chapter. The first was option value, which arises because of uncertainty about the future demand for, or supply of, a natural resource. In many cases, however, the sign of option value is indeterminate. Also the relevance of deriving separate estimates of option value is questionable. The main implication of the option value literature is that practitioners should focus on deriving estimates of option price, which is equal to compensating or equivalent surplus after allowing for the effects of uncertainty.

The second type of non-use value considered was quasi-option value. Quasi-option value reflects the value of deriving further information about the benefits of preservation or development by waiting. It was argued that while it is not possible to derive an accurate estimate of quasi-option value, respondents may make a subjective assessment of quasi-option value. This may form part of the reason why respondents are willing to pay to preserve a natural resource.

The third component of non-use value considered was existence value. Existence value reflects peoples’ willingness to pay to preserve a resource even though it involves no in situ use. It was suggested that existence value may exist for unique and non-unique goods, and even for non-environmental goods. It was also suggested that existence value could be negative. The magnitude of existence value was shown to depend on both the supply of and demand for natural resources. Motives behind existence values were found to matter. They can affect the magnitude, sign and estimability of existence value. The main implication from the review of motives is that problems may be experienced with estimating existence values if certain types of motives are prevalent.
The overall value of preserving a natural resource was then discussed. This was defined as total economic value, and is the sum of use value, option value, quasi-option value and existence value. It was argued that this concept of value is appropriate for use in stated preference studies as respondents generally form overall estimates of their willingness to pay to preserve a natural resource. Moreover, estimating total economic value has the advantage that double-counting is less likely. This concept of value will principally be used in this thesis, although non-use value is expected to be the major component of total economic value.

Next, different measures of total economic value were evaluated. Compensating surplus was asserted to be the relevant measure for deriving estimates of willingness to pay given that respondents do not have property rights for the improvement. Similarly, equivalent surplus, or willingness to accept compensation, is the correct measure if respondents have property rights for the subsequent situation. These measures are superior to Marshallian consumer surplus because they produce unique results and allow for income effects.

Lastly, the issue of aggregating individual preferences was discussed. Various approaches have been suggested, but all have their limitations. The Kaldor-Hicks potential compensation principle does not necessarily produce unique outcomes, and requires that compensation be paid for there to be certainty that a new alternative will improve welfare. Aggregation of ordinal preferences has been shown to be impossible. The most common approach used to aggregate preferences is to assume full cardinal comparability, and that welfare weights and the marginal utility of income are constant across individuals. The main implication from these assumptions is that when interpreting aggregate estimates of total economic value it should be remembered that distributional effects have
not been incorporated and hence these estimates do not fully represent welfare changes.

Thus, with the values associated with preservation identified, the appropriate unit of measurement defined, and a process for incorporating the results into policy established, the context of the thesis has been set in place.

\footnote{It is difficult to separately estimate use and non-use values in any survey. It is anticipated that use values may comprise part of the total economic value that is estimated in this thesis. Some evidence of this is provided in Chapter 7.}
Chapter 3 Techniques for Estimating Non-use Values

3.1 Introduction

In Chapter 2, the theoretical basis for estimating the value of changes in the quality or quantity of natural resources was reviewed. In this chapter, alternative techniques that can be used to estimate non-use values are reviewed. The class of techniques that can be used to estimate non-use values are known as stated preference (SP) techniques. The characteristic feature of SP techniques is the use of surveys in which attempts are made to elicit directly from respondents their willingness to pay (or accept compensation) for a change in environmental quality or some other policy change. It is not possible to use techniques based on the notion of weak complementarity, such as the travel cost or hedonic price methods, because non-use values are not revealed either directly or indirectly in conventional markets.

The SP technique most widely used to estimate non-use values is the CVM. The CVM, however, has several limitations. It can be cumbersome and expensive when being used to estimate the values of a wide range of potential environmental outcomes. Separate sub-samples of the affected population must be surveyed to estimate the value of each particular policy alternative. In addition, the capacity of the CVM to provide valid

\footnote{Sections 3.3-3.6 of this chapter have also been reported in Morrison, Blamey, Bennett and Louviere (1998).}
estimates of non-use values has been debated in the literature (Kahneman and Knetsch 1992; Diamond and Hausman 1994; Hanemann 1994, 1995; McFadden 1994, Smith 1996).

In response to these limitations, some researchers have proposed the use of conjoint analysis to estimate the value of changes in environmental quality (eg Mackenzie 1993; Gan and Luzar 1993; Johnson and Desvousges 1997; Rolfe and Bennett 1996b). Conjoint analysis is a generic term coined by Green and Srinivasan (1978) to describe a variety of different SP techniques, such as those that involve the use of ratings scales, rankings, paired comparisons, and choices. Specifically, Green and Srinivasan (1978, p.104) defined conjoint analysis as:

any decompositional method that estimates the structure of a consumer's preferences given his/her overall evaluations of a set of alternatives that are prespecified in terms of levels of different attributes.

An attractive feature of conjoint techniques is the more explicit focus on attributes, which may result in more discriminating responses than with the CVM. Conjoint techniques also have the advantage that they provide much more information about the structure of people's preferences. Hence the results from a single survey can be used to estimate the value of multiple alternatives.

The CVM and four conjoint techniques—CM, contingent ranking, contingent rating and paired comparison—are reviewed in this chapter to determine their potential for estimating non-use values. The use of the CVM is reviewed in Section 3.2, and the four conjoint techniques are evaluated in Sections 3.3 to 3.6. Conclusions are offered in Section 3.7.
3.2 Contingent Valuation

The use of the CVM to estimate the value of improved environmental quality was first suggested by Ciriacy-Wantrup (1947). It was initially trialed by Davis (1963), Ridker (1967), Brown and Hammack (1974) and Randall, Ives and Eastman (1974). Since then, the number of studies has grown exponentially, with Carson, Wright, Carson, Alberini and Flores (1994) producing a CVM bibliography containing 1674 entries. In Australia, the growth has been less dramatic, with the ENVALUE database listing 26 studies (Morrison, Groenhout and Moore 1995).

3.2.1 Description

Like all SP techniques, CVM questionnaires contain several well defined elements including a description of the study site, details of the proposed changes (including a method of payment), an elicitation question and a series of socioeconomic and attitudinal debrief questions. The earliest CVM applications typically used open-ended questions such as ‘what is the maximum increase in taxes you would be willing to pay to achieve outcome X?’ (eg Gramlich 1973, Brown and Hammack 1974). However, this format was considered to be too cognitively demanding for respondents (Brookshire et al 1986) and subject to strategic biases. The most common elicitation format for the CVM is now the dichotomous choice or referenda format\(^{27}\) (Bishop and Heberlein 1979)\(^{28}\). As shown in Table 3.1, under this format respondents are asked whether they would be willing to pay a fixed amount to achieve a given environmental outcome.

\(^{27}\) The dichotomous choice CVM format is known as the ‘referenda’ format when it is designed to be similar to a referendum where, given a proposal, respondents are asked to cast a ‘yes’ or ‘no’ vote.
The amount that respondents are asked to pay is varied across different sub-samples. The data are then analysed using binary logit or probit regression techniques and welfare estimates are derived using the resulting model.

Table 3.1: The dichotomous choice CVM format

<table>
<thead>
<tr>
<th>Do you support the proposal to improve water quality at a cost of $50 per household, or do you oppose the proposal? (tick one box)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I support the proposal at a cost of $50</td>
</tr>
<tr>
<td>I oppose the proposal at a cost of $50</td>
</tr>
</tbody>
</table>

The dichotomous choice version of the CVM is consistent with the random utility model (RUM), which states that consumers seek to maximise utility in choice, but there is randomness in their choices due to the researcher’s inability to establish all influences on choice (Thurstone 1927, Manski 1977). The RUM assumes that utility is inherently unobservable by analysts; and hence the utility of a good is decomposed into (1) an observable component, which consists of the quantity of the non-market good (x) and individual characteristics (s), and (2) an unobservable or random component (e). Thus, the indirect utility of good i can be expressed as $V_i(x, s) + e_i$, where $V(\cdot)$ is the deterministic component of utility. In the RUM, the probability of choosing alternative i from some set of competing alternatives, A, can be written as shown in equation (3.1). The probability of choosing the ith alternative rather than the jth alternative is equal to the probability that the deterministic (V) plus random utility (e) of i is greater than j.

Other elicitation formats include iterative bidding (eg Randall, Ives and Eastman 1974), payment cards (eg Smith, Desvousges and Fisher 1983), double-bounded dichotomous choice (Hanemann 1985), polychotomous choice (Ready, Whitehead and Blomquist 1995) and dissonance-minimisation formats (Blamey, Bennett and Morrison forthcoming).
\[ P(i | j \in A) = P( (V_i + e_i) > (V_j + e_j) ) \] \hspace{1cm} (3.1)

The dichotomous choice format is favoured more than other CVM formats because of its realism (it can be designed to be similar to a referenda – see Arrow et al 1993), its behavioural basis (the RUM) and its incentive compatibility\(^{29}\) (it is less susceptible to strategic bias; Hoehn and Randall 1987). A panel of economic experts, set up by the National Oceanic and Atmospheric Administration in the United States, concluded that the dichotomous choice format CVM ‘can produce estimates reliable enough to be the starting point for a judicial process of damage assessment, including lost passive-use values’ (Arrow et al 1993, p.4610). The NOAA panel provided, however, only qualified support for the use of CVM and recommended that various guidelines be adhered to in applications.

### 3.2.2 Welfare estimation

Indirect utility is assumed in the CVM, to be a function of the level of the non-market good of interest, a composite good, prices and socioeconomic characteristics:

\[ V = V(x, z, p, s) \] \hspace{1cm} (3.2)

where \( V \) is indirect utility, \( x \) is the quantity of a non-market good, \( z \) is a composite of all other goods, \( p \) is a vector of prices and \( s \) is a vector of socioeconomic characteristics.

Assuming that the only personal characteristic in \( s \) is income and that the level of the composite good and prices are constant, compensating surplus is found by reducing income under the new alternative until the individual is indifferent between the two alternatives:
\[ V(x^0, m) = V(x', m-cs) \]  

where \( x^0 \) is the original quantity of the non-market good, \( cs \) is compensating surplus and \( m \) is income.

### 3.2.3 Other issues

Despite the official certification provided by the NOAA panel, the CVM remains a controversial technique. Ciriacy-Wantrup’s (1952) prediction that the use of the CVM would be criticised now appears prophetic\(^\text{30}\).

Various criticisms have been made of the CVM, including that respondents (1) act strategically to manipulate the final outcome; (2) treat CVM questionnaires as if they are hypothetical; (3) are insensitive to the scope of the good being purchased (perfect embedding or part-whole bias); (4) tend to ‘yea-say’; and (5) that willingness to pay and accept compensation differ by greater amounts than predicted by theory. These criticisms have been reviewed in detail in Mitchell and Carson (1989) and Morrison, Blamey, Bennett and Louviere (1996).

### 3.3 Choice Modelling

CM was developed originally in the marketing and transport literature by Louviere and Hensher (1982) and Louviere and Woodworth (1983). More recently it has been applied in the environmental context. For example, Adamowicz, Louviere and Williams (1994) used CM to estimate water-based recreational values; Eom (1994) estimated the value of pesticide risk.

---

\(^{29}\) Hoehn and Randall (1987) show that in a referendum model with individually defined costs, truth telling is the optimal strategy for respondents. Thus there is no incentive to act strategically. This contrasts with the open-ended CVM where strategic behaviour can be an optimal strategy.

\(^{30}\) Ciriacy-Wantrup (1952) suggested that the following objections to the use of CVM would be made: non-additivity of individual preferences, problems of ‘lumpiness’ in
reduction in food products; Adamowicz, Swait, Boxall, Louviere and Williams (1996) estimated the non-use value for preserving caribou habitat; and Rolfe and Bennett (1996b) estimated the non-use value associated with preserving rainforests.

### 3.3.1 Description

In CM questionnaires, respondents are presented with a series of choice sets, each containing usually three or more resource use alternatives (profiles). An example of a choice set is shown in Table 3.2. Respondents are asked to choose their preferred alternative from each choice set. The alternatives in the choice sets are defined using a common set of attributes (i.e., resource use characteristics such as water quality, number of waterbirds, etc.), the levels of which vary from one alternative to another. The alternatives used in the choice sets are developed using experimental design techniques, which are discussed in Chapter 4.

**Table 3.2: Example of one choice set in a choice modelling questionnaire**

<table>
<thead>
<tr>
<th>Water quality</th>
<th>Alternative 1</th>
<th>Alternative 2</th>
<th>Alternative 3 (the status quo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>50,000</td>
<td>100,000</td>
<td>30,000</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>20,000 ha</td>
<td>60,000 ha</td>
<td>20,000</td>
</tr>
<tr>
<td>Area</td>
<td>$10</td>
<td>$40</td>
<td>$0</td>
</tr>
</tbody>
</table>

extra-market goods, strategic bias, problems from not valuing other extra market goods simultaneously, and that the exercise is simply too academic.

It is generally assumed that the multiple responses of each individual are independent. However, it is possible that responses will be correlated, even if respondents are encouraged to treat each choice set independently. Adamowicz, Swait and Louviere (1997) reported current research that showed correlated responses will only impact the efficiency of parameter estimates and not their unbiasedness. They also suggested that it is possible to introduce a correction factor to enable the estimation of more accurate standard errors.
In CM, as well as other conjoint techniques, a Lancastrian approach to demand theory is used. Lancaster (1966) argued that goods could be decomposed into a set of 'attributes' or characteristics. For instance, a car could be considered to be simply the sum of its component parts ie 4 wheels, a chassis, an engine etc. This contrasts with the CVM where the goods themselves, rather than the goods' attributes, are what is assumed to be valued by individuals. With this concept of demand, indirect utility can be expressed as shown in equation (3.4). Utility is also assumed to be additively separable in x and z so that the marginal rate of substitution between any of the attributes in x is independent of the level of consumption of z.

\[ V = V(x, z, p, s) \] (3.4)

where \( V \) is indirect utility, \( x \) is a vector of attributes, \( z \) is a composite of all other goods, \( p \) is a vector of prices, and \( s \) is a vector of socioeconomic characteristics.

Assuming that the only personal characteristic in \( s \) is income (m) and that \( z \) and \( p \) are constant, compensating surplus is defined as follows:

\[ V(x^0, m) = V(x^1, m-cs) \] (3.5)

where \( x^0 \) is the original vector of attributes and \( cs \) is compensating surplus.

Similar to the dichotomous choice CVM format, CM is consistent with the RUM (see equation 3.3). The type of model that is used to analyse choice data depends on the assumption made about the distribution of the error terms \( (e_i \text{ and } e_j) \). The most common assumption is that the error terms are independently and identically distributed (IID) and follow a Gumbell distribution. This assumption results in McFadden's (1974) conditional logit model, which is often called the multinomial logit (MNL) model. The MNL model has the following form, where \( V_{ij} \) is the deterministic utility component of alternatives i and j, and \( \lambda \) is a scale parameter:

\[ P(i|A) = \frac{\exp(\lambda V_{ij})}{\sum_j \exp(\lambda V_{ij})}, \forall i, j \in A \] (3.6)
The indirect utility function $V_i$, which represents the utility of the different alternatives in the MNL model, can have different functional forms. The simplest functional form involves an additive structure. Additive structures only include the attributes from the choice sets eg:

$$V_i = C + \sum \beta_k X_k \quad \text{where } i = 1, \ldots, N \text{ and } k = 1, \ldots, N \quad (3.7)$$

where $C$ is an alternative specific constant, $\beta$ is a coefficient, and $X$ is a variable representing an attribute from a choice set.

The alternative specific constants ($C$) show the effect of systematic but unobserved factors on respondents' choices. In other words, while the $X$ variables show the effect of deterministic variables (i.e., the attributes in the choice sets), the constants capture the unobserved factors that explain choice. Technically they reflect the mean of the differences in the error terms (Ben-Akiva and Lerman 1985). It is possible to have $J-1$ alternative specific constants in a MNL model (where $J$ is the number of alternatives). This is because the constants are based on differences between the alternative options and the base option.

More complex specifications, which include socioeconomic and attitudinal variables, are possible. Socioeconomic and attitudinal variables cannot be included directly into utility functions as they are invariant across alternatives in a choice set. Instead they are included interactively, either with the alternative specific constant ($C$), or with one of the attributes ($X$) (see Swallow, Weaver, Opaluch and Michelman 1994) eg:

$$V_i = C + \sum C_S h + \sum \beta_k X_k + \sum \beta_k S h X_k \quad (3.8)$$

where $i=1, \ldots, N$, $k=1, \ldots, K$, $h=1, \ldots, H$, $C$ is an alternative specific constant, $\beta$ is a coefficient, and $X$ is a variable representing an attribute from a choice set, and $S$ represents socioeconomic or attitudinal variables.
CM can be used to produce theoretically correct welfare estimates. Where a MNL model is estimated that involves several different goods, compensating surplus can be estimated using the following expression (Hanemann 1984)\(^{32}\):

\[
CS = - \frac{1}{\beta_M} \left\{ \ln\left( \sum_i \exp^{V_0} \right) - \ln\left( \sum_i \exp^{V_1} \right) \right\}
\]  

(3.9)

where \(\beta_M\) is the coefficient for the monetary attribute and is interpreted as the marginal utility of income, \(V_0\) represents the utility of the initial state, and \(V_1\) represents the utility of the subsequent state (for each of the alternatives).

When only a single good is modelled, equation (3.9) reduces as follows:

\[
CS = - \frac{1}{\beta_M} (V_0 - V_1)
\]

(3.10)

Point estimates of the value of a unit change in an attribute can be found by estimating implicit prices (also known as ‘part-worths’). Implicit prices are equivalent to the marginal rates of substitution between environmental (or other) attributes and the monetary attribute. Implicit prices are not, however, welfare estimates which are equal to compensating surplus. Implicit prices are calculated as follows, if utility is a linear function of all attributes:

\[
IP = \frac{\beta_A}{\beta_M}
\]

(3.11)

where IP is the implicit price, \(\beta_A\) represents the coefficient of the \(A\)th non-monetary attribute, and \(\beta_M\) represents the monetary attribute.

\(^{32}\) An equivalent formula has been derived by McConnell (1995).
Equations 3.9 to 3.11 may produce inaccurate estimates of compensating surplus if a constant base alternative has not been included in each of the choice sets. The appropriate form for the constant base alternative will depend on the context of the survey. In the context of estimating non-use values it may represent the status quo at the time of the survey (eg 'continue current situation'), or it may represent, for example, what will happen in the longer term if no changes to policy occur. In the context of estimating use values, it may simply be 'I would choose not to visit' or something similar.

There are several reasons why excluding a constant base alternative may affect value estimates. First, if a constant base alternative is excluded from choice sets, respondents cannot indicate that they would not choose to support any of the alternatives under consideration. Omitting a constant base alternative is therefore likely to affect the choice probabilities of the other alternatives and model coefficients. Secondly, model coefficients may be biased when based on trade-offs between alternatives that are not within the respondent's feasible choice set, particularly if indifference curves are not homothetic. Biased estimates of willingness to pay may result. Thirdly, value estimates derived from trade-offs relative to a base alternative may be different from value estimates derived from trade-offs solely between non-base alternatives. Empirical evidence suggests that respondents value losses and gains differently (Knetsch 1990, Peterson et al 1996). Hence there may be a discontinuity in respondents' valuation of changes in environmental quality around the status quo\textsuperscript{33}. By not estimating values relative to a base alternative, respondents' perceptions of what is a gain and what is a loss may change, potentially affecting value estimates. A fourth reason for including a constant base alternative is that it may improve the plausibility of the questionnaire\textsuperscript{34}. By including a base alternative, a choice set is similar to a referendum question, which increases its realism (see Arrow et al 1993). For these reasons it is critical

\textsuperscript{33} This is effectively another reason why indifference curves may not be homothetic.
to include a constant base alternative when estimating non-use values for cost-benefit analysis.

It should, however, be noted that in some CM applications it may not be appropriate to include a base alternative if the goal of the study is not the estimation of compensating variation or surplus. This may be the case if the goal of the study is to determine which alternatives are likely to receive the largest amount of support in the community. For example, where it has been proven that a new landfill site is needed and information is sought on the relative importance of different environmental quality indicators (eg Opaluch, Swallow, Weaver, Wessells and Wichelns 1993).

### 3.3.3 Other issues

An important aspect of the MNL model is the embedded scale parameter, \( \lambda \), which is inversely proportional to the standard deviation of the error distribution. In the case of the Gumbell distribution, it can be shown that \( \lambda^2 = \pi^2/6\sigma_e^2 \), where \( \sigma_e^2 \) is the variance of the error distribution (Swait and Louviere 1993). The scale parameter (\( \lambda \)) cannot be identified separately from the \( \beta \) vector of estimates in any specific data source; hence it is usually arbitrarily set to one. This scaling generally has no effect on the value estimates derived from a MNL model\(^{35} \), because the scale parameter can be eliminated by dividing the \( \beta \) estimates for all other explanatory variables by one reference \( \beta \). This is seen in the following implicit price:

\[^{34} \text{In Section 6.2.3 it is noted that participants in focus groups requested inclusion of a constant base alternative for this reason.}\]

\[^{35} \text{Assuming the model effects are strictly linear in price, and that the scale parameter is not estimated as a function of socioeconomic or attitudinal variables.}\]
IP = \lambda \beta_{env} \lambda \beta_m = \frac{\beta_{env}}{\beta_m} \tag{3.12}

where IP is an implicit price, \lambda is a scale parameter, \beta_{env} is a coefficient for an environmental variable, and \beta_m is a coefficient for a monetary variable.

However, differences in the scale parameters should be considered when comparing two or more models because variance differences impact on the parameter magnitudes. That is, even if the underlying preference process is identical, model parameters from two different choice data sources will be unequal if the random component variances are unequal. It is possible to identify the ratio of scale parameters in two or more data sets, although the scale parameter cannot be identified in any particular model. Valid model comparisons between choice models can be undertaken by estimating the scale ratio, and testing whether the two model parameter estimates are equal after taking the variance differences into account (Swait and Louviere 1993).

One non-trivial limitation of the MNL model is the requirement that the responses are independent of irrelevant alternatives (IIA). The IIA property results from the IID error assumption, which requires that the probability of choosing one alternative over a second depends only on the utility of the respective alternatives. Hence, ratios of choice probabilities for pairs of alternatives are independent of the existence or inclusion of other alternatives. Violations of the IIA property may occur for a number of reasons. For example, the inclusion of close substitutes in a choice set may lead to violations when their unobserved effects are correlated. Specification errors may also lead to violations of this property. The existence of random taste variations (ie heterogeneous preferences) may lead to IIA violations when the heterogeneity is not allowed for in the model. Various tests for the existence of IIA violations are examined in Chapter 4. If an IIA violation is found and it is not possible to modify the existing MNL model to remove the violation (eg by including individual
characteristics), more complex models that relax some or all of the IID assumptions can be used. These models are also reviewed in Chapter 4.

A further issue with CM, and indeed all conjoint techniques, is how respondents react to the complexity of the task. In contrast to the CVM, where respondents are typically required to choose between two alternatives, conjoint techniques typically require the evaluation of larger number of alternatives. This can be expected to result in additional task complexity when compared to CVM. The complexity of a given conjoint task depends *inter alia* on the number of alternatives in the task as a whole and within any sub-task, and the number of attributes used to describe the alternatives. The complexity of conjoint tasks, in conjunction with the bounded rationality of respondents, may give rise to the use of simplified decision strategies (or heuristics). These heuristics can take many different forms. They are often associated with non-compensatory decision strategies, such as lexicographic preference orderings and elimination by aspects (Tversky 1972; Payne and Braunstein 1978), although they can take other forms. Blamey, Bennett, Louviere, Morrison and Rolfe (1997) reported evidence of the use of a causal heuristic in which respondents seek to impose a causal structure on attributes with a view to basing their choices on the more fundamental, causally prior, environmental attributes. The use of heuristics, however, tends to be reduced when tasks are simplified. For example, Payne and Braunstein (1978) and Mazotta and Opaluch (1996) found that reducing task complexity reduced the random error component of choices. Therefore, it may be possible to minimise problems related to the use of heuristics by carefully designing and piloting questionnaires in order to reduce task complexity. As will be seen in subsequent sections, problems relating to task complexity are by no means limited to CM applications.

Compared to the CVM, CM would be expected to have several advantages. The first is that the more explicit focus on attribute differences may result in more discriminating responses. This could
reduce the incidence of biases such as yea-saying (Blamey, Bennett and Morrison forthcoming). The second is that more information is gathered about the structure of respondents’ preferences; hence it is possible to value multiple alternatives using a single survey. This is also an attractive feature for benefit transfer, as it means that environmental changes at the two different sites can be more closely matched (see Chapter 8). Third, it may be possible to value simultaneously multiple goods within a CM exercise (see Section 4.2). However, these features do come with a cost. CM questionnaires tend to be more cognitively demanding for respondents, and hence there may be greater potential for the use heuristics. The process of questionnaire design and data analysis is also more demanding for researchers.

3.4 Contingent Ranking

Contingent ranking was one of the first conjoint techniques to be used to estimate the value of improved environmental quality. It has been used to estimate the value of improved air quality in urban and rural areas (Rae 1983, 1984), improved water quality (Smith and Desvousges 1986), hazardous waste risk reduction (Smith, Desvousges and Freeman 1985), reduced diesel odours from motor vehicles (Lareau and Rae 1987) and the preservation of forests (Garrod and Willis 1997).

3.4.1 Description

Respondents rank three or more resource use alternatives from most to least preferred in contingent ranking applications, as shown in Table 3.3. Similar to CM, each of the alternatives contains a common set of attributes. The alternatives are usually generated using experimental design techniques. Respondents are either required to rank the full set of
An advantage of contingent ranking suggested in the literature (eg Beggs, Cardell and Hausman 1981) is that, relative to other SP techniques, it provides additional statistical information compared with a single choice of the most preferred alternative. That is, information is obtained about respondents' first choices, as well as their subsequent choices. However, the benefits of using information from lower and middle ranks is uncertain, as will be discussed below.

**Table 3.3: Example from a contingent ranking questionnaire**

<table>
<thead>
<tr>
<th></th>
<th>Alternative 1</th>
<th>Alternative 2</th>
<th>Alternative 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water quality</td>
<td>fair</td>
<td>good</td>
<td>poor</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>50,000</td>
<td>100,000</td>
<td>50,000</td>
</tr>
<tr>
<td>Area of wetland</td>
<td>60,000 ha</td>
<td>60,000 ha</td>
<td>20,000 ha</td>
</tr>
<tr>
<td>Household cost</td>
<td>$40</td>
<td>$70</td>
<td>$10</td>
</tr>
</tbody>
</table>

The use of contingent ranking to estimate the value of improved environmental quality was stimulated by the seminal study by Luce and Suppes (1965). They demonstrated that contingent ranking could be consistent with the RUM, and that statistical models derived from the RUM could be used to analyse the data. They also provided researchers with a theoretically sound way to analyse the results of a ranking exercise based on the RUM. Previously, researchers analysed the data using ordinary least squares regression (eg Whitmore and Cavadias 1974). Luce and Suppes (1965) showed that ranking alternatives was equivalent to the probability of choosing the good with the highest utility from a choice set A, multiplied by the probability of choosing the good with the next highest utility out of the remaining goods, etc. This can be seen in the following conditional probability distribution:
Pr(U_1 > U_2 > ... > U_H for H ≤ A) = \prod_{h=1}^{H} \left[ \frac{e^{v_h}}{\sum_{i=1}^{H} e^{v_i}} \right]  

(3.13)

where \(i\) represents a resource use alternative, \(A\) is a set of alternatives, \(h\) the rank, and \(H\) the total number of rankings (which is equivalent to the lowest ranking).

Following from this theoretical derivation, response data from contingent ranking exercises are usually analysed using rank ordered logit models (Beggs et al 1981).

3.4.2 Welfare estimation

The alternatives ranked by respondents in past contingent ranking exercises have not always included a constant base alternative (eg Smith and Desvousges 1985). Similar to CM, the exclusion of a constant base alternative in a contingent ranking exercise may result in inaccurate welfare estimates. Mackenzie (1993) demonstrated the equality of implicit prices derived using CM and contingent rankings when constant base alternatives are excluded. Using a common data set, he estimated implicit prices using both ordered and binary logit models, and found that mean values were statistically equivalent.

As with CM, it is possible to include a constant base alternative in every set of alternatives ranked by respondents (eg Lareau and Rae 1987, Garrod and Willis 1997). However, it is possible that the welfare estimates derived using these techniques will diverge when a base alternative is included. This is because contingent ranking includes data from the trade-offs between all alternatives, regardless of whether each trade-off involves a base alternative. The use of data on trade-offs between all alternatives implies that marginal rates of substitution are averaged across the indifference map. In contrast, with binary choice models at least one
alternative in each trade-off is always the base alternative. The marginal rates of substitution are therefore always calculated relative to the base alternative. If indifference curves are not homothetic, the exclusion of these other trade-offs may result in different model coefficients, implicit prices, and estimates of compensating surplus or variation. The existence or extent of this bias in estimates derived using the contingent ranking method, however, has not yet been demonstrated.

3.4.3 Other issues

Whether contingent ranking exercises are more cognitively burdensome than CM exercises will clearly depend on the context of choice and the way the task is structured. Whether one approach is more demanding than another, and potentially more susceptible to the use of various heuristics, is hence an empirical question. Several studies have, however, identified a form of bias that is unique to ranking tasks. It has been found that the variance of alternatives change depending on where they have been ranked (e.g., Chapman and Staelin 1982; Hausman and Ruud 1987; Ben-Akiva, Morikawa and Shiroshi 1991). This may result in violations of the IID assumption associated with the rank ordered logit model, and may cause model coefficients, and hence estimates of value, to be biased.

For a given ranking exercise where alternatives are ranked A>B>C>D and Alternative ‘C’ is the base alternative, the trade-offs used in estimating the model are: A>B, A>C, A>D, B>C, B>D, and C>D. In contrast with a binary logit model that has a constant base alternative, only three trade-offs are considered: A>C, B>C, C>D. The other trade-offs (i.e., A>B, A>D, B>D), which are not related to movements from the base alternative, are not included in the analysis.

Following similar reasoning it is possible that there may be divergences between estimates derived using binary and MNL models. However, further research is needed to clarify this issue.

As reported in Section 3.5, some evidence is available from a study by Roe, Boyle and Tiesl (1996). They converted ratings data into rankings and binary choice data. They compared estimates of compensating variation derived using both approaches. The rankings data were analysed using an ordered logit model, and the choice data using a binary logit model. A base alternative was included in all binary choices and rankings. The welfare estimates were consistently 2-3 times larger in the rankings model.
Increasing variance at lower or middle ranks may result from respondents paying less careful attention to either lower or middle ranked alternatives. It may also result if respondents resort to the use of heuristics, especially if rankings are complex. For example, respondents might rank the highest and lowest utility profiles first, and randomly rank those in the middle; or they might rank only the top one or two and randomly rank the remaining alternatives. Increasing variance at lower and in middle ranks due to the use of heuristics may explain why some of the earlier contingent ranking studies that focused on environmental issues obtained counter intuitive results (e.g. Rae 1983, 1984). Both Smith and Desvousges (1986) and Lareau and Rae (1987) conjectured that the reason that these studies produced counter-intuitive results was because respondents may have been overwhelmed by the task of ranking a large number of alternatives.

Hausman and Ruud (1987) and Ben-Akiva et al (1991) have demonstrated the effect on model coefficients of including data from the lower ranks of a ranking exercise. In Hausman and Ruud’s (1987) study, respondents ranked eight alternatives each having eight attributes. Each alternative represented a different cellular phone. Seven different models were estimated. The first model used data only from the first rank, the second model used data from the first and second rank, the third model used data from the first, second and third rank and so on. Using a likelihood ratio test, Hausman and Ruud (1987) found that the model based on data from all ranks was not equal to a model that excluded the bottom three ranks. They also compared the implicit prices generated by the seven models. Except for two continuous variables, implicit prices were found to be inconsistent over ranks. Ben-Akiva et al (1991) also tested for the equality of model coefficients across ranks. Their study was slightly different to that of Hausman and Ruud (1987). It involved the choices between alternative transport options. Respondents were asked to rank four alternatives that each had three attributes. Separate models were

---

39 Hausman and Ruud (1987) also examined implicit prices in the study by Rae (1983) and found that they were inconsistent.
estimated only using data from each rank (ie for rank 1, 2 or 3). They separately compared models estimated at different ranks using likelihood ratio tests and found that they were different.

Hausman and Ruud (1987) proposed use of a heteroscedastic ordered logit model to account for differences in variance. This model allows for differences in variance through a scale parameter that varies across rankings. They found that, in general, the scale parameter increased with rank. However, it was found not to increase monotonically. They did not test whether models estimated using ranks of different depth were equivalent after allowing for variance differences. Ben-Akiva et al (1991) also compared MNL models after allowing for differences in scale. They concluded that the data from only the top two ranks were consistent after allowing for differences in variance.

Whether simplification of ranking exercises would produce more homogeneous variances across ranks is an empirical question which can be expected to depend on the context of the valuation exercise. The evidence from Hausman and Ruud (1987) and several of the earlier ranking studies which focused on the valuation of environmental goods, suggests that inconsistencies are likely to occur when eight alternatives are ranked, even if differences in variance are taken into account. Lareau and Rae (1987) conjectured that the use of only four alternatives, may lead to more plausible results. However, the results from Ben-Akiva et al (1991) suggest that problems may still be faced even with very simple ranking exercises.

3.5 Contingent Rating

The contingent rating approach originated in psychology (Anderson 1982), and has been widely applied in marketing. For example, Wittink and Cattin (1989) reported that there were about 400 commercial conjoint
studies undertaken each year in the USA during the early 1980s, 49% of which used contingent rating. Contingent rating has been used to estimate environmental values in applications such as recreational fishing (Roe, Boyle and Teisl 1996), recreational hunting (Mackenzie 1992, 1993; Gan and Luzar 1993), preservation of bandicoots (Jakobsson, Kennedy and Elliott 1995), and improved ground water quality (Stevens, Barrett and Willis 1997).

3.5.1 Description

In a typical contingent rating application, a series of resource use alternatives are evaluated by respondents, one at a time, using a ratings scale. Respondents do not compare the different profiles per se, but rather rate each separately on some latent dimension like “intention to visit”. Table 3.4 provides an example of one possible alternative in a series that respondents might evaluate.

Table 3.4: Example from a contingent rating questionnaire

| Please circle one of the numbers below to show your preferences for the following alternative: |
|-----------------------------------------------|-------------------|
| Water quality                                | fair              |
| Number of waterbirds                         | 50,000            |
| Area of wetland                              | 60,000 ha         |
| Household cost                               | $40               |
| 1     2     3     4     5     6     7     8     9     10 |
| Weakly Preferred                             | Strongly Preferred|

In contingent ratings applications indirect utility is assumed to be a function of a vector of attributes of the good of interest and a vector of socioeconomic characteristics etc. (see equation 3.14). Utility is assumed to be related to individuals’ ratings via a transformation function $\phi(.)$: 63
\( r_i(x, z, p, s) = \phi[U_i(x, z, p, s)] \) \hspace{1cm} (3.14)

where \( r_i \) is a rating for good \( i \)

A theoretical basis for the traditional contingent rating approach has been proposed using information integration theory and/or social judgment theory (Anderson 1982, Lynch 1985, Louviere 1988a). Both paradigms are concerned with the process by which individuals evaluate and integrate attribute information to form holistic evaluations of combinations of attribute levels. Respondents are assumed to value each piece of information about an option presented to them separately, and then integrate the values to form a holistic evaluation of a bundle of information, which is expressed using a cardinal ratings scale. The individual's true evaluation of the information cannot be observed, hence it is assumed that the observed rating is linearly related to the unobserved latent dimension with an associated (IID normal) error distribution. However, it is not clear whether these assumptions are reasonable and whether individuals' ratings can be transformed into utility. As will be discussed below, this is because of the use of cardinal ratings scales and the difficulties associated with comparing ratings across individuals.

The data collected from contingent ratings questionnaires have traditionally been analysed using ordinary least squares regression (OLS), particularly in marketing applications. Respondents' ratings are regressed against the attribute of the good, as follows:

\[ R = \beta_0 + \beta_1 X_1 + \ldots + \beta_N X_N \] \hspace{1cm} (3.15)

where \( R \) represents respondents' ratings, and the \( X \)'s represent attributes used to describe the alternatives evaluated by respondents.
In order to apply OLS to estimate the parameters, respondents' ratings must satisfy cardinal measurement properties. The cardinality assumption implies that a one unit difference is the same regardless of the numerical value of the ratings scale. For example, an increase from '4' to '5' on a ten point ratings scale implies the same increase in utility as an increase from '9' to '10'. If ratings are not cardinal, OLS regression yields biased and inefficient estimates (Mackenzie 1992). However, it is not necessary to assume that respondents' ratings are cardinal when modelling ratings data. It is possible to assume, although with some loss of information, that ratings data have only ordinal significance, and to model the data using ordered logit or probit models. Ordered models have been used in some recent environmental applications (eg Gan and Luzar 1993; Mackenzie 1992, 1993; Jakobsson et al 1995).

A further complication with using ratings scales is that respondents may use ratings in different ways (Schuman and Presser 1981). For example, two respondents may assign ratings of '8' and '10' respectively to an alternative, even though the welfare implications are equivalent for both individuals. In other words, each respondent may use a different metric when evaluating alternatives. Roe et al (1996) suggested one way of reducing the extent of this problem (see also Stevens et al 1997) is to allow for differences in centring points (ie origin). In each questionnaire, respondents are asked to rate an alternative that represents the status quo. This information is then used to create a dependent variable for the remaining alternatives based on differences from the status quo. Similarly, the attributes for each alternative are represented as deviations from the status quo. The subsequent model is described as a 'ratings difference model' and Roe et al (1996) contended that it 'removes this centring noise from the data because all differences are deviations from the status-quo rating' (p.146). Roe et al (1996) and Stevens et al (1997) both reported results where the ratings difference model yielded more significant explanatory variables than the traditional ratings approach, suggesting that the difference model had some success in reducing noise due to different
centring points. While this appears to be an improvement on the traditional model, the ratings difference model may still be less than ideal. Centring points are only part of the metric that respondents may use when evaluating alternatives. The scale units that respondents use to represent increments in utility are also critical, and cannot be corrected using a ratings difference model. For example, if the two respondents described above also gave the alternative representing the status quo a rating of '4', then the ratings difference model would not be able to correct for any noise from using different scales. An alternative approach would be to estimate separate models for each individual; however, this would require the collection of large amounts of data from each person which would not be feasible in many surveys.

### 3.5.2 Welfare estimation

In traditional applications of the contingent rating method, estimates of implicit prices have been derived by calculating marginal rates of substitution (eg Gan and Luzar 1993). However, as noted by Roe et al (1996), the traditional ratings approach cannot be used to estimate the welfare gains from resource use changes. This is because implicit prices are not theoretically correct estimates of welfare changes.

Roe et al (1996) and Stevens et al (1997) used the ratings difference approach to derive estimates of welfare change. Compensating surplus is determined by subtracting money from income until the ratings difference is equal to zero (see equation 3.16).

\[40 \text{ ie the ratings difference for the two individuals would be '4' and '6', even though the difference in utilities are identical for both individuals.}\]
\[ \Delta r_{1,0} = r_1(x_1, m-cs) - r_0(x_0, m) \]  
(3.16)

where \( \Delta r_{1,0} \) is the difference in ratings, the status quo is represented by alternative 0, and prices and the level of the composite good are assumed to be constant.

One issue that arises with the ratings difference approach is that no clear 'cut point' is defined for when an alternative would be chosen. While respondents might rate a new alternative more highly than the base alternative, it may not be the case that it is chosen. For example, respondents may think a new alternative is better value than the base alternative, and hence give it a higher rating, but still choose the base if they cannot afford the new alternative. Another explanation for this phenomenon is that respondents may think an alternative is a better one, but not choose it if they believe that the transactions costs involved with implementing the new alternative are high (see Swait, Louviere and Williams 1994). This may cause trade-offs, and hence model coefficients implied by ratings, to be biased. This is effectively a problem with the assumption in equation (3.14) that ratings can be transformed into utility. Stevens et al (1997) provided some evidence of the extent of this problem, as discussed below.

Several other methods have been proposed in the literature for deriving welfare estimates from ratings data for use in cost-benefit analysis. These methods involve analysing the data using a CM framework. All of these methods suffer to some extent from the same problem faced by the ratings difference model: that there is uncertainty about what rating an alternative must receive before it is 'preferred' to the status quo.

The first method, suggested by Louviere (1988b), involves assuming that every rating above a middle rating constitutes a 'yes' vote and that ratings below the middle rating constitute a 'no' vote. This method assumes that the ratings that signify 'yes' and 'no' votes are the same for each person. This may be a reasonable assumption if choice based labels are used to
indicate whether a respondent is in favour or against a proposal. Otherwise the designation of what constitutes a ‘yes’ or ‘no’ vote through the assignment of choice significance to the middle rating is subjective.

The second procedure was used by Roe et al (1996). They compared the rating given to each alternative with that of the status quo to determine whether a respondent would choose a particular alternative. For example, if a respondent gave the status quo a rating of ‘5’ and another alternative a rating of ‘6’, then they would choose the other alternative in a binary choice situation. Similarly, respondents would be assumed to choose the status quo for any alternatives given ratings of less than five. A problem arises if respondents give any other alternatives a rating of ‘5’, in which case their choice would be indeterminate. It is also not clear whether a single rating of the base alternative is sufficient to maintain the salience of this option throughout the entire ratings exercise. Roe et al (1996) compared estimates of compensating variation for five scenarios estimated using the ratings difference approach and the binary choice approach just described. They also converted the ratings data to ranking data and estimated an ordered logit model. They found that the results for the ratings difference model and the binary logit were statistically equivalent. However, the welfare estimates derived using the rankings data were between two to three times larger, thus demonstrating the effect of including inappropriate trade-off data when estimating model parameters.

A third procedure was used by Stevens et al (1997). They assumed that alternatives given a rating of ten only would be chosen in preference to the status quo. Respondents were told to only give alternatives a rating of ten if they ‘would definitely vote in favour of...the program (p.236)’. Stevens et al (1997) used this approach because they believed that respondents would not necessarily choose an alternative option just because it was rated more highly than the base option. They compared the value estimates generated using traditional ratings, ratings difference and binary logit and found that the ratings models resulted in substantially larger
estimates of willingness to pay. Stevens et al (1997, p.234) contended that:

the traditional rating and ratings difference value estimates are larger, in part because some respondents would not actually buy the commodity being valued. Our binary response model is defined in terms of whether individuals would pay...results derived from this model may therefore be more reliable...

### 3.5.3 Other issues

As with the other conjoint approaches, the complexity of ratings tasks depends on the characteristics of the task and the context in which it is applied. Whether ratings are more difficult than choices or ranking *ceteris paribus* is, similarly, an empirical question. On one hand, ratings might be more difficult than choices or orderings since additional information regarding cardinal differences between alternatives is required. On the other hand, ratings scales are often less closely linked to choice behaviour and some of the difficult trade-offs inherent in choice decisions may be avoided.

Similar to the other techniques, contingent ratings is susceptible to some unique problems associated with cognitive burden. One such problem arises from the imperfect information respondents have when they commence a ratings exercise about the utility space spanned by the set of alternatives. The ordered logit model, which is often used to analyse ratings data, requires the assumption of IID error terms. Violations of this assumption can occur, in practice, when respondents alter the metric they use to rate alternatives as they work their way through the ratings exercise. The errors associated with later ratings may differ to that of earlier ratings as a result. This may occur if the quality of the alternatives rated changes unexpectedly, or if there are learning effects. Errors may also result for the ratings difference approach because the base alternative is only rated once,
and respondents may lose sight of the reference level of utility provided through the base alternative$^{41}$.

3.6 *Paired Comparison*

The final SP approach considered in this chapter is paired comparison. The earliest use of the paired comparison approach for estimating environmental values involved recreation applications (Sinden 1974). Since then it has been used to value morbidity risk reductions (Magat, Viscusi and Huber 1988; Viscusi, Magat and Huber 1991; Krupnick and Cropper 1992; Johnson and Desvousges 1997), landfill sites (Opaluch et al 1993) and environmentally friendly electricity generation (Johnson and Desvousges 1997).

3.6.1 Description

The name 'paired comparison' has in the past referred to any study where respondents are presented with two alternatives and are asked to indicate their preferences between them. In some studies respondents are asked to choose their preferred alternative (eg Viscusi, Magat and Huber 1991; Krupnick and Cropper 1992; Opaluch et al 1993; and Peterson et al 1996), while in other studies respondents are asked to rate their preferences between the alternatives (eg Magat et al 1988; Johnson and Desvousges 1997). As the first of these forms is conceptually similar to CM, in this section we only review the later form. An example of this type of question is shown in Table 3.5.

$^{41}$ This may result in problems similar to what was described in Section 3.3.2 regarding respondents' different valuation of gains and losses. In that section it was noted that respondents value gains and losses differently, and excluding the base alternative may affect respondents' perspective on what constitutes gains and losses.
Table 3.5: Example from a paired comparison survey

Please indicate which of the two alternatives you prefer most by circling one of the numbers below:

<table>
<thead>
<tr>
<th></th>
<th>Alternative 1</th>
<th>Alternative 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>poor</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>50,000</td>
<td>100,000</td>
</tr>
<tr>
<td>Area</td>
<td>20,000 ha</td>
<td>60,000 ha</td>
</tr>
<tr>
<td>Household cost</td>
<td>$10</td>
<td>$40</td>
</tr>
</tbody>
</table>

**Strongly Prefer**

Alternative 1

1 2 3 4 5

**Strongly Prefer**

Alternative 2


The resource use alternatives evaluated by respondents in paired comparison exercises have often been generated using computer algorithms. Computer algorithms, such as ‘Adaptive Conjoint Analysis’ randomly select an initial pair of profiles for a respondent to evaluate (Green, Krieger and Agarwal 1991). Based on the respondent’s evaluation, the algorithm then successively presents the respondent with pairs of profiles that are close in utility to find points of indifference more efficiently. Points of indifference provide the most information about attribute values. Whilst intuitively appealing, problems with non-convergence have been experienced in some applications of Adaptive Conjoint Analysis (Johnson and Desvousges 1997).

In paired comparison, as with other SP techniques, utility is assumed to be a function of the attributes of the good of interest and socioeconomic characteristics etc. (see equation 3.4). The utility difference between the two alternatives in a paired comparison is estimated by calculating the difference between two indirect utility functions (Johnson and Desvousges 1997):

\[
\Delta U_{ij} = V^i(x, m) - V^j(x, m)
\]

(3.17)

where \(\Delta U_{ij}\) is the difference in utility, \(V^i\) is the indirect utility of the alternative on the left hand side, and prices and the level of the composite good are assumed to be constant.
When applying paired comparison, it is assumed that the change in ratings is related to the change in utility using a transformation function, as follows. This transformation function is subject to the same assumptions regarding cardinality and interpersonal comparisons of ratings scales as contingent rating.

$$\Delta r_{ij} = \phi(\Delta U_{ij})$$

(3.18)

where \(\phi\) is a transformation function

The behavioural basis of paired comparison responses is unclear. The RUM does not apply to ratings because only the information about which option is preferred is relevant. Information integration theory might provide a basis for formulating a difference model for the integration process. This would require the extra assumption that respondents can separately evaluate two alternatives, and estimate the difference between them using a cardinal ratings scale. Unfortunately, it is unclear how to define a theory of errors for this process.

When analysing paired comparison data, each observed rating is usually transformed by subtracting the middle category (e.g., five on a nine-category scale) from it, and regressing these difference scores against the matrix of differences in attribute levels, as follows:

$$R_{ij} = \beta_0 + \beta_1(X_1^i - X_1^j) + \ldots + \beta_1(X_N^i - X_N^j)$$

(3.19)

where \(R_{ij}\) is the ratings difference and \(X_N^i\) is the ‘nth’ attribute of the ith alternative

Several studies have treated the ratings scales used to compare alternatives as having cardinal significance because of the use of OLS estimation procedures (e.g., Magat et al. 1989). As discussed under contingent ratings, this assumption may not hold in practice. However, it is also possible to
use ordered modelling techniques which only require the assumption that ratings have ordinal significance (eg Johnson and Desvousges 1997).

3.6.2 Welfare estimation

Compensating surplus is estimated in paired comparisons by varying income until the difference in ratings is equal to zero. The paired comparison approach is therefore conceptually similar to the ratings difference model that was reviewed in the previous section. The main difference being that, in most cases, ratings are not differenced using the status quo as the second alternative. Therefore estimated parameters are based on trade-offs across utility space. As Johnson and Desvousges (1997, p.86) commented, this is not appropriate when using values in cost-benefit analysis where information is needed about trade-offs ‘relative to a common starting point, which generally is the status quo’. It is, however, possible for one alternative to represent the current situation and remain constant (eg Segal 1995), although it is more common for both to vary.

Respondents do not indicate a categorical preference between maintaining the status quo and choosing a new alternative in paired comparison exercises. Similar to the ratings difference approach, no information is gathered on the ‘cut-point’ for when respondents would actually choose a new alternative. Therefore trade-off data, and resulting welfare estimates, may be biased if changes in ratings cannot accurately be transformed into changes in utility.

3.6.3 Other issues

Similar to contingent ratings, problems may arise because of the non-comparability of ratings across respondents. Respondents may use ratings scales in different ways, as discussed previously. This could be expected
to increase variance and potentially cause bias in model parameters. In order to minimise problems associated with differing variance, Johnson and Desvousges (1997) used an heteroscedastic ordered probit model in which a separate scale parameter was estimated for each individual. However, allowing for differences in variance appeared to have little effect on model results. Johnson and Desvousges (1997) also estimated dichotomous choice models where ratings were transformed to categorical ‘yes/no’ choices. They found that ‘surprisingly, suppressing intensity information has little effect on overall statistical significance. Significance levels actually improve for many parameters’ (p.95). These results indicate that it may be more beneficial to gather choice rather than ratings data.

Finally, paired comparison would be expected to be more cognitively demanding than many of the other SP techniques. Respondents are faced with what amounts to a two stage decision process: they must decide not only which option they would prefer to choose, but also how strongly they feel about it. Paired comparison is thus susceptible to the use of heuristics, and possibly more so than some of the other techniques.

3.7 Conclusion

Five SP techniques have been evaluated in terms of their suitability for use in estimating non-use values in this chapter. The first technique evaluated was the CVM. The CVM is the most widely used SP technique for estimating non-use values. The dichotomous choice version of the CVM was reported to have several advantages, including that it is a similar to a referenda, it has a strong behavioural basis, and it is capable of producing welfare estimates that are consistent with consumer theory. However it also has several limitations. It is arguably prone to bias, it is relatively costly to use when valuing multiple policy options, and it has become a controversial technique.
The second technique evaluated was CM. This technique has the advantage that it does not require any cardinality assumptions for the response measure, and it is consistent with the RUM. It is also capable of producing welfare estimates that are consistent with demand theory. However, it does have limitations. The MNL model, which is most often used in CM applications, requires the assumption of independently and identically distributed error terms. In practice this assumption may be violated, necessitating modifications to the MNL model, or use of an alternative model. It should be noted, though, that this assumption may be required for the remaining conjoint techniques. CM may also be subject to the use of heuristics which, in some cases, may violate the axioms of the RUM.

The next technique evaluated was contingent ranking. Like CM it does not require any cardinality assumptions for the response measure, and it is theoretically consistent with the RUM. However, several concerns have been raised about using contingent ranking to estimate environmental values. The first is that it may not produce welfare estimates that are consistent with demand theory. While it has not yet been demonstrated, it is possible that the inclusion of irrelevant trade-off data may lead to biased welfare estimates. Secondly, empirical evidence suggests that the data collected from even the simplest of contingent ranking exercises may be inconsistent. It is possible that using the data from different ranks will result in IIA violations and biased model coefficients and, hence, implicit prices.

Contingent rating is also problematic. The traditional approach requires the assumption that response measures have cardinal significance due to the use of OLS regression, which is inappropriate. It is not possible to derive theoretically correct welfare estimates using the traditional contingent ratings approach. Recently there have been attempts to remedy some of these problems. The ratings difference model has been used to increase the comparability of responses, and has been partly successful.
However, some questions remain about the validity of the welfare estimates derived from using the ratings difference approach.

The final technique examined was paired comparison. This technique suffers from similar problems to contingent rating. It requires the comparison of ratings across respondents, and welfare estimates may not be consistent with demand theory. The behavioural basis of paired comparison is unclear.

In summary, several themes have emerged from this review. The first is that conjoint techniques potentially have several advantages over the CVM, including more discriminating responses and the capacity to value multiple alternatives in a single survey. The second is the recognition that not all of the conjoint techniques necessarily produce theoretically correct welfare estimates, although further research is needed on this issue. Third, not all of the techniques have a well defined behavioural basis, particularly the ratings difference and paired comparison approaches. Fourth, it appears that the data gathered using each of the techniques may not be of the same quality, either because of problems with interpersonal comparisons of ratings scales or the use of heuristics. While all of the conjoint techniques are prone to the use of heuristics, ratings and rankings data appear to be especially susceptible. On the basis of this review, CM appears to have most potential for producing valid welfare estimates, and estimates suitable for benefit transfer. For this reason, CM is used in this thesis instead of the CVM or any other conjoint technique to estimate non-use values. In the following chapter issues in applying CM are examined.
Chapter 4  
Methodological Issues in Choice Modelling  

4.1 Introduction  

In Chapter 3, CM was compared with the CVM, contingent ranking, contingent rating and paired comparison. On the basis of the comparison, it was concluded that CM was the most appropriate technique for use in this thesis.  

The objective of this chapter is to review a number of methodological issues that are relevant to the application of CM to estimate non-use values. This review is necessary in order to provide a background to the application of the CM technique that is detailed in the following chapters. In order to achieve this objective, a detailed investigation of the main stages in a CM application is carried out, and the key methodological issues faced by researchers at each of these stages are reviewed.  

Broadly, there are four main stages in any CM application. The first stage is the design of the questionnaire. This includes the design of the choice sets. A large number of issues relating to choice set design must be considered. Many of these can affect the success of a CM application. They include whether to use labels, the specification of a constant base alternative and the selection of attributes and attribute levels. These issues are reviewed in Section 4.2.
After the questionnaire has been developed it is necessary to select an experimental design. An experimental design is used to generate the different choice sets that are evaluated by respondents when answering the questionnaire. The choice of experimental design can affect the type of model which can be estimated, the functional form that can be used in the model, the specification tests that can be used to evaluate the model and the sample size. Therefore, experimental design is of central importance. Issues relevant to experimental design are explored in Section 4.3.

The third stage involves the selection of an appropriate statistical model. The most common model used in CM is the MNL model. The MNL model is derived by assuming that the error terms of the RUM are independently and identically distributed. The error terms may not be independently and/or identically distributed in many situations. It may then be necessary to use an alternative statistical model, such as the multinomial probit, nested logit or heterogeneous extreme value models. The appropriateness of each of these alternative models depends on the nature of the IID violation. Issues pertaining to model selection are examined in Section 4.4.

The final stage of a CM application is model evaluation. Model evaluation involves an assessment of the overall robustness of a model, and the use of specification tests such as of the applicability of the IIA property. The model evaluation stage is necessary to provide an indication of the accuracy of a final model, and to indicate whether an alternative statistical model or model structure would be more appropriate. Model evaluation is explored in Section 4.5. Conclusions are offered in Section 4.6.

4.2 Choice Set Design

A choice set contains two or more resource use alternatives (or profiles) which are defined using a common set of attributes (resource use
characteristics). As described in Chapter 3, in a CM questionnaire respondents are presented with a series of choice sets and asked to choose their preferred alternative from each set.

The design of the choice sets is one of the most important tasks in the development of a CM questionnaire. This is because the choice set is the main vehicle for conveying information about different resource use alternatives. Also, poorly designed choice sets can have undesirable properties. They can affect whether welfare estimates are consistent with economic theory. Overly complex designs can also create confusion amongst respondents and stimulate the use of inappropriate heuristics.

There are a number of decisions that need to be made when designing choice sets, all of which can potentially have substantial effects on the final results. The main decisions facing researchers when designing choice sets are reviewed in this section. These issues are pertinent to the design of the questionnaires used in this thesis, which are described in Chapter 6.

### 4.2.1 Framing

The term 'framing' is used in the CVM literature to refer to the context established for payment. A strategy to ensure an appropriate framing of an issue typically involves asking respondents to rank the importance of a number of relevant non-environmental and environmental issues which are all candidates for government funding (e.g., Bennett, Blamey, and Morrison 1997). The objective of such a framing question is to stimulate respondents to think about the opportunity costs of a decision to support a dichotomous choice referenda. While the valuation exercise is still conducted within a partial equilibrium framework, this approach encourages respondents to consider substitutes and complements.
With CM there are opportunities for providing a more complete frame than what is typically possible with the CVM. This is because of the potential to include different goods within the same choice set. This is particularly useful when dealing with recreational choice (e.g., Morey, Rowe and Watson 1993; Kaoru 1995). An advantage of this approach is that the saliency of other alternatives, and hence the opportunity cost associated with choosing any one alternative, is maintained throughout the questionnaire.

When estimating non-use values it would also be ideal to consider multiple goods. This is particularly important when there are multiple projects on the political agenda (see Bishop and Welsh 1992). Rolfe and Bennett (1996b) and Rolfe, Bennett and Louviere (1998) have investigated the possibility of including multiple goods in choice sets when estimating non-use values. They have conducted several surveys that have focused on the preservation of rainforests in Australia and overseas. This approach encourages respondents to think explicitly about substitute goods. However, it may not be appropriate to limit respondents in the number of projects that they can choose to pay to preserve. Respondents may be willing to pay for multiple public goods at a time. As a result, welfare estimates derived from such studies may be underestimates of willingness to pay. Despite these reservations, this type of application could provide useful information about the relative importance of different public goods to the community.

42 One possibility for deriving more accurate welfare estimates is to imitate the California Income Tax referenda (Bennett and Carter 1993). In these referenda, respondents are asked questions about their willingness to pay for several different projects (see also Hoehn and Loomis 1993). Another possibility is to develop modelling techniques that allow respondents to choose one or more goods within choice sets.
4.2.2 Labelled versus generic alternatives

Either labelled or generic alternatives can be included in choice sets. Labelled alternatives are specified in some way. The labels may represent different sites where multiple goods are included within a choice set (e.g., Great Barrier Reef, Snowy Mountains, Great Ocean Road etc.). Or they may represent different scenarios where a single good is being considered in a choice set, (e.g., preserve 10% of remnant vegetation, preserve 30% of remnant vegetation). Labels often provide information about a good that is different to what can be communicated through the attributes. In contrast, generic alternatives are unlabelled. They are usually titled, for example, ‘alternative 1, 2 and 3’.

Whether to use labels depends on the context of the CM exercise. The use of labelled alternatives would be appropriate if the goal of the application is to determine the proportion of the population that would support a policy proposal if a referendum were held. In this case, the use of labels is consistent with notions of predictive validity (Blamey, Rolfe, Bennett and Morrison 1997). In a referendum, such labels are likely to be used to ‘sell’ a policy; hence surveys should similarly use labels. However, the generic format may be more suitable if the primary goal of the application is the estimation of non-use values. One reason is that value estimation is constrained by the use of labels. It may be necessary to undertake a new survey if the government wanted to value a new alternative that did not correspond to any of the labels. A second reason is that the use of labels may stimulate the use of undesirable heuristics. For example, labels may act as indicators for certain attributes. The label ‘preserve 30% remnant vegetation’ may indicate to respondents that the number of non-endangered species may increase. As a result, respondents may not consider this attribute when making their choices, causing it to be insignificant. Evidence of the use of this heuristics was found in Blamey, Bennett, Louviere, Morrison and Rolfe (1998).
Another issue in designing choice sets is whether and how to include a constant base alternative. In Chapter 3 it was concluded that it is necessary to include a constant base alternative to derive theoretically correct welfare estimates. Hence a constant base alternative should be included if the objective of an application is to estimate non-use values for cost-benefit analysis.

In CM applications reported in the economics literature, a common approach has been to include an alternative which states ‘I would not choose any of these alternatives’ of ‘I would not go hunting’ etc. (eg Adamowicz, Louviere and Williams 1994; Boxall, Adamowicz, Swait, Williams and Louviere 1996; Rolfe and Bennett 1996). The attribute levels of the base alternative are, therefore, not explicitly defined. This approach is appropriate if the objective of the application is the estimation of use values, such as in the case of recreational choice, or in transport and marketing applications. However, when estimating non-use values it would seem more appropriate to describe this option as, for example, ‘Continue the current situation’ and, importantly, list the corresponding attribute levels. Unless listed, individuals are permitted to judge for themselves what levels are involved. This may affect the way that respondents make their choices. Moreover, leaving out the attributes has implications for the modelling process. The MNL model is a ‘difference in attributes’ model. If no attribute levels are defined for the constant base, the MNL model treats the levels for the base alternative as equal to zero, and then takes differences. This can affect welfare estimates.

43 It may not always be appropriate to specify the base alternative as the current situation. For example, the appropriate base alternative might be the current situation in five years time.
The careful selection of attributes to include in choice sets is critical for the development of a robust choice model. At a basic level, candidate attributes can be selected by determining which attributes are relevant to respondents and which are important from a policy perspective\textsuperscript{44}. However, the final selection requires consideration of some more subtle problems (Blamey, Rolfe, Bennett and Morrison 1997).

Some relevant environmental attributes may be perceived by respondents to be causally prior to other environmental attributes. For example, a change in water quality in a wetland may be believed by respondents to cause a change in the number of waterbirds breeding. If respondents have this belief, they may focus on the ‘indicator’ attribute and discount information about the other attributes\textsuperscript{45}. The inclusion of an ‘indicator’ attribute may make other attributes redundant. Hence it may be desirable to include either the indicator or the other attributes, but not both.

A second problem facing researchers is determining whether to disaggregate attributes. For example, is it sufficient to have as an attribute the number of species, or should this be specified as two attributes: the number of endangered species and the number of non-endangered species? Some guidance may be found through asking respondents in focus groups whether they actually distinguish, for example, between the two different types of species.

Finally it is necessary to consider respondents’ cognitive burden when selecting attributes. Respondents may seek to simplify the exercise by

\textsuperscript{44} Generally it is not possible to include attributes that are only relevant from a policy perspective and not relevant to respondents as they are unlikely to be significant explanatory variables. However, some attributes of marginal relevance to respondents, but important for policy, could be included.

\textsuperscript{45} A test of the effect of causal attributes was reported in Blamey, Bennett, Louviere, Morrison and Rolfe (1997). The authors found some support for the hypothesis that inclusion of a causal heuristic can affect respondents’ choices.
using heuristics (Mazotta and Opaluch 1995, Swait and Adamowicz 1996) or choose not to complete the questionnaire if choice sets become too difficult to answer because of the number of attributes included. According to Carson, Louviere, Anderson, Arabie, Bunch, Hensher, Johnson, Kuhfeld, Steinberg, Swait, Timmermans and Wiley (1994) the average questionnaire includes about seven attributes. In the non-use context, respondents typically know much less about environmental goods and may find a questionnaire difficult to answer if seven or more attributes are included. Hence the use of fewer attributes may be appropriate.

4.2.5 Selection of attribute levels

After the set of attributes has been selected, levels must be set for each attribute. The first decision faced by researchers when selecting attribute levels is whether they should be specified as increments or absolute amounts. For example, if employment is included as an attribute, it could be included as the total number of jobs, or alternatively, the number of jobs lost or gained. It is difficult to know which method of defining attribute levels is preferable. Whether increments or absolute amounts should be used depends on what respondents find easier to understand, whether respondents consider information about the absolute levels important and information availability.

After deciding how attribute levels should be specified, the attribute ‘space’ needs to be determined. The attribute space reflects the magnitude and range of the attribute levels. Green and Srinivasan (1978) suggested, in the context of conjoint analysis more generally, that it should be as large as possible but within the range of plausibility. Minimising the range of a particular attribute will potentially reduce the significance of a particular attribute. Meyer and Eagle (1982), in a study of shopping centre choice using student respondents, showed that reduced variability in the levels of
'major' attributes can affect the importance of less important attributes. They suggest that this occurs because insufficient variation in major attributes cause respondents to focus on less important attributes when choosing between alternatives.

Once the attribute space has been determined, it is then necessary to select the number of attribute levels to include. Attributes can have any number of levels. They can be symmetric, which means that each attribute has the same number of levels, or asymmetric. Most non-market valuation studies have included at least three levels, and both symmetric and asymmetric designs have been used. By including more attribute levels, more information about the effect of an attribute on utility can be derived. Including more than two levels would appear to be essential if attributes are expected to be non-linear, as three levels are necessary to estimate quadratic terms. The disadvantage of including more levels is that it increases the size of the experimental design. In addition, respondents may eventually ignore what they perceive to be minor differences in attribute levels if too many levels are included (Pearmain, Swanson, Kroes and Bradley 1991).

Little research has been conducted into the appropriate number of levels to include. A few marketing studies have, however, demonstrated that increasing the number of attribute levels may increase the importance of an attribute (Currim, Weinberg and Wittink 1981; Wittink, Krishnamurthi and Nutter 1982; and Steenkamp and Wittink 1994).

After selecting the attribute space and the number of levels, the final step is to assign the intermediate attribute levels. The most common approach used in the literature is to use equal increments. Pearmain et al (1991), however, recommended the use of unequal increments as it creates a larger number of potential trade-offs. Other alternatives are to select attribute levels so that they follow a logarithmic progression (eg Rolfe and Bennett 1996), or to select certain deciles from a normal distribution of a given
attribute (eg Elrod, Louviere and Davey 1992).

4.3 Experimental Design

After designing the questionnaire, the next stage in a CM application is the creation of an experimental design. An experimental design is used to create the choice sets that will be shown to respondents. It does this by systematically varying the attributes in the choice sets. For CM, experimental design is a two stage process. First the alternatives are generated, and then they are combined into choice sets. This experimental design process is used for designing the questionnaires detailed in Chapters 6 and 7.

An example of a simple experimental design is shown in Table 4.1. There are three attributes (increase in water rates, wetland area, number of waterbirds) in this example, and each attribute has two levels. Using these three attributes it is possible to derive $2^3$ or 8 alternatives. The set of eight alternatives, which includes every possible combination of attributes, is known as a full factorial.
Full factorials have several desirable properties. The first is that they are orthogonal. Orthogonality refers to multicollinearity between the attributes. There is no correlation between the attributes in a full factorial, so they are said to be completely independent or orthogonal. As a result, the use of a full factorial will allow the determination of the separate importance of each attribute. The second property relates to identification. With a full factorial design it is possible to estimate all main effects and all two way and higher interactions. A main effect refers to the contribution of a change in one of the attributes on the relevant dependent variable (eg utility). Main effects designs imply that additive functional forms will be estimated, as follows:
\[ V = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \]

where \( V \) is utility and \( X_1, X_2 \) and \( X_3 \) represent different attributes

Interactions show how the main effects are modified by one or more other variables. This is shown in the following equation. A two way interaction occurs when the magnitude of a main effect changes at different levels of a second attribute (eg \( X_1X_2 \)), and a three way interaction occurs when the magnitude of a two way interaction changes at different levels of a third attribute (eg \( X_1X_2X_3 \)), as follows:

\[ V = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1X_2 + \beta_5 X_2X_3 + \beta_6 X_1X_3 + \beta_7 X_1X_2X_3 \]

As the number of attributes and attribute levels increases, so does the number of alternatives within a full factorial. For example, if there are two attributes at two levels, and two at three levels, there would be \( 2^2 \times 3^2 \) or 72 alternatives in a full factorial. In a CM application involving non-use values it would be common to have at least six attributes, each with three levels. The full factorial for such an application would have \( 3^6 \) or 729 possible alternatives. Because of the large number of alternatives to evaluate, and the need to have multiple evaluations of each alternative, it is usually not possible to use the full factorial.

The most frequently used approach to reduce the number of alternatives is to use a fractional factorial design. A fractional factorial is a part or a fraction of the full factorial. They are derived by making certain assumptions about the existence of two way and higher interactions. Most fractional factorial designs continue to maintain the property of orthogonality, however, the identification properties of the full factorial are compromised because of the assumptions made in deriving the fractional factorial. Therefore, depending on the fractional factorial selected, not all two way or higher level interactions will be estimable. Fractional factorial designs can be found in various design catalogues, including Box and Hunter (1961), Addelman (1962a, 1962b), Hahn and Shapiro (1966) and
Dey (1985). Some statistical computer packages such as SPSS, SAS and MINITAB are also capable of generating orthogonal fractional factorials.

The number of alternatives in a fractional factorial will probably still be more than can be evaluated by any one respondent, except for the simplest of experimental designs. It is possible, however, to systematically or randomly block the alternatives in the fractional factorial into different sub-sets. Blocking is usually done by creating additional blocking variables or by random assignment. Each respondent then only has to evaluate the number of alternatives in a single block. This, however, requires the assumption that individuals allocated to each block have identical preferences. It also implies that larger sample sizes will be required as individuals cannot evaluate all alternatives.

Broadly, there are four main types of fractional factorial designs, each differing in the strength of the assumptions that are made about the existence of interactions. Each of these types of designs are said to have a different resolution, as follows (Mason, Gunst and Hess 1989):

**Resolution 3 designs**: allow estimation of main effects, but assume that all interactions are negligible. Resolution 3 designs are said to be unprotected. This means that any significant interactions could potentially be confounded with main effects.

**Resolution 4 designs**: are similar to Resolution 3 designs except that main effects are protected from any significant interactions. Hence the existence of significant interactions will not bias estimates of main effects.

**Resolution 5 designs**: it is possible to estimate main effects and some or all two way interactions in these designs. Main effects are protected from two way interactions, but not higher order interactions. Two way interactions can also be confounded with higher order interactions.

**Resolution 7 designs**: all main effects, two way and three way interactions are mutually unconfounded.
The Resolution 3 designs are the simplest fractional factorial designs to use. These designs have the fewest number of alternatives to evaluate, and hence a smaller sample size is usually required. They tend to be commonly used in commercial applications (e.g. Centre for International Economics 1997). In the context of marketing applications, Louviere (1988a, p.40) contends that the following generalisations can be made about the importance of main effects and interactions:

(a) Main effects explain the largest amount of variance in response data, often 80% or more; (b) two-way interactions account for the next largest proportion of variance, although this rarely exceeds 3%-6%; (c) three-way interactions account for even smaller proportions of variance, rarely more than 2%-3% (usually 0.5%-1%); and (d) higher-order terms account for minuscule proportions of variance.

4.3.1 Combining resource use alternatives to form choice sets

The experimental designs discussed so far can be used to generate different resource use alternatives. In CM applications, however, respondents are required to choose their preferred option from two or more options. Some method is, therefore, needed to combine the resource use alternatives in order to form choice sets.

Two main approaches are used to combine resource use alternatives. These are the sequential and simultaneous approaches (Bunch, Louviere and Anderson 1993). In the sequential approach, the alternatives that have been generated using a fractional factorial design are placed in choice sets using an assignment strategy. Examples of this approach include \(2^j\) block assignment, balanced incomplete block, foldover, shifted pairs, shifted triples and shifted quadruple designs. In the simultaneous approach, fractional factorial designs are used to create alternatives and choice sets.
simultaneously. The main simultaneous approach is the $L^{MN}$ design. Each of these techniques are now explained in turn.

(i) $2^J$ block assignment

After creating a set of $J$ alternatives using a fractional factorial design, each of the $J$ alternatives is treated as a two level attribute to determine whether they are present or absent in any particular choice set (e.g., Louviere and Woodworth 1983). A main effects fractional factorial is then used to assign the alternatives into choice sets. A feature of this type of design is that it is possible to conduct mother logit tests for violations of the independence of irrelevant alternatives property, provided that a design that allows the estimation of 2-way interactions has been used (Louviere and Woodworth 1983). A disadvantage is that the design creates choice sets of unequal size and that plans are available for only certain values of $J$. If, however, the objective of the study is to estimate availability effects (i.e., the presence or absence of a product on market share), this type of design may be appropriate.

An example of a $2^J$ design is shown in Table 4.2. There are four alternatives that need to be combined into choice sets in this example. A half fraction of a $2^4$ design could be used to assign these alternatives to choice sets, as follows. Note that in choice set 3 there is only one alternative present. This would still be a useful choice set if there is a base alternative included in each choice set, in which case the respondent still has a choice between two alternatives. However, choice set 7 would obviously need to be discarded as it contains no alternatives.
Table 4.2: A $2^{j}$ experimental design

<table>
<thead>
<tr>
<th>Choice set</th>
<th>Alternative 1</th>
<th>Alternative 2</th>
<th>Alternative 3</th>
<th>Alternative 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Present</td>
<td>present</td>
<td>absent</td>
<td>present</td>
</tr>
<tr>
<td>2</td>
<td>Present</td>
<td>present</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>3</td>
<td>Present</td>
<td>absent</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>4</td>
<td>Present</td>
<td>absent</td>
<td>present</td>
<td>present</td>
</tr>
<tr>
<td>5</td>
<td>Absent</td>
<td>present</td>
<td>absent</td>
<td>present</td>
</tr>
<tr>
<td>6</td>
<td>Absent</td>
<td>present</td>
<td>present</td>
<td>absent</td>
</tr>
<tr>
<td>7</td>
<td>Absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>8</td>
<td>Absent</td>
<td>Absent</td>
<td>present</td>
<td>present</td>
</tr>
</tbody>
</table>

(ii) Balanced incomplete block designs

A balanced incomplete block design involves, as its name suggests, *blocking* alternatives into choice sets. It is described as incomplete because it is not possible to include every possible alternative in each block (or choice set in this case). It is described as balanced because each alternative occurs in the design an equal number of times.

Balanced incomplete block designs are often characterised using the following vector of parameters $(v, b, r, k, \lambda)$, where $v$ is the number of alternatives, $b$ is the number of blocks, $k$ is the size of the blocks, $r$ is the number of choice sets that each alternative appears in, and $\lambda$ is the number of choice sets that each pair of alternatives appears together in (Bunch et al 1993).

Similar to the $2^j$ designs, only certain balanced incomplete block designs are available, and only for relatively small designs. Bunch et al (1993) reported that balanced incomplete block designs are available for the following combinations of $(v, b, r, k, \lambda)$. 
Balanced incomplete block designs can be found in Cochrane and Cox (1957) and Mason et al (1989).

(iii) Foldover designs

Each of the J resource use alternatives in a fractional or full factorial represents the first alternative in each choice set in a foldover design. The second alternative in each choice set is then found by taking its exact foldover. For example, if an alternative has levels 001, the second alternative would be 110. Or if there are three level attributes and the levels of the first alternative was 021, the levels of the second alternative would be 201.

One of the limitations of foldover designs, as well as the shifted designs, is that each attribute in each choice set is different. This increases the number of trade-offs that respondents are required to make when evaluating options. This may increase respondent burden and, potentially, the use of heuristics (Mazotta and Opaluch 1995).

(iv) Shifted pairs, triples and quadruples

Shifted pairs, triples and quadruple designs are similar to foldover designs except that modulo arithmetic is used to create the second (or third or fourth) alternative(s) in each choice set. This approach, which was developed by Bunch et al (1993), basically involves adding a constant to each of the attributes in the original factorial design.

### Table 4.3: Available balanced incomplete block designs

<table>
<thead>
<tr>
<th>v</th>
<th>b</th>
<th>r</th>
<th>k</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>14</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>80</td>
<td>15</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
To create a second set of alternatives in a choice set, shifted pairs are used. To each attribute, \( L-1 \) (or \( \text{mod} \ L \)) is added (where \( L \) is the number of levels of each attribute). So if each attribute has three levels, \( 3-1= 2 \) would be added to each of the original attributes to create the new set of alternatives. In other words, if the levels of the attributes in the original alternatives were 012, then the levels of the second set of attributes would be 201. Note that when \( L=2 \), the shifted pair design is equivalent to the foldover design.

Shifted triples (and shifted quadruples) are found by adding \( L-1 \) to the attributes of the second (third) alternative in a shifted pair (triple) design.

(v) \( L^{MN} \) designs

In an \( L^{MN} \) design, each attribute from each alternative in a choice set is given a column within a fractional or full factorial design. Hence, if there are two alternatives with two attributes, a design with four columns would be required. Each row within the factorial design represents a choice set. This can be seen in Table 4.4, which shows an \( L^{MN} \) design where there are two alternatives in each choice set, each with two attributes at two levels. The first two columns represent the attributes of the first alternative, and the third and fourth columns represent the attributes of the second alternative.
Table 4.4: An $L^{MN}$ design

<table>
<thead>
<tr>
<th>Choice set</th>
<th>Alternative 1 Attribute 1</th>
<th>Attribute 2</th>
<th>Alternative 2 Attribute 1</th>
<th>Attribute 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

In $L^{MN}$ designs, ‘$M$’ refers to the number of alternatives in each choice set, ‘$N$’ is the number of attributes, and ‘$L$’ is the number of levels of each attribute. The title is derived from the fractional or full factorial which is required for this design. In the above example, with two alternatives, two attributes and two levels, $L^{MN} = 2^2 = 2^4$. Hence a $2^4$ fractional factorial is needed to create this $L^{MN}$ design. It is also possible to have asymmetric $L^{MN}$ designs. If there were two attributes at two levels and two at three levels, and two alternatives per choice set, a $3^2 \times 2^2 = 3^4 \times 2^4$ design could be used.

One of the advantages of $L^{MN}$ designs is that it is possible to estimate ‘cross-effects’ in MNL models (Lazari and Anderson 1994). Cross effects occur when the attributes in one alternative affect the probability that a second alternative will be chosen. These cross effects should be equal to zero if the IIA property holds. This is because the utility of each alternative should only depend on the level of its own attributes and not on the presence or absence of other alternatives. The capacity to estimate cross effects, therefore, permits testing of the IIA property.
Given that there are five different approaches that can be used to develop experimental designs for choice, on what basis can a researcher choose between them? There are six main criteria that can be used to make this choice. These are discussed below.

(i) Identification

The first factor to consider when choosing an experimental design is the type of model that is going to be estimated and the expected model structure. The following generalisations can be made about what the different designs can be used to estimate:

- main effects can be estimated using any of the designs;
- attribute interactions can be estimated using all of the designs (except for shifted or foldover designs) as long as an appropriate fractional factorial has been selected;
- quadratic and cubic terms can be estimated using all of the designs except for shifted or foldover designs; and
- cross-effects can only be estimated using $L^M N$ designs.

(ii) Power

When selecting an experimental design it is necessary to ensure that the design has sufficient power to allow model estimation. Power refers to the number of degrees of freedom that are left to estimate error terms after allowance has been made for all of the variables in a model. Many commercial conjoint studies have insufficient power to estimate error terms. Wittink and Cattin (1989) reported that the average commercial conjoint study uses an experimental design that contains 16 alternatives, and 8 attributes that each have 3 levels. To estimate a model based on this average design $(3-1)\times8 = 16$ degrees of freedom would be required. Hence, on average, there are no degrees of freedom left to estimate error.
The available degrees of freedom can be calculated by subtracting the following from the total number of alternatives evaluated by respondents:

- \( L-1 \) for each attribute, where \( L \) is the number of levels in an attribute;
- \((L_1-1)(L_2-1)\) for each attribute interaction;
- \((L_1-1)(L_1-1)\) for each quadratic term; and
- \( L-1 \) for each cross effect.

(iii) Efficiency

Another consideration when selecting an experimental design relates to design efficiency. Design efficiency reflects the information about attribute trade-offs that can be derived from an experimental design. More efficient designs provide a greater amount of information about these trade-offs, and will \textit{ceteris paribus} lead to more precise estimates of model coefficients.

The efficiency of experimental designs for choice models is different from experimental designs for linear models (Carson et al 1995). This is because choice models are based on utility differences. Hence, design efficiency for a choice model is higher when there are few zero attribute differences between alternatives in an experimental design. For example, a foldover design where there are fewer zero differences is more efficient than an \( L^{MN} \) design, as shown in Table 4.5. Notice that all of the attribute differences are different from zero with the foldover design.
Table 4.5: Attribute differences in a foldover and an $L^{MN}$ design

### Foldover design

<table>
<thead>
<tr>
<th>Choice Set</th>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### $L^{MN}$ design

<table>
<thead>
<tr>
<th>Choice Set</th>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute 1</th>
<th>Attribute 2</th>
<th>Attribute differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Exact measures of the efficiency of an experimental design ($X$) can be derived using the variance of the Fisher information matrix ($X'X)^{-1}$. Various measures of design efficiency have been used, including 'A-efficiency', 'D-efficiency' and 'G-efficiency' (Kuhfeld, Tobias and Garratt 1994). All of these measures focus on the variance of the information matrix or the size of the eigenvalues. These measures can be used to test the relative efficiency of candidate designs. Bunch et al (1993) reported preliminary research on the relative efficiency of different generic main effects designs.

(iv) **Cognitive burden**

Cognitive burden is a significant issue in SP studies involving the estimation of non-use values. Frequently, respondents have only a limited knowledge about the environmental goods that are evaluated, and hence
they can quickly become fatigued and resort to the use of heuristics as a means of simplifying the answering of choice sets (Mazotta and Opaluch 1995).

One way to reduce respondent fatigue is to minimise the number of trade-offs they are required to make in each choice set. While this will reduce efficiency, it may not reduce variance if respondents make choices that more accurately reflect their preferences. Designs such as the $2^J$, $L^{MN}$ and balanced incomplete block designs which do not have trade-offs at every attribute would be favoured from this perspective.

(v) Cost

Cost is a concern in most studies when selecting an experimental design. Smaller designs are cheaper because fewer surveys are required. However, there is usually a trade-off between reduced cost and greater inaccuracy. If there is little a priori information about a model's functional form it may be necessary initially to use a large experimental design which allows the estimation of a greater number of model parameters. For subsequent models, when there is knowledge of which model parameters are insignificant, there can be greater confidence in using simpler and cheaper experimental designs.

(vi) Availability

The final consideration when selecting an experimental design is a pragmatic one. Often there may only be a few designs available if large numbers of attributes and attribute levels are to be used. Because of limits on the availability of designs, it may be necessary to compromise on one of the above criteria.
One further issue that is worthy of investigation is how to deal with dominated or implausible alternatives. A dominated alternative occurs when all of the attributes of an alternative are worse than all of the attributes of a second alternative in the same choice set. For example, Alternative 1 dominates Alternative 2 in Table 4.6. It has a higher level of environmental quality, and it is cheaper. Carson et al (1995, p.356) warned against using designs that include dominated alternatives because ‘the respondent choices do not reveal information about trade-offs between the levels of different attributes’.

Table 4.6: A choice set with a dominated alternative

<table>
<thead>
<tr>
<th></th>
<th>Alternative 1</th>
<th>Alternative 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wetland area</td>
<td>1000 km²</td>
<td>500 km²</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>2000</td>
<td>500</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>1000</td>
<td>200</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$20</td>
<td>$50</td>
</tr>
</tbody>
</table>

An implausible alternative is one that makes no sense. Consider the following example. In Alternative 1 there are 1000 waterbirds breeding but only 500 waterbirds in the wetland.

Table 4.7: A choice set with an implausible alternative

<table>
<thead>
<tr>
<th></th>
<th>Alternative 1</th>
<th>Alternative 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wetland area</td>
<td>1000 km²</td>
<td>500 km²</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>500</td>
<td>2000</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>1000</td>
<td>200</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$20</td>
<td>$50</td>
</tr>
</tbody>
</table>

There are several methods of dealing with dominated and implausible alternatives. In some cases it may be possible simply to ignore the problem, especially if there is uncertainty about what is a dominated or
implausible alternative (eg Centre for International Economics 1997). It may often be the case that alternatives found by a researcher to be dominated or implausible are not viewed in the same way by respondents.

One option for reducing implausibility is to include within a questionnaire an explanation of why seemingly implausible options are actually possible. It may also be possible to include explanations for the inclusion of dominated alternatives.

Another option is to delete dominated or implausible options (Pearmain et al 1991). This may be necessary if implausibility is believed to be a more pervasive problem in the study, and it is not possible to explain why these options have been included. However, deleting options reduces the orthogonality of an experimental design, and may affect the identifiability of model parameters.

The two attributes can be combined into a composite attribute if implausibility results from a correlation between two attributes (Green and Srinivasan 1978). For example, if the two attributes are waterbird numbers and breeding, the three levels of the new attribute could be as follows:

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500 waterbirds, 200 breeding</td>
</tr>
<tr>
<td>2</td>
<td>1000 waterbirds, 500 breeding</td>
</tr>
<tr>
<td>3</td>
<td>1500 waterbirds, 1000 breeding</td>
</tr>
</tbody>
</table>

One of the limitations of combining two attributes into a single attribute in this way is that it is not possible to estimate the separate value of an increase in each of the original attributes. It is only possible to estimate the value of a change from Level 1 to Level 2 etc. A further modification, however, can enable estimation of separate values. This involves using composite attributes, but with a larger number of levels, as follows. Both attributes can then be modelled separately and separate values found for each attribute. Care needs to be taken to maintain the orthogonality of the
experimental design when using this approach.

<table>
<thead>
<tr>
<th>Level</th>
<th>Waterbirds</th>
<th>Breeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>1000</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>1500</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>1500</td>
<td>500</td>
</tr>
<tr>
<td>6</td>
<td>1500</td>
<td>1000</td>
</tr>
</tbody>
</table>

4.4 Model Selection

After the questionnaire has been designed and completed by respondents, the process of model selection begins. There are a number of models available to researchers. Each of these models results from different assumptions about the distribution of the random components of the RUM; specifically whether they are independently and/or identically distributed. The choice of model, therefore, depends primarily on the error structure of the alternatives being evaluated.

If the error terms are IID, then it is appropriate to use the MNL model which is described in Chapter 3. However, if violations of this assumption occur, it may be necessary to use an alternative model which allows for non-IID error terms, such as the multinomial probit, nested logit or heteroscedastic extreme value models. The characteristics of each of these other models are now reviewed in turn. In Chapter 7 (and Appendix 5) attempts are made to use these models to overcome IIA violations.

4.4.1 Multinomial probit

The multinomial probit model is derived by assuming that the errors terms of the RUM have a multivariate normal distribution. The use of this
distribution means that the assumption of IID error terms (and hence the IIA property) is not required, and the error terms of different alternatives can be correlated or have unequal variance. The model can allow for taste variation across respondents as well as covariance between alternatives (Hensher and Johnson 1981; Ben-Akiva and Lerman 1985). This may be useful if respondents have quite different preferences for a good, or if some alternatives are substitutes.

Despite these advantages, the multinomial probit has been infrequently used. Examples of its use include Hausman and Wise (1978), Miller and Lerman (1981) and Chintagunta (1992). The more general error structure creates significant computational difficulties as well as potential inaccuracies. Hensher and Johnson (1981) suggested that a nested logit, or a MNL model that includes socioeconomic interactions to allow for taste variation, is likely to be more practical than the multinomial probit. However, Carson et al (1994) have suggested that computational improvements are likely to lead to an increase in use of the multinomial probit model.

### 4.4.2 Nested logit

Nested logit models have frequently been used to analyse choice of recreational sites (Morey et al 1993; Kaoru 1995), transport modes (Forinash and Koppleman 1993) and consumer goods (Kamakura, Kim and Lee 1996). The formal specification of the nested logit model is reported in Greene (1993).

Unobserved components of utility are assumed to be shared between certain alternatives in nested models; hence their error structures are correlated and not independent (Ben-Akiva 1973). In other words, the utility of alternatives A, B and C can be described as follows:
Note the common error term $\varepsilon_{AB}$. Because of this common error term, the errors for alternatives A and B will be correlated, and distributed differently to alternative C.

A ‘tree-structure’ needs to be pre-specified to use a nested logit model. Tree structures reflect the existence of homogenous sets of alternatives which have correlated error structures. They can have multiple levels. An example of a simple tree structure for transport choice that has two levels is depicted in Figure 4.1. All of the alternatives are shown in the branches at the bottom of the structure (Plane, Train, Car, Bus). These alternatives are then grouped at the next level using the limbs of the tree (ie Air, Ground). Individuals are assumed to choose an alternative at the lowest levels of a tree; hence they indicate on which limb they are located.

**Figure 4.1: A tree structure for transport choice**

While the nested logit model is relatively simple to estimate, and is likely to be useful in a number of circumstances, it does have limitations. Bhat (1994) suggested the following three problems. First, the number of
structures to estimate when searching for the best structure increases rapidly as the number of alternatives increase. Second, the alternatives may be difficult to partition into subsets if the differences between alternatives can be represented by a continuum. Third, different respondents may perceive different structures, and no single structure may be appropriate for all respondents.

4.4.3 The heteroscedastic extreme value model

The heteroscedastic extreme value model can be used to evaluate alternatives when the error terms are independent but not identically distributed (eg Allenby and Ginter 1995, Swait and Adamowicz 1996). This model could be used if the variance of unobserved or observed characteristics differs across alternatives. For example, in a transport application, if the unobserved characteristic ‘comfort’ varies greatly for automobiles but only slightly for trains, then the two different modes may have different variances (Bhat 1995). Or in the case of consumer goods, the error terms associated with the effect of price increases may differ across goods resulting in different variances (Dellaert, Brazell and Louviere 1997).

Formally, the heteroscedastic extreme value model can be described as follows (see Dellaert et al 1997):

\[
P(j|X) = \frac{e^{\lambda(p)X}}{\sum e^{\lambda(p)X}}
\]

where the scale parameter, \( \lambda \), is a function of the parameter vector \( p \) which can describe product, context and individual characteristics.

A feature of this model is that variables can be entered into the vector of independent variables \( X \) which affects mean utility, or the vector \( p \) which
affects variance—and, therefore, affects utility indirectly—or both (Swait and Adamowicz 1996). For example, it is possible to enter a price variable in both $X$ and $p$. Price would then have a direct and an indirect effect on utility and the net effect of a price change on utility would depend on the direction and magnitude of each of the separate effects. A price decrease would typically have a direct positive effect on utility, but could also have a negative effect if the decrease makes products more similar, the choice more difficult, and hence increases variance. The capacity to model how variables affects both mean utility and variance enables a much wider range of hypothesis tests than is possible using the MNL model.

The heteroscedastic extreme value model is simpler to compute than the multinomial probit model. Compared to the MNL model, only $J-1$ extra parameters (where $J$ is the number of alternatives) must be computed in the variance-covariance matrix, in contrast with $[J*(J-1)/2]-1$ extra parameters with the multinomial probit model (Bhat 1995).

Before moving onto model evaluation, one general problem with model selection should be noted. As detailed in the next section, tests are available to determine the existence of IIA violations. Yet these tests do not reveal the cause of IIA violations. Hence, there is often uncertainty about which of the above three models is the most appropriate. Each of these models, therefore, represent ways of dealing with a problem with a largely unknown cause. The major cause of IIA violations, though, is believed to be unequal variance of the error terms. Hence in most cases the nested logit or heterogeneous extreme value model are likely to provide adequate solutions to IIA violations (Louviere, pers.comm.).

4.5 **Model Evaluation**

The capacity to evaluate models is of fundamental importance when estimating choice models, and any statistical model. Models must be
evaluated so that the best fit for the data can be found. Model evaluation is necessary for determining whether a model has been misspecified or whether an inadequate or poor quality data set has been collected. Researchers need a knowledge of a model's robustness so that they can determine what level of confidence can be had in the results.

Researchers typically go through a two step process when evaluating a statistical model. The first step is to examine the summary statistics to get an overall picture of the performance of the model. This includes looking at the significance of individual model coefficients (particularly those with theoretical relevance), and summary statistics that indicate the overall explanatory power and significance of the model. The second step involves determining whether the model is correctly specified. Correct model specification requires the inclusion of all relevant variables and exclusion of irrelevant variables. It also requires that the assumptions behind a specific model (e.g., IID error terms) hold. If they don't, then a different model in which different assumptions are made may be required.

The summary statistics that are relevant to the MNL model will be briefly reviewed in the next sub-section. Specification problems that often affect the MNL model, and tests that can be used to identify these problems, are then reported. These statistics and specification tests are used in Chapter 7 to evaluate the choice models estimated in this thesis.

### 4.5.1 Summary statistics

Most statistical packages provide a number of statistics that are relevant for model evaluation. Asymptotic t-statistics and corresponding probability values are provided to indicate whether individual coefficients are statistically different from zero. Most packages also provide estimates of the final log-likelihood, the log-likelihood when there are no
coefficients in the model, and the log-likelihood when only constants are
included. The log-likelihood indicates the probability, or likelihood, that a
set of coefficients would have generated an observed sample. When
undertaking maximum-likelihood estimation, algorithms attempt to
maximise the log-likelihood; hence larger log-likelihoods represent models
with better explanatory power.

The log-likelihood statistics can be used to test the overall significance of a
model. This test, which is conceptually similar to the F-test in ordinary
regression, is known as the likelihood ratio test. It involves comparing the
log-likelihoods of restricted and unrestricted models, and can be used more
generally for hypothesis testing. To test the overall significance of a
model, the final log-likelihood is the unrestricted value (L_u), and the log-
likelihood with no coefficients is the unrestricted model (L_R)^{46}. The test
statistic for the null hypothesis that all of the coefficients are equal to zero
is as follows:

\[ -2(L_R - L_u) \sim \chi^2 \text{ with } r \text{ degrees of freedom} \]

where \( r \) is the number of restrictions on the parameters made when computing
\( L_R \).

The log-likelihood statistics are also used to provide an indicator of the
goodness-of-fit of the model. Various measures of goodness of fit have
been proposed in the literature. They are conceptually similar to the r-
squared in normal regression. Cragg and Uhler (1970) proposed a measure
which is known as the Cragg-Uhler R^2 or sometimes as the pseudo R^2. It
is calculated as follows:
Cragg-Uhler $R^2 = \frac{L_u^{2/n} - L_r^{2/n}}{1 - L_r^{2/n}}$

Where $L_u$ is the unrestricted likelihood, $L_r$ is the restricted likelihood and $n$ is the number of observations

Another measure of goodness of fit was proposed by McFadden (1974). This measure is known as McFadden’s $R^2$ or also as pseudo-$R^2$ (or $\rho^2$). The McFadden $R^2$ is equal to one minus the ratio of the final log-likelihood and the log-likelihood with no coefficients, as shown in the following equation:

$$\text{McFadden } R^2 = 1 - \frac{\ln(L_u)}{\ln(L_r)}$$

Where $\ln(L_u)$ is the unrestricted log likelihood, $\ln(L_r)$ is the restricted log likelihood

The adjusted $R^2$ includes a correction factor to allow for the addition of extra coefficients. Hensher and Johnson (1981, p.51) suggested that ‘values of rho squared of between 0.2 to 0.4 are considered extremely good fits so that the analyst should not be looking for values in excess of 0.9 as is often the case when using $R^2$ in ordinary regression.’

Another measure of goodness of fit that is sometimes used is the percentage correct predictions (Maddalla 1983). This measure indicates how well a model predicts the data. It is useful an ordinal measure for comparing different models. However, the indicator should be interpreted with caution. This is because high values do not necessarily indicate a robust model; it is possible that high values will be derived simply because of the nature of the data.

---

Note 46 The log-likelihood when the constant is included can also be used for the restricted model. Ben-Akiva and Lerman (1985) suggested that this is a more informative test of
4.5.2 Specification errors

Models may still be prone to specification errors even though they appear to be robust in terms of the summary statistics. There are five main specification errors that can affect MNL models (Horowitz 1981). Depending on the structure of the model, they can result in either biased or inefficient coefficients.

(i) Inclusion of irrelevant variables

The first specification error is the inclusion of an irrelevant variable. If a linear in parameters functional form is used, inclusion of irrelevant variables does not bias parameter estimates, but can reduce efficiency. If a non-linear functional form is used, the parameters may also be biased in certain circumstances (Horowitz 1981).

(ii) The existence of non-IID error terms

Violations of the IID assumption occur for two main reasons (Hensher and Johnson 1981). First there may be correlations between the unobserved components of utility of the alternatives, causing the error terms to be correlated. Second, observed and unobserved attributes may not be independent, causing differences in the distribution of the error terms of the alternatives. Violations of the IID assumption should result in biased parameter estimates.\(^{47}\)

(iii) Random taste variations

Simple MNL models assume that respondents preferences are

significance, because the null-hypothesis that all coefficients are equal to zero is almost always rejected.\(^{47}\) Horowitz (1981b), however, found that even when the variances of the error terms of different alternatives differed by a factor of 100 and the error terms had correlations of 0.99 the parameter estimates were consistent, apart from an arbitrary scale factor.
homogeneous and, hence, model parameters are assumed to be fixed across individuals. Heterogeneity can be allowed for by including socioeconomic and other interactions. However, if tastes are not fixed, and no allowance is made for heterogeneity, then the model will be specified incorrectly (McFadden, Tye and Train 1977). This is because the variables where there is a taste variation (i.e., heterogeneous preferences) are specified as fixed, whereas they are subject to a distribution. Hence model predictions may be incorrect. Random taste variations should, in principle, result in biased parameter estimates and cause non-IID error terms.

(iv) Omission of relevant explanatory variables

Unlike standard regression analysis, the omission of a relevant explanatory variable does not necessarily cause a specification error. This is because the alternative specific constants in a MNL model are expected to capture the mean values of any unobserved relevant variables (Ben-Akiva and Lerman 1985). Whether omission of a relevant variable causes a specification problem depends on the relationship between the omitted relevant and included variables. Horowitz (1981b) reported that the model coefficients will be biased unless:

- the omitted variable is distributed independently of the included variables;
- either the omitted variable must have equal mean values in all alternatives, or else alternative specific constant terms must be included in the utility function to represent the effects of the alternative-specific means of the omitted variable;
- either the omitted variable must be IID across alternatives, or else a model that permits non-IID random utility components must be used;
- the omitted variable must have the same distribution in the population for which the forecasts are made as in the population from which the values of the model's parameters are estimated; and
- the omission of the relevant variable must not substantially alter the parametric form of the random component of the utility function.

---

Horowitz (1981b) also reported that numerical experiments with a linear in parameters MNL model found that random taste variations result in consistent parameter estimates.
The error terms are correlated with the explanatory variables

This problem may occur because of the omission of relevant variables (see Horowitz's (1981) first point above). The existence of this problem should lead to biased parameter estimates.

4.5.3 Specification tests

The final step in model evaluation involves conducting specification tests to determine the adequacy of model structure. These tests can be used to diagnose the problems identified above.

The most powerful tests of the MNL model are for compliance with the IIA property. This is because IIA violations result from most of the specification problems listed above. Therefore, if there are no IIA violations one can be reasonably sure that a model is robust.

Two main approaches for testing for violations of the IIA property have been suggested in the literature. The first approach involves estimation of a mother (also called universal) logit model (McFadden et al 1977, Louviere and Woodworth 1983). The probability of a particular alternative being chosen is a function not only of the attributes of that alternative, but of competing alternatives in the mother logit model. The influence of the attributes of one alternative on the choice of a second alternative are known as cross-effects. The existence of significant cross-effects implies that the utility of each alternative in a choice set is not independent of the utility of all the other alternatives. Hence there is a violation of the IIA property. This test is usually conducted by estimating mother logit and MNL models and testing the null hypothesis that the MNL model is the true model using a likelihood ratio test (McFadden et al 1977).
The second main approach involves a comparison of the coefficients of a full MNL model with a restricted model from which one alternative has been removed. If the IIA property holds—so that the probability of choosing one alternative over a second alternative is independent of the attributes of a third alternative—then consistent parameter estimates should be found in the full and restricted models. A limitation of this test is that it requires that the models not include alternative specific constants. This is because the alternative specific constant always has a value of zero when an alternative is deleted, thus causing a Hessian singularity. The problem with deleting alternative specific constants is that they can assist in dealing with IIA violations (Train 1986). Various statistics have been developed for this test. Hausman and McFadden (1984) developed the following test statistic for the null hypothesis $\hat{\beta}_F = \hat{\beta}_R$, where $\hat{\beta}_F$ is the vector of coefficients for the full model, $\hat{\beta}_R$ is the vector of coefficients for the restricted model, and $\Sigma_{\hat{\beta}_F}, \Sigma_{\hat{\beta}_R}$ are the respective variance-covariance matrices:

$$HM = (\hat{\beta}_R - \hat{\beta}_F)' (\Sigma_{\hat{\beta}_F} - \Sigma_{\hat{\beta}_R})^{-1} (\hat{\beta}_R - \hat{\beta}_F)$$

which is asymptotically $\chi^2$ distributed with $K$ degrees of freedom, where $K$ is the number of coefficients in $\hat{\beta}_R$.

Small and Hsiao (1982) have developed an alternative test statistic that involves the calculation of log-likelihood ratios$^{49}$. This statistic is as follows:

---

$^{49}$ An example of the use of this statistic is provided in Ben-Akiva and Lerman (1985).
\[
    SH = \frac{1}{1 - N_1/(\alpha N)} \left\{ -2 \left[ L_R(\hat{\beta}_R) - L_R(\hat{\beta}_R') \right] \right\}
\]

where \( N \) is the number of observations in the unrestricted model, \( N_1 \) is the number of observations in the restricted model, and \( \alpha \) is a scalar.

This statistic is asymptotically chi-squared distributed with \( K \) degrees of freedom, if it can be assumed that \( \sum_{\hat{\beta}_y}, \sum_{\hat{\beta}_R} \) differ by a scalar \( \alpha \). In practice \( \alpha \) can be assumed to be equal to 1, and this test treated as a screening test. A more rigorous, but more complex, test statistic derived by Small and Hsiao (1982) can also be used if IIA violations are found.

4.6 Conclusion

Methodological issues related to applying CM have been considered in this chapter to provide background information for the application of the CM technique in following chapters.

The first issue was the design of choice sets. While choice sets are only one part of a questionnaire, they are one of the most critical. There are many decisions for researchers to make, including whether to use labels, whether to include a constant base alternative and which attributes to include. Some of these decisions are fairly clear cut when estimating non-use values (eg whether to include a constant base alternative). However, many of the decisions are context specific. Nonetheless, several recommendations can be made for the design of choice sets in this thesis. These are that:

- generic choice formats are likely to be more suitable for the estimation of non-use values than labelled formats;
- constant base alternatives should be included in each choice set;
- the attributes of the constant base alternative should be specified when estimating non-use values for cost-benefit analysis;
• the magnitude and range of attribute levels should be as large as possible, but within the range of plausibility.

The next issue considered was experimental design. A number of different ways of developing experimental designs were discussed. Each of the designs have advantages and limitations. Six criteria were suggested which provide some guidance in selecting designs, including identification, power, efficiency, cognitive burden, cost and availability. However, the criteria are generally conflicting; hence, the experimental design process typically involves compromise and must be undertaken on a case by case basis. The main implication from the discussion of experimental design is that researchers must decide which criteria are most important to them as it is generally not possible to find an ideal design.

The third issue was model selection. Three alternative models to the simple MNL were reviewed that relax part or all of the IIA assumption. These models have been reviewed as they may be required if IIA violations are detected in the MNL models reported later in this thesis. One of the models reviewed, the multinomial probit, is considered to be computationally burdensome and hence does not appear to be a suitable alternative if IIA violations are found. The remaining two models—the nested logit and heterogeneous extreme value logit—may provide a practical means of dealing with IIA violations. However, knowing which model is most appropriate for dealing with IIA violations is likely to remain elusive until better methods of determining the source of IIA violations are devised.

The final issue considered was model evaluation. Various summary statistics were reported, many of which are similar to what are used in ordinary regression. These statistics can be used to get an indication of the explanatory power of the model. The specification errors experienced by the MNL are, however, frequently different from ordinary regression. Hence it is not possible to interpret the effect of specification errors in the
same way as ordinary regression. For example, the omission of relevant explanatory variables need not lead to a specification error in MNL models, while random taste variations may lead to specification errors. Specification tests were also reported. It was concluded that IIA tests were relatively powerful as they can detect a number of specification errors. The mother logit test for IIA violations was recommended as it does not require the deletion of alternative specific constants.
Part 2  The Choice Modelling Surveys

The objective of the second part of this thesis is to describe the development of and the results from three separate CM surveys and to test the first two hypotheses. The data from these three surveys are used in the third and final part of the thesis to test several benefit transfer hypotheses.

In Chapter 5, the first chapter in Part 2, the two case study sites used in the surveys—the Macquarie Marshes and the Gwydir Wetlands—are described. Both of these sites are major wetlands in New South Wales that have been substantially degraded because of reduced instream flows resulting from the development of irrigated agriculture. Options are available to the New South Wales Government to improve the quality and size of both of these wetlands. Determining the value of these options to the wider community is the objective of the surveys described in Chapters 6 and 7.

In Chapter 6 the development of the questionnaires is described. The results from eight focus groups and a pretest that were undertaken to develop the questionnaires for both sites are presented. Issues discussed in this chapter include choice set design, selection of information for the questionnaires and minimisation of the use of heuristics. The final design of the questionnaires used in the surveys is also reviewed. The actual questionnaires used in the focus groups and in the surveys are presented in Appendices 1 to 3. Further information about the testing of alternative choice set designs is reported in Appendix 4.
In Chapter 7 the results from the three separate surveys are presented. The first survey, undertaken in Sydney, and the second survey, undertaken in Moree, both focused on the Gwydir Wetlands case study. The third survey, undertaken in Sydney, involved the Macquarie Marshes case study. Several different models are estimated using the data from each survey. The first two hypotheses in the thesis are tested at the end of Chapter 7. Further information on specification tests and results from use of alternative models are reported in Appendix 5.

In Chapter 8 empirical issues associated with aggregation are considered. Two types of approaches for dealing with divergences between sample and population characteristics when aggregating individual level estimates are considered. The first approach involves adjusting mean sample estimates based on the characteristics of the populations, while the second approach involves making assumptions about the preferences of non-respondents. Both of these approaches are tested in this chapter to demonstrate how policy relevant results might be obtained, and what factors are most likely to influence the results during aggregation.
Chapter 5 Description of Case Study Sites

5.1 Introduction

Large scale irrigation began in the Macquarie Valley after the construction of Burrendong Dam in 1967 and in the Gwydir Valley after the construction of Copeton Dam in 1976. Irrigated agriculture is one of the largest components of total agricultural production in both valleys.

A consequence of increasing use of water for irrigation is a reduction in the amount of water reaching wetlands in the Macquarie and Gwydir Valleys. The Macquarie and Gwydir Valleys contain extensive wetlands that are recognised for their national and even international significance (Department of Water Resources 1991, 1993). These wetlands and their biota have been substantially affected by changes to the flow regime caused by regulation and increased use of water for irrigation (Keyte 1994, Kingsford and Thomas 1995).

There are, therefore, trade-offs between the use of water for consumptive and instream uses. A relevant question for decision makers is whether it would be more beneficial to the community to allocate extra water to irrigation or to allocate extra water to the Macquarie Marshes and Gwydir Wetlands. The benefits of allocating extra water to irrigation can be found by estimating the change in surplus resulting from increased agricultural production. For wetlands, the benefits of extra water can be found by estimating changes in use and non-use values.

40 The case study sites reviewed in this Chapter are described in greater detail in Morrison and Bennett (1997).
As discussed in Chapter 3, it is necessary to use SP techniques to estimate non-use values. Therefore, these two sites provide a suitable context for testing whether it is possible to use CM to estimate non-use values. If CM is found to be capable of producing valid estimates of non-use values, these estimates could then be used to help in determining which water allocation options are likely to provide the greatest net benefits to the community.

In this chapter, an overview of the irrigated agricultural industries and wetlands in the Macquarie and Gwydir Valleys is provided. The past and present water use in the two valleys is then explained. The resulting impacts of reduced instream flow on the Macquarie Marshes and Gwydir Wetlands are detailed.

5.2 The Macquarie Valley

5.2.1 Site description

The Macquarie Valley is part of the Murray-Darling Basin. It starts at the Great Dividing Range near Bathurst and spans across the western plains to Bourke (See Figure 5.1). The Macquarie Marshes, which are the main wetlands in the Macquarie Valley, are about 50 km north of Warren. Burrendong Dam, which is upstream from Wellington, is the largest dam in the Macquarie Valley. It has a storage capacity of 1,189,000 ML and an average yearly flow of 475,000 ML.
The total revenue from agricultural production in the Macquarie Valley in 1993 was $642 million, including $292 million in crop production (Department of Land and Water Conservation forthcoming). The area of crops and pastures was 639,321 hectares (ha), of which 15% was irrigated. A total of 10,188 people are employed in agriculture, forestry and hunting in the Macquarie Valley (Australian Bureau of Statistics 1995). Cotton production generates the highest total revenue of the irrigated crops in the Macquarie Valley ($100m in 1993, DLWC forthcoming). Other irrigated crops include cereals, oilseed, citrus, lucerne, pasture, vegetables and vines.
Water allocations in the Macquarie Valley were originally made on the basis of the area that was licensed for irrigation. Licences to extract water from regulated streams could be received upon application to the Department of Water Resources. The allocation of water to irrigation is about 600,000 ML, which is approximately 89% of total allocations (DWR 1991).

A volumetric allocation scheme was introduced in the Macquarie Valley in 1981 because of shortages in the supply of water for irrigation. Under the scheme, irrigators were given an annual allocation of water according to their licensed area. In those years when the amount of water available is insufficient to supply irrigators with 100% of their annual allocations, water is allocated in proportion to available water. Irrigators in the Macquarie Valley receive their full allocation in about 65% of years (DWR 1991).

5.2.3 The Macquarie Marshes

The Macquarie Marshes are the main wetland areas in the Macquarie Valley. They consist of semi-permanent wetland, slightly elevated lower floodplain and more elevated higher floodplain areas. The area of the Marshes was approximately 130,000 ha during flooding in 1990. Within the Macquarie Marshes there is a Nature Reserve covering 18,150 ha. The Nature Reserve is listed as a Wetland of International Importance under the Ramsar Convention.

The Macquarie Marshes are known for their waterbirds. Kingsford and Thomas (1995) estimated that between 10,000 and 300,000 waterbirds use the Macquarie Marshes every October, depending on the extent of flooding. Seven waterbirds and eight other birds listed as endangered by the NSW National Parks and Wildlife Service, as well as other waterbirds
protected under international migratory agreements, have been sighted in
the Macquarie Marshes (Kingsford and Thomas 1995).

The Macquarie Marshes have been used for grazing cattle and sheep since
the mid 1800s. Cunningham (1996) estimated that the value of grazing at
the Macquarie Marshes is about $5.2-7.4 million p.a. in terms of gross
margins.

### 5.2.4 Impact of regulation and consumptive use of water
on the Macquarie Marshes

Increased regulation and consumptive use of water has affected the
environment of the Macquarie Marshes. The effects on different indicators
of wetland quality and quantity are as follows (NPWS and DLWC 1995).

(i) Flashing

Kingsford and Thomas (1995) reported that the area affected by large
floods has contracted by 40-50% between 1944 and 1993. This is
primarily because of reduced flows into the Marshes, although erosion of
channels has increased the amount of flow required for flooding. The
frequency of high and medium floods reaching the Marshes has decreased
(Brereton 1994) and the proportion of low flows has sharply increased
(NSW EPA 1995). The natural flow regime consisted of high winter and
low summer flows, but this has been reversed.

(ii) Vegetation

Changes in the flooding regime has caused declines in wetland vegetation.
Brander (1987) estimated that, in a part of the southern Marshes, the area
of river red gums declined by 55% between 1934 and 1981. Brereton
(1994) reported that water couch, which is important feed for stock and waterbirds, decreased by 40% in the Monkeygar and Bulgeraga Creek area between 1949 and 1991. Exotic species have affected large areas in the Marshes.

(iii) Waterbirds

Kingsford and Thomas (1994) found a significant decline in the number of species and the density of waterbirds between 1983 and 1993. More recent results also show that there has been a significant decline in the number of waterbirds in the Macquarie Marshes (Kingsford, pers.comm.). The frequency of waterbird breeding has declined from every year to every four years on average (Johnson, pers.comm.).

(iv) Native fish populations

The native stocks of fish in the Macquarie Marshes appear to have declined in abundance and diversity, while introduced species—such as carp—have increased. The decline in native fish stock may have been affected by streamflow regulation which has resulted in releases of colder water from Burrendong Dam, reduced flows entering the Marshes and reduced flow variability (Swales 1994).

(v) Water quality

The Macquarie Marshes improve downstream water quality by filtering impurities from the water (DLWC 1995a). Filtering occurs when flow levels are large enough for water to leave the channels within the Marshes and spread overland through vegetated areas. Water remains within channels during low flow periods and erosion of channel beds and banks occurs. As a result, downstream nutrient and turbidity concentrations have often been high. The increase in low flow periods because of regulation would have negatively affected downstream water quality.
Pesticide levels are generally lower in the Macquarie Valley than in the Macintyre, Gwydir and Namoi Valleys which also contain large cotton growing areas (DWR 1994).

5.2.5 Reallocation of water in the Macquarie Valley

The NSW Government decided in September 1995 that extra water should be allocated to the Macquarie Marshes. An additional 75,000 ML per year of general security wildlife allocation was allocated to the Marshes, and a limit of 50,000 ML a year was placed on the use of flows from unregulated tributaries and dam spills. Median flow to the Marshes is expected to increase from 315,000 ML to 380,000 ML per year due to these changes (DLWC and NPWS 1996).

Several positive impacts for wetland quality can be expected because of this extra water. It is likely that waterbird breeding will increase from what occurred before the reallocation and the health of most of the vegetation in the area of semi-permanent and lower floodplain will be stabilised. However vegetation in the higher floodplain will receive little benefit from the extra water. The area of the Marshes will fall from 120,000 to 100,000 hectares in the long-term, even with the reallocation (Johnson, pers.comm.).

The reallocation will benefit many graziers and some irrigators within and downstream of the Marshes as they will receive extra water. However, irrigators upstream of the Macquarie Marshes will experience costs. Powell (1995) estimated that a reduction of the supply of water for irrigation by 115,000 ML p.a. would decrease value added due to agricultural production by $14 million per year. DLWC (1995b) estimated that the value of gross margins in the region from adoption of the Water
Management Plan would decline by $55 million in net present value, or $4.4 million per year.

5.3 The Gwydir River Valley

The Gwydir Valley has a similar history to the Macquarie Valley in terms of its development for irrigated agriculture and its declining wetlands. The regulation of water stimulated the rapid development of a large irrigated agricultural industry. The irrigated agricultural industry has also put pressure on existing water supplies, with the reliability of the supply of water for irrigation now one of the lowest in the State. Accompanying the growth in irrigated agriculture has been a decline in the amount of water reaching the Gwydir Wetlands (Keyte 1994).

5.3.1 Site description

The Gwydir Valley is also part of the Murray-Darling Basin, to the north of the Macquarie Valley (see Figure 5.2). The Gwydir River starts at the Great Dividing Range near Armidale and flows to Collarenebri. Copeton Dam, the main regulatory structure in the valley, is east of Bingara. It has a capacity of 1,364,000 ML and an annual average inflow of 445,000 ML. Downstream of Moree the Gwydir divides into two rivers: the Lower Gwydir and Gingham Watercourses. The main wetlands in the Gwydir Valley are along these Watercourses.
The main economic activity in the Gwydir Valley is agriculture. Agriculture was dominated by cattle and sheep grazing and the cultivation of cereal crops prior to the construction of Copeton Dam in 1976. Since then, irrigated agriculture has become dominant. The most widespread irrigated crop is cotton, although other crops, such as vegetables, soybeans, sorghum and pecan nuts, are also irrigated. Cotton production in the Gwydir Valley is about a quarter of total production in NSW ($133 million in 1989/90, DWR 1993). Total employment in agriculture in the Gwydir Valley is estimated to be 3057 people in the 1991 census (ABS 1996).
Prior to the construction of Copeton Dam, it was assumed that 56,000 ha would be licensed for irrigation. However, 86,000 ha had been licensed by 1979. The total allocation of water for irrigation is now over 500,000 ML, equal to 97% of total allocations (DWR 1993). Similar to the Macquarie Valley, a volumetric allocation scheme for regulated water was introduced in 1981. Irrigation water supply reliability of 100% allocation is only 5 years in 10 (Keyte 1994). Irrigators have responded to this uncertainty by increasing on-farm storage capacity for harvesting tributary flows. The capacity of on-farm storage totalled 300,000 ML in 1994.

5.3.3 The Gwydir Wetlands

The wetlands on the Lower Gwydir and Gingham Watercourses are terminal, except during major floods when water reaches the Barwon River. Similar to the Macquarie Marshes, the Gwydir Wetlands contain semi-permanent wetlands, low lying floodplain and the higher level woodlands. Bennett and Green (1993) estimated that the Gwydir Wetlands currently have an area of 102,000 ha.

The Gingham and Lower Gwydir wetlands provide important waterbird habitat and breeding areas. Large numbers of colonial waterbirds breed in the Gwydir Wetlands. Over 100,000 colonial waterbirds bred in the Gwydir Wetlands in floods during 1995/96 (McCosker 1996). Many thousands of non-colonial nesting waterbirds were also observed to be breeding. During this breeding event five endangered waterbird species and seven species protected under international migratory agreements were observed.

The Gwydir Wetlands have supported a substantial grazing industry since the early 1900s. A survey of Gingham Watercourse landholders in 1991
found that there were 103,000 sheep and 24,500 head of cattle on surrounding properties (McCosker and Duggin 1992).

5.3.4 Impact of regulation and consumptive use of water on the Gwydir Wetlands

(i) Flooding

The flooding regime in the Gwydir Wetlands has changed markedly since the completion of Copeton Dam. Bennett and Green (1993) found that, prior to regulation, flows of at least 100 GL/month would have occurred on the Gwydir River in 192 months during a 93 year period. Since regulation they estimate that there has been a 70% reduction in floods of this size. With current regulation, flooding of 100 GL/month would be expected in 58 months during a 93 year period.

Reduced flooding in the Gwydir Wetlands, however, does not solely result from reduced inflows. Similar to the Marshes, erosion and channelisation have increased the amount of flow required before water will leave river channels and go overland.

(ii) Waterbirds

The abundance and diversity of waterbirds in the Gwydir Wetlands appears to have decreased since regulation. Debus (1989) compared species lists compiled during Royal Australian Ornithological Union campouts held in 1933 and 1988. He concluded that the diversity of waterbirds was still intact, but the abundance of colonial nesters and other waterbirds had declined. Evidence from local residents also suggests a decline. Since 1979, 58 waterbird species have been recorded on a
property in the Gwydir Wetlands. This had decreased to 45 species in 1993 and the population of remaining waterbirds had declined (Keyte 1994).

(iii) Vegetation

The main impact from the use of water for irrigation has been on the semi-permanent wetland area. This area previously comprised about 20,000 ha. Before flooding in 1995/96 the area had contracted to about 2000 ha (McCosker 1994). This has been accompanied by an increase in the area of exotic species.

The area of lower floodplain was originally about 80,000 ha in the Lower Gwydir and Gingham Watercourses (McCosker and Duggin 1993; McCosker pers.comm). The 1995/96 floods covered about 55,000 ha, and flooded the majority of the lower floodplain.

(iv) Native Fish

No empirical studies have been undertaken on native fish populations in the Gwydir Wetlands.

(v) Water quality

Currently there is no water quality monitoring within the Gwydir Wetlands. Water quality monitoring stations upstream of the Gwydir Wetlands at Yarraman Bridge and Brageen Crossing provide an indication of downstream water quality.

Water quality is of relatively poor quality at the Gwydir Wetlands. Nutrient concentrations are at levels which can lead to eutrophication. For

41 A flow of 100 GL/month will flood 20,000 ha of the Gwydir Wetlands (Bennett and Green 1993).
example, total phosphorus levels at Brageen Crossing were above 0.05 mg/L more than 75% of the time (Keyte 1994). ANZECC (1992) water quality guidelines suggest that total phosphorus levels in rivers should not exceed 0.01-0.1 mg/L. Levels of pesticides at Brageen Crossing have frequently exceeded levels specified by ANZECC (1992) for aquatic protection. The most common pesticide detected is endosulfan. In samples taken during 1991/92, endosulfan had a median concentration of 0.055 µg/L. The ANZECC (1992) guideline level for aquatic ecosystem protection is 0.01 µg/L.

5.3.5 Reallocation of water in the Gwydir Valley

In September 1995 the NSW Government announced changes to the allocation of water in the Gwydir Valley. While no environmental flow allocations from regulated water were announced, rules were introduced to limit use of off-allocation flows. Under these rules, primary wetland areas were given priority for water except at times of greatest need for crops.

Yatawara and Hill (1996) estimated the annual reduction in irrigators’ gross margins resulting from several reallocation options in the Gwydir Valley using hydrologic and economic models. The first option analysed involved restrictions on the use of all tributary and off-allocation flows. This option was found to reduce gross margins by about $26.5 million p.a.. Under the second option the use of off-allocation only was restricted; this reduced gross margins by $11.8 million p.a.. For the final option the use of off-allocation was restricted by 50%; this reduced gross margins by $2.2 million p.a..

42 Off-allocation is a period when water extracted from a system is not debited against irrigators’ announced volumetric allocation because natural system inflows downstream of the storage, or storage overflows, significantly exceed immediate consumptive demands (DLWC and NPWS 1996).
5.4 Conclusion

Extensive areas of irrigated agriculture have been developed in the Macquarie and Gwydir Valleys. Accompanying this development has been a decline in the amount of water reaching the Macquarie Marshes and Gwydir Wetlands, which has affected wetland quality and quantity. While in many instances there is a lack of rigorous scientific analysis, a number of impacts can be identified. As a result of reduced instream flows, there appears to have been reductions in wetland area, reductions in the species and numbers of waterbirds, poorer water quality, reductions in the health of wetland vegetation and declines in fish populations.

Because of the competing nature of alternative uses of water in the Macquarie and Gwydir Valleys, significant improvements in wetland quality and quantity are likely to occur only if there is a reduction in the amount of water allocated to irrigated agriculture. This will involve costs to the community. Offsetting these costs are a number of benefits, including use benefits from increased cattle production, increased tourism and improved water quality, as well as non-use benefits accruing to the wider community.

Information on the magnitude of these use and non-use benefits, as well as accurate estimates of the opportunity costs from reductions in agricultural production, will assist decision makers seeking socially optimal trade-offs involving the allocation of water between irrigation and wetland uses. While information is available from conventional markets for goods and services to estimate the cost to the community from reductions in agricultural production and many of the use benefits associated with improved wetland quality or quantity, little information is available on the non-use values of improved wetland quality or quantity. Therefore there is a rationale for using SP techniques to provide estimates of these non-use benefits. This is the focus of the next two chapters.
Chapter 6
Questionnaire Design and Development

6.1 Introduction

The key element of any CM application is the questionnaire in which survey respondents are faced with a number of choices between alternative resource use options (Morrison, Blamey, Bennett and Louviere 1996). The design of the questionnaire is therefore critical to a successful application of CM. As discussed in Chapter 3, a CM questionnaire is made up of a number of components, including:

- a description of the study site;
- details of the proposed changes;
- a sequence of choice sets made up of combinations of site attributes at specified levels; and
- a series of socioeconomic and attitudinal debrief questions.

Information must be gathered from external sources to design these components. This includes information about the case study sites, which has been reported in Chapter 5. The design process must also involve input from prospective respondents. Which attributes are regarded as important, the simplicity with which information needs to be presented, the clarity of presentation and the existence of bias are all features of questionnaire design that can only be assessed through feedback from respondents. An effective way of securing this feedback is the use of focus group sessions.

43 The first three sections in this chapter and Appendix 4 have also been reported in
A focus group is a planned discussion involving usually between eight to ten participants (Krueger 1988). The participants are guided by an experienced facilitator and the groups are held in a relaxed and non-threatening environment where participants are encouraged to share openly their opinions and attitudes about a specific topic. Many focus groups are held in specially designed rooms where the people in the group are seated around a table. For SP studies, focus groups are usually between one and a half and two hours in duration (Desvousges and Smith 1988, Rolfe and Bennett 1995).

Partly because of the dynamics that result from having a group of eight to ten people discussing a single issue, focus groups can be used to reveal many of the factors which drive people’s decision making processes, as well as many of the potential problems with draft questionnaires (Krueger 1988). People are generally less inhibited within focus groups sessions compared to personal interviews, possibly as a result of other people sharing their views, and are more likely to express their opinions. Also, opinions expressed by others tend to stimulate people to become more critical. An advantage of using focus groups is that many problems with draft questionnaires, which may not be obvious in an in-person interview, may be revealed, and if a survey does work in a challenging focus group setting, there can be greater confidence that it will work when it is put into the field.

A sequence of focus groups was carried out to explore design issues for the CM questionnaires to be used to estimate the value of environmental improvements at the Macquarie Marshes and Gwydir Wetlands. The results of these focus groups, and the questionnaires developed using these results, are detailed in this chapter. The results from the Macquarie Marshes focus groups are reported in Section 6.2 of this chapter. The results from the Gwydir Wetlands focus groups are presented in Section 6.3. In Section 6.4, the questionnaires used in the Macquarie Marshes and

Morrison, Bennett and Blamey (1997a).
Gwydir Wetlands surveys are described. Conclusions from the focus groups are presented in Section 6.5.

6.2 The Macquarie Marshes Focus Groups

Two main objectives were addressed in conducting the Macquarie Marshes focus groups. The first was to gather information relevant for designing a questionnaire, such as which attributes to include in the choice sets. The second objective was to test a draft questionnaire. Several alternative choice set formats were also tested, and these are reported in Appendix 4.

The results from the four focus groups for the Macquarie Marshes survey are presented in this section. The logistics for focus groups are initially outlined to show how participants were chosen for the groups. The results regarding attribute selection are reported in Section 6.2.2. Results from the evaluation of the draft questionnaire are detailed in Section 6.2.3.

6.2.1 Macquarie Marshes focus group logistics

Two focus groups were held in Sydney on 20 and 21 November 1996 and two were held in Dubbo about a week later on 2 and 3 December 1996. All of the focus groups were facilitated by the author of this thesis.

The first Sydney focus group was held between 6-8 pm in North Sydney, while the second group was held between 10 am-12 noon in Parramatta. The aim of holding the Sydney focus group in two different locations was to achieve a more representative sample of the Sydney population. The Sydney groups were held in specially designed focus group rooms that contained a one way mirror, so that observers could watch the focus groups without disrupting the sessions, and video and cassette recording facilities.
The Dubbo focus groups were held at the same location: the first from 6-8 pm on Monday 2 December, and the second between 10 am-12 pm on Tuesday 3 December. The Dubbo groups were held in a conference room within a motel and participants were seated around a table. Both groups were video and audio recorded.

A professional recruitment agency (Applecom Research) was contracted to select a sample of 10 people that was representative of the population in terms of age and sex for each of the focus groups. People approached as prospective respondents were told that the focus groups involved the discussion of an ‘environmental issue’ and payment was mentioned after they had agreed to attend. The two Sydney groups were selected through a random telephone survey about a week before the groups were held. The agency selected the people for the two groups in Dubbo by ringing contacts in Dubbo and asking them to invite people they knew. This method is not ideal, but is a pragmatic approach given the difficulty of recruiting participants for focus groups in country towns. The age and sex statistics for each of the focus groups is shown in the following table:

**Table 6.1: Age and sex statistics for the Sydney and Dubbo focus groups**

<table>
<thead>
<tr>
<th></th>
<th>Sydney FG#1</th>
<th>Sydney FG#2</th>
<th>Dubbo FG#3</th>
<th>Dubbo FG#4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number female</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Number 20-30 years</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Number 31-50 years</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>Number 51-65 years</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
</tbody>
</table>
At the beginning of each focus group respondents were asked what information they would need to know to decide whether to support an option to allocate more water to the Macquarie Marshes. This questioning was designed to determine which attributes should be included in the choice sets. The information requested by participants can be broadly grouped into two main themes: information about the Macquarie Marshes, and information about the impact of increasing flows to the Macquarie Marshes on irrigators and townships in the Macquarie Valley. Specifically, participants wanted answers to the following questions:

- What benefit are the Marshes for wildlife?
- What is the effect of increased water on flora and native animals?
- What is the effect of extra water on the landscape?
- What species of fish have been affected?
- What are the current land uses around the Marshes?
- What will happen to farmers around the Marshes if there is increased flooding?
- What is the size of the Marshes?
- If the Marshes weren't there, would birds move to other wetlands?
- What would happen to migratory birds if the Marshes weren't there?
- What has been the effect of reduced water on erosion?
- Is there potential for tourism at the Marshes?
- Do the Marshes improve water quality?

These questions indicate several potential attributes. These are wetland area, native vegetation, wildlife, waterbirds and water quality. Several other questions were asked to help determine more precisely which wildlife attributes should be included in the choice sets. It appeared that participants were interested in knowing the effect on rare and endangered bird species, waterbird populations and the number of waterbirds breeding. Participants were also keen to know whether there would be adverse impacts on irrigators and communities in the Macquarie Valley. This

---

44 This is explained in greater detail in Morrison et al (1997a).
suggests a need to include an attribute in the choice model to reflect the impact of any reallocation on employment, and, possibly, regional income.\footnote{Blamey (1996) discusses the factors that affect respondents' beliefs and decisions in the context of Schwartz's (1977) norm activation model. One of the factors identified is whether 'initiatives are distributively fair' (p.24). Including a non-environmental attribute would provide respondents with some indication of the effects on the distribution of wealth from the adoption of a new policy option.}

### 6.2.3 Reactions to the draft questionnaire

Participants were asked to complete a draft questionnaire that had been prepared prior to the focus groups (see Appendix 1). Participants were then asked a series of questions about how they answered the choice sets or their reactions to the information at the beginning of the questionnaire.

**Bias**

The initial reaction of most participants showed that they thought the draft questionnaire was biased. One reason given was that the only option in the choice sets was for increased water to the Marshes (see Table 6.2). Participants felt that they could only select improved wetland quality; that they could not choose to continue the current situation or reduce wetland quality. An option ‘Choose neither A nor B’ was included, which in effect allowed participants to choose not to have more water allocated to the Marshes. However, many participants did not notice this option.

It doesn’t give you any alternative but to say yes to more water to the Marshes. (S1\footnote{S1 represents the first Sydney focus group, S2 the second Sydney focus group, D1 the first Dubbo focus group and D2 the second Dubbo focus group.})

There’s not enough room...to make a valid decision...You’re tied down in what you want to say. (D2)
**Table 6.2: Draft Macquarie Marshes questionnaire choice set format**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in wetland area</td>
<td>40,000 ha</td>
<td>80,000 ha</td>
</tr>
<tr>
<td>Increase in area of healthy river redgums</td>
<td>5000 ha</td>
<td>10,000 ha</td>
</tr>
<tr>
<td>Increase in number of waterbirds</td>
<td>80,000</td>
<td>160,000</td>
</tr>
<tr>
<td>Increase in waterbirds breeding</td>
<td>20,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Increase in waterbird species</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
</tr>
</tbody>
</table>

I would choose A [ ]  I would choose B [ ]  I would choose neither A or B [ ]

Another reason was a lack of information about the irrigators and local towns:

...it hasn’t done any impact studies on the impact on the community. (S1)

...you’ve mentioned nothing about the farmers...and the effect farmers are having on the land...For people to make an informed decision they need both sides of the story...positive and negative effects of irrigation and of the Marshes. (D2)

The photos also proved to be a source of bias:

These photos - you’ve got the really nice greenie ones and then...each time the one that looks like hell. (D1)

A final reason given was an excessive concentration on waterbirds. Three waterbird attributes were included in the draft questionnaire. Having so many waterbird attributes suggested to people that there may be a pre-existing agenda:
To me water quality was the big issue that should be addressed and I couldn’t understand why birds kept getting mentioned. (D2)

From these responses it appears that several measures could be used to reduce perceived bias in the choice sets, such as (1) including the base alternative within choice sets, (2) including an option where respondents can choose reduced wetland quality, (3) including an employment or regional income attribute, and (4) reducing the number of waterbird attributes. In terms of information, details could be reported about the irrigators and townships. For the photos, a less biased approach may be to just include a photo of what the Marshes look like when they are in a healthy state and avoid extreme contrasts.

Confusion

A number of respondents found answering the choice sets in the questionnaire confusing. Some participants found it difficult to think about seven attributes:

...I can only focus on one issue... (S1)

...You had to really concentrate, that’s what I was finding...it was easy to get confused. (D1)

Participants were also confused by the changing attribute levels across the choice sets:

Confused with all those figures in the tables for a start...because all the other figures were switched and swapped around there...it looked like trick questions, it was confusing. (D2)

You’ve got an increase in wetland area of 40,000 ha in Option A, Option B is 80,000 ha. Flip over to the next question and its 40 and 80 again...so the first two lines are the same and then they’re just swapped over...that’s where the confusion comes from - what are you getting at? (S2)
Other participants became confused because of the number of choice sets:

The first time I was OK but towards the end I was confused because it gave me so many different options and I thought, 'Did I choose the same back on the other page?' (D1)

Confusion had several effects on the way participants answered the questionnaire. A few participants questioned the seriousness of the survey:

I started and then I was almost going to stop when I looked at the first few. I thought, was this for real? (D2)

Other participants sought to remedy their confusion by using heuristics to answer the choice sets. In particular, they would try to focus on just one or two attributes. This strategy is explored more fully in Section 2.3.4.

I found it rather confusing to answer. At first I'm really studying each alternative and trying to consider it, and by question 10, I'm just looking at the cost of the good. (S1)

Water quality was the main one I focused on, regardless of the cost or anything...To look at it too carefully is just incredibly confusing. (S1)

Several options are available for reducing confusion. The most obvious options are to reduce the number of attributes and choice sets. Another option, as discussed in Chapter 4, is to use a less efficient experimental design, such as an $L^M^N$, where not all of the attributes in choice sets have different values across alternatives.

Implausibility

This problem has been identified in other conjoint analysis applications (Pearmain et al 1991; Steckel, DeSarbo and Mahajan 1991). The main reason participants gave for implausibility was that they could not understand why certain attribute levels went together:
A lot of the groupings are improbable...like poor water and more species and more bird breeding going together. (S1)

If you’ve got 80,000 ha of extra wetland why is water quality only fair? (S1)

Participants were not told in the questionnaire how each of the alternatives in the choice sets would come about. Rather the following statement was included in the questionnaire:

To keep matters simple, we do not describe how each management option would come about. For this reason, some options appear a little unusual, or even unrealistic. Bear in mind that there are a lot of ways that water can be managed, so that options which seem odd are actually quite possible.

There are several possible ways of reducing implausibility caused by combinations that don’t make sense to respondents. One way is to explain why different alternatives, that may appear odd, could occur. If participants understood why seemingly implausible alternatives could be possible, they may be less confused by unexpected changes in attribute levels. Explaining why unusual alternatives are possible has an additional advantage. A number of participants were irritated because there was no explanation of how the alternatives were derived. Some participants felt that they were being manipulated. Others commented that the information would probably influence how they would answer:

1 Are they really saying it for that reason or is it because they want to keep the information from us so that we can’t make a more informed decision.
2 It says bear in mind that there is lots of ways that water can be managed but they’re not telling us what they are.
3 I don’t know what has brought about these choices and as I went along I was getting more and more sceptical...I didn’t know what to tick and why. (S2)

Another possibility for removing this form of implausibility is to create composite attributes or use correlated experimental designs, as described
in Section 4.3.3. A third possibility is simply to tell respondents that the options presented in the choice sets are hypothetical and are designed to find out which attributes are most important to people.

Not specifying the actual level of the attributes, but changes in attributes associated with each option, also appeared to add to implausibility. Participants identified problems that they would not have cited if they had been more aware of the actual levels of the attributes:

If you’ve got the smaller area, and you’ve got 160,000 birds and you’ve got new species coming in and they’re all breeding like mad and you don’t have a lot of trees then you’ll have another type of disaster, not just lack of water. (S1)

Implausibility also appeared to cause participants to focus on one or two attributes:

I didn’t think that they were very realistic so for just about every one I took the cheapest. (S1)

I actually got quite irritated with the questions...because there are so many and they are unrealistic...in the end I was just looking for good water and the price didn’t worry me. (D1)

**Indicator attributes**

In the focus groups, many participants concentrated on a single attribute when selecting their preferred option from each choice set. For most participants this attribute was water quality. Under this strategy they typically argued that if one attribute was of good quality then the other attributes would ‘follow’, even though a close examination of the choices would have revealed a contradiction:

...my main focus was water quality because I thought that if the quality of the water was good then everything else would just fall into place. (S2)
...if you don't have healthy redgums the birds can't be supported anyway. (S2)

I went for water quality because if the water quality is good all the rest will take care of itself basically. (D2)

The use of this 'indicator attribute' strategy is a serious problem because, in effect, participants are dismissing the information about the other attributes in the choice sets. Even if in an alternative, say, the levels of all the other attributes fell but water quality increased, participants would believe that, in the long-run, the levels of the other attributes would actually increase. The result is that it may not be possible to determine the importance that participants attach to each of the attributes.

There appear to be several reasons why respondents are using this heuristic. Two reasons were given in the previous sections: implausibility and confusion. If participants consider that the survey is implausible (and bias could have the same result) they will not give it careful attention and only focus on one or two attributes. If participants see it as confusing, they may use a strategy whereby they focus solely on one or two attributes to simplify how they answer the choice sets (see also Blamey et al 1997). Another possibility is that participants may have genuinely believed that the attributes were related. Several options for minimising the use of indicator attributes follow from these explanations. These are: (1) reduce implausibility and bias, (2) reduce confusion and (3) ensure that participants appreciate what causes changes in each of the attributes so that they are less likely to use one attribute as an indicator for other attributes.

**Adequacy of information**

Participants were asked whether the information in the questionnaire was adequate. Many questions were asked about irrigation in the Macquarie Valley:
• Will the purchase of water be one-off?
• Will there be pressure on farmers to sell the water?
• Is there any excess water in Burrendong Dam?
• Will farmers sell water if they don’t have enough?
• Do farmers currently pay for their water?
• What are the benefits of irrigated agriculture to the surrounding community and the economy?
• What will be the effect of the existing reallocation on farmers and on the Marshes?
• Why and how were the farmers given the water in the first place?
• How much water do the irrigators currently get?
• What are the negative effects of irrigation on the land and on water quality?
• What type of crops are irrigated?

The lack of information about the irrigators appeared to cause some participants to be concerned about the negative effects of a reallocation. Note the following comment:

We don’t have any information about the farmers or the irrigators so we keep coming up with reasons about why we shouldn’t be concentrating on the Marshlands. We keep getting back to how it will affect farmers because we don’t have this information. (S1)

Many participants also found the information about the water trading market ambiguous and difficult to understand, and asked the following questions:

• How does the water trading market work?
• What is the market’s geographic boundary?
• Who trades in the market?
• From where will the extra water come?
• How much water do the irrigators actually need?
• Will the water be purchased each year?
Other issues

Several other issues involving the draft questionnaire are briefly discussed.

• Reminder statement

In the draft questionnaire the following reminder statement was included:

When thinking about which option you prefer, keep in mind your available income and all the other things that you have to spend money on. It is possible that in the future other environmental projects may cost you additional money.

This statement appeared to have the desired sobering effect on participants:

1 Its frightening isn’t it.
2 ...sobering.
3 Certainly made you think about...how much its going to cost. (D2)

• One-off levy

Questions were asked to determine whether participants believed that the levy on water rates, which was used as the payment vehicle in the draft questionnaire, would be one-off. Similar to other studies (eg Bennett, Blamey and Morrison 1997), it was found that participants doubted that the levy would be one-off. As a result, when participants agree to pay a certain amount, they are probably exhibiting a much higher willingness to pay. The following comments were made in the focus groups:

I believe that it would be twice or three times or once every 5 years. (S1)

1 They’d need more.
2...$30 ongoing. (S2)

It would be like the petrol levy...3 by 3 by 3 by 3... (D1)
Summary

In summary, several serious problems were encountered in the draft questionnaire. Participants thought that the questionnaire was biased, they found answering the choice sets confusing, they found some of the alternatives in the choice sets implausible, the information about irrigation and the water trading market inadequate, and they doubted that payment would only be one-off. In addition, many respondents appeared make use of indicator attributes when answering choice sets. The use of this heuristic appears to have been motivated in part by other problems with the questionnaire such as implausibility, bias and confusion. It is possible that the use of this heuristic may be reduced by minimising these problems.

6.3 The Gwydir Wetlands Focus Groups

Similar to the previous set of focus groups, the primary objective of the Gwydir Wetlands focus groups was to trial a draft questionnaire. The questionnaire was similar to the one used in the Macquarie Marshes focus groups, but was modified on the basis of the conclusions reached in the previous section\textsuperscript{47}.

The logistics of the focus groups are reported in Section 6.3.1. In Section 6.3.2 the main differences between the Macquarie Marshes and Gwydir Wetlands questionnaires are shown. Participants’ reactions to the questionnaire are reported in Section 6.3.3.
6.3.1 Gwydir Wetlands focus group logistics

Two focus groups were held in Sydney on 16 and 17 April 1997 and two were held in Moree on 18 April 1997. All of the focus groups were facilitated by the author of this thesis\(^\text{48}\). The first Sydney focus group was held in North Sydney between 6-7.30 pm, while the second group was held in Parramatta between 10-11.30 am in the same locations as the Macquarie Marshes focus groups. The focus groups held in Moree were conducted in a conference room in the Skillshare offices between 2-3.30 pm and 6-7.30 pm. All four of the focus groups were audio and video recorded.

A professional recruitment agency (Applecom Research) was again contracted to select participants for the Sydney focus groups, using a similar approach to that described in Section 6.2.2. The participants for the Moree groups were recruited through a contact in Moree\(^\text{49}\). The age and sex statistics for each of these groups are shown in Table 6.3:

<table>
<thead>
<tr>
<th>Table 6.3: Age and sex statistics for the Sydney and Moree focus groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney FG#1 FG#2</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Number female</td>
</tr>
<tr>
<td>Number 20-30 years</td>
</tr>
<tr>
<td>Number 31-50 years</td>
</tr>
<tr>
<td>Number &gt;51 years</td>
</tr>
</tbody>
</table>

\(^{47}\) The Gwydir Wetlands questionnaire was also professionally edited by Mr Ed Highley of Arawang Pty Ltd.

\(^{48}\) Russell Blaney assisted in facilitating the first of the Sydney focus groups, and Jeff Bennett assisted in facilitating the two Moree groups.

\(^{49}\) This proved to be a much more cost-effective means of recruitment, and was identical to the approach used by Applecorn Research in Dubbo.
6.3.2 Differences between the Macquarie Marshes and Gwydir Wetlands questionnaires

A number of changes were made to the Gwydir Wetlands questionnaire based on the findings of the Macquarie Marshes focus groups. The most significant of these changes are reviewed in this section. A copy of the questionnaire trialed in the Gwydir Wetlands focus groups is contained in Appendix 2. The main changes were as follows:

**Framing**

Respondents were asked to rate the importance of improving wetland quality relative to the importance of five other environmental issues in the draft Macquarie Marshes questionnaire. Because of the difficulty expressed by some participants in providing this rating, a simpler but more comprehensive framing was used in the draft Gwydir Wetlands questionnaire (see Bennett et al 1997).

**Information about irrigation**

A page of information was included about irrigation in the Gwydir Valley to reduce bias. This included information on production, employment, water use and water trading.

**Information on why attributes change independently**

In the Macquarie Marshes questionnaire no explanation was given for why attributes change independently. The following paragraph was included in the Gwydir Wetlands questionnaire:

The removal of water for irrigation has altered the natural pattern of flooding. This has had a different effect on each aspect of the Gwydir Wetlands. Wetland vegetation requires regular floods of various sizes to remain healthy. Waterbird breeding requires flooding of sufficient size at a suitable time of year. Endangered
species require the preservation of certain habitats through regular flooding. These different flooding needs mean that water managers must choose which aspect of the Wetlands to preserve if extra water is available.

**Photos**

In the Gwydir Wetlands questionnaire no photos were included in the questionnaire. Rather, participants were given a sheet containing four different photos of the Gwydir Wetlands. All of these showed the Gwydir Wetlands in a healthy state. It was envisaged that one of these photos, together with a photo of a cotton farm, would eventually be used on the cover of the questionnaire.

**Experimental design**

Main effects L^{MN} experimental designs were selected from Hahn and Shapiro (1966) for the Gwydir Wetlands questionnaires. This design was used because it is simpler for respondents to answer, even though it is a less efficient design. Choice sets with dominated options were deleted to reduce implausibility.

**Choice sets**

A constant base alternative was included in each choice set as shown in Table 6.4. All of the attributes (except household cost) were expressed as absolute values rather than as increments, and labels were used.
Table 6.4: The non-balanced choice set format

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water rates</td>
<td>no change</td>
<td>$100 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Employment</td>
<td>2800</td>
<td>2780</td>
<td>2720</td>
</tr>
<tr>
<td>Wetland area</td>
<td>400 km²</td>
<td>800 km²</td>
<td>500 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 4 years</td>
</tr>
<tr>
<td>Endangered and protected species</td>
<td>12</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

Attributes

The number of environmental attributes was reduced from six to three in the draft Gwydir Wetlands questionnaire. The three waterbird attributes were replaced by two attributes: frequency of waterbird breeding, and number of endangered and protected species using the wetland as habitat. Frequency of waterbird breeding was used because of a lack of information about actual numbers of waterbirds breeding, and because it was thought that it would be easier for respondents to comprehend. Endangered and protected species was used because participants in the Macquarie Marshes focus groups indicated that it was an important attribute. Area of healthy river redgums was excluded because it was similar to the wetland area attribute, there was an obvious need for parsimony, and because wetland area was thought to be more useful for benefit transfer. Water quality was deleted because the Gwydir Wetlands are terminal and it is therefore not an outcome that would be achieved by increased flows to the Gwydir Wetlands.
Bias

The initial reactions in the Gwydir Wetlands groups revealed that the extent of bias had been reduced in comparison to the draft Macquarie Marshes questionnaire. In only one of the four groups was the initial reaction that the questionnaire was biased, and less than half of the participants in this group expressed this reaction. None of the participants in the Sydney groups noted the existence of bias.

"I think it is good because I really don't know much about it and it brings it to your attention that there are problems out there that I'm not really aware of... (S3)"

"I thought that it was quite reasonable (S4)"

"Little bit one sided isn't it... biased from the start (M2)"

"It could have been termed inflammatory in certain parts (M2)"

More careful probing revealed that many of the participants in the Moree groups, who believed that the questionnaire was biased, thought that the bias was subtle and difficult to identify:

1 I think that its pro-wetlands
2 But I don’t see how you’d get away from it
3 Its just the issue
4 You could talk more about the economic benefits of the irrigation for Moree (M1)

1 Yes...its just my initial impression. They...have some stuff later on about cotton. But it still seemed more against it.
2 I didn’t think it was biased. It presents both sides...

---

S3 represents the first Gwydir Wetlands focus group in Sydney; S4 the second Gwydir Wetlands focus group in Sydney; M1 the first focus group in Moree; and M2 the second focus group in Moree.
Some participants were able to identify parts of the questionnaire that were causing bias. A few participants thought it was biased because the uncontrolled flooding in the wetlands area that occurred before Copeton Dam was constructed was not mentioned:

Q: So you’d want to know the beneficial effects of the dam [for the wetland]?
A: Oh, absolutely, yes...I think that Sydney people should get that sort of thing because they don’t understand irrigation. Its like criticising motherhood when you criticise not having water for wetlands. (M1)

Another source of bias identified was the phrase ‘The removal of water for irrigation’:

Just your first sentence there ‘The removal of water for irrigation’. It hasn’t been totally removed...its very confrontationist...it makes people think that its taken all the water. (M1)

The opening paragraph in the questionnaire may also be a source of bias, although other participants questioned this:

1 The opening paragraph says that the wetlands have fallen because of Copeton Dam irrigation. That basically slaps the irrigators in the face when you start...
2 How cushy do you want to make it really. If you’re just laying out the facts and you’re a sensible person, surely you can make your own judgment. (M2)

Another source of bias identified was in the paragraph on water use:

Varying the water use. Then it goes on to say ‘However this is supplemented by...’
Is this making the reader feel that they’re always going to get their 100%...? (M2)

A final source of bias identified relates to the rationale given for why farmers would choose to sell water:
Confusion

While bias had been considerably reduced in the draft Gwydir Wetlands questionnaire, many participants still found the questionnaire confusing:

...it was bloody confusing. (S3)

I think it’s good to have the opportunity to give your view on such things because it’s easily done but this survey is not so easily done. (S4)

Reading this is a bit complicated for me and a lot of Aboriginal people. (M1)

However, the extent of the confusion was notably less than with the Macquarie Marshes questionnaire. Some participants thought that, while the questions were difficult, they were able to express their preferences:

If you look at...[the] $100 increase...and you say ‘well I’m not [prepared to pay it]’ so you don’t...If you’d like to see the area doubled and birds breeding there every two years then there’s an option there...I didn’t find it that difficult in that if you sat down and looked to all the different options...You nut out what you want and try and find your best fit. (M1)

Pretty much the same in regard to the hip pocket nerve but also I took into consideration the area and the species... excepting one case...where the $100 increase brought great changes or in one case absolutely nothing for a $30 increase. (S4)

Participants gave various reasons for the existence of this confusion. Some participants found the choice sets difficult to understand:

1 You need to go over it several times. I changed a couple of my answers after I perused it.2 I did that too...I crossed out the ones which I had actually ticked and
changed them to a different box after I had read further...and discovered that there were several options. (S3)

What has tricky questions got to do with what we feel should be done?...You want to ask a...straight out question whether we want more water to go to the watercourse or whether you don’t...but I tell you I couldn’t work out this. (M1)

1 Damned confusing...all these boxes and things. When I first looked at them I thought they were all the same...It was only when I came back and had a look that I thought, oh, its not quite the same.
2 It was hard...its a very subtle change that you’re...looking for. (M1)

Other participants found it difficult to understand how the water trading market worked. In particular, there was confusion about the difference between charges for the use of water and costs associated with the permanent purchase of water licences:

1 Why has water become so expensive?
2 I got a little bit confused...farmers paid about $5/ML for the water they use each year. I found it difficult seeing the relationship...to the price of the water rights.
Q: Water rights: that came across as a yearly amount?
Sort of seems to me that it is per year (S4)

These responses suggest that there is a need to simplify choice sets, and that the information presented in the questionnaire needs to be simplified and clarified further.

Some participants (mostly from S4) indicated that confusion resulted because their general understanding and knowledge about the Gwydir Wetlands was limited. These participants suggested that they were not capable of responding to this survey. They thought that the survey would be more appropriately completed by people in Moree51. Some participants in Moree had a similar attitude: they did not think that Sydney people should be completing the questionnaire.

---

51 This response has been identified in other situation involving the estimation of non-use values (eg Bennett, Blamey and Morrison 1997)
1 Should be done by the people living in that area
2 They know the area better and probably have seen the changes and what the effects have been on bird migration there and they'd be better to make comments (S3)

1 They're also extremely specific for people like us who are fairly removed from the situation
2 Its nice to have been asked these questions but I'm not the person to ask, because if you ask me about televisions I can tell you how they work because that's my job, but I'm not an expert on how much area you need for a heron to successfully reproduce and survive... (S4)

1 I'd rather have someone tell me 'look there are so many numbers of birds and this is the cycle that we need...I'm worried that a whole lot of people who don't really know anything about it will put all these forms in and then from that data comes the magic number which could be totally wrong...
2 ...you've got to have some specialist understanding of the whole situation cause otherwise you can't make an acceptable comment. (S4)

On an initial reading, these comments may cast doubt on the appropriateness of using CM outside the region in which an environmental issue is occurring. However, it is possible that part of the problem is that participants were asked how they would like the Gwydir Wetlands to be managed. If the emphasis was instead placed on finding out what aspects of the wetland are important to people, or what they would like to see preserved, this problem might be minimised. Respondents may be much happier to reveal their preferences rather than stating what they think should be done.

Implausibility

There was less evidence of implausibility compared to the draft Macquarie Marshes questionnaire. However, it was apparent that a few participants did not understand the relevance of the paragraph about attribute independence because it was included too early in the questionnaire:
I kept wondering if there was a big area why aren’t there more birds?... I found it a really confusing thing...Now I’ve read it and noted it and now I understand where you’re coming from. I kind of didn’t notice it. What you’re saying is it depends on water releases, its not just a matter of size of the wetlands area.

Other participants commented that they found it difficult to treat each of the choice sets as independent. One remedy for this problem could be to put each choice set on a separate page:

Trick questions...how you answered question four related directly to how you should answer question six. The increase in water rates was $30 whereas in this one it was $100...with the same wetland area...breeding cycle...endangered species and employment. (S3)

Like they were meant to be totally independent of each other but you sort of got a feeling that if you answered one way for one [it excluded you from answering a certain way for other options]...They weren’t really separate entities. (S4)

Participants found some of the information presented to be implausible. In the Moree groups some questioned whether the government would actually purchase water from irrigators rather than simply reduce their allocations:

Option 2: is that...fair dinkum? I thought like if they want the water they’d just take it. (M1)

Q: Did you believe it when you read it, did you think it was reasonable?
[Several nos]
1 I didn’t think they’d buy it
2 I did (M1)

Q: Did you believe it...the government buying rights?
[Two yeses]
1 I don’t really know
2 I’m against the whole concept (M2)
Some participants doubted that irrigators would actually agree to sell water, except in a fairly narrow set of circumstances. This suggests that further explanation needs to be given about why irrigators would agree to sell water.

I've never known a farmer to have too much water (M1)

1 ...I don't think that the government has got a hope in hell of buying any back to be honest.
2 It all hinges on the fact that the growers will sell the water because they want the money and its generally not the case...
3 ...the people that...have licences to sell are probably people who want to get out of the industry; [people]...who have gone through pretty rough times with the drought and have got to sell; and unused licences...that would be the three categories. (M2)

A few respondents also questioned whether it was possible to save water through increased efficiency. They typically noted that drip irrigation had not proven very successful. This suggests a need to elaborate more carefully how water could be saved.

My theory was with the drip irrigation and its not necessarily working properly...(M2)

Indicator attributes

There was little evidence of the indicator attribute problem in the draft Gwydir Wetlands questionnaire, suggesting that the steps taken to minimise this problem had been effective.

Adequacy of information

Participants asked for extra information about the following issues. While this information may be relevant for some participants, what is perhaps more salient, given that many participants are already being overloaded
with information, is whether the extra information is relevant for the majority of participants.

- History

Several respondents in the Moree focus groups thought that the questionnaire should include more information about the history of the Gwydir Wetlands. A couple of participants noted that uncontrollable flooding of the wetlands and surrounding farms occurred before Copeton Dam. They thought that the flood mitigation benefit of Copeton Dam should be explained. Other participants thought that the benefits of the wetland for grazing should be more fully explained; in particular, the reliability of feed around the wetlands.

- What has happened to the land that used to be wetlands?

Participants were told that the area of the Gwydir Wetlands had fallen from 2000 km$^2$ to 400 km$^2$. One participant queried what had happened to the 1600 km$^2$ that had ceased to be wetland. A related question was whether this area could again become wetland.

- Origin of the information in the questionnaire

Several participants questioned the origin of the information used in the questionnaire. They were unsure of its validity, and wanted references to guarantee this:

I really don't know whether there were 19 bird species listed. I mean how many people know that?...I wouldn't know if somebody told me that 225 bird species were found in the area. I'd [just] think there must have been. (M1)

Some of these statistics you've got. I don't know how you got them...(M1)
• Employment

A couple of participants in the first Moree focus group thought that more people were employed in irrigation:

1. . . . there's vastly more people employed in irrigation.
2. You've got 1600 employed here. I'm sure that I passed 1600 trucks just getting here. (M1)

In contrast, a couple of participants in the second Moree group suggested that irrigation had been associated with a shift from full-time to casual or seasonal employment. While it would be difficult to describe trends in employment in Moree, these comments suggest a need to include references for the employment and other statistics used in the questionnaire.

Summary

In summary, the changes made in this questionnaire to minimise bias, implausibility and the use of indicator attributes appeared to work relatively successfully. However, some respondents still had difficulty understanding how attributes could be independently varied and confusion still appeared to be a problem. To some extent confusion remained a problem because of the lack of knowledge that people in Sydney had about water related problems in the Gwydir Valley. This is true of many of surveys in which non-use values are estimated and suggests that these surveys need to be focused on people's preferences rather than the management of natural resources.
6.4 The Macquarie Marshes and Gwydir Wetlands
Questionnaires

In this section, the final changes made to the questionnaires are described\textsuperscript{52,53}.

Preferences about options versus views on management

Several changes were made to emphasise that the survey sought information about respondents' preferences about different options, rather than their views about how the wetlands should be managed. First, the titles were changed from 'The Management of the Macquarie Marshes (or Gwydir Wetlands)...' to 'Options for the Macquarie Marshes (or Gwydir Wetlands)...'. Second, the title of the introductory paragraph was changed from 'Your views on the management of the Macquarie Marshes (or Gwydir Wetlands)' to 'Your views on the Macquarie Marshes (or Gwydir Wetlands)'. Other references about how the community would like to see the wetlands 'managed' were also deleted.

Framing

The frame used in the final questionnaires was modified to reduce the amount of information in the questionnaire. The first question used in the Gwydir Wetlands draft questionnaire was deleted, and only the second question was used.

\textsuperscript{52} Copies of the questionnaires are in Appendix 3.
\textsuperscript{53} A pre-test of fifty surveys was undertaken in Sydney. However, only very minor modifications were made on the basis of the pretest.
Information about the Gwydir Wetlands

Much of the information about the Gwydir Wetlands was included under the map that accompanied the questionnaire\textsuperscript{54}. A number of changes were made to reduce and clarify the information in the questionnaire. The main changes were as follows:

- information about the separate areas of core and floodplain wetland were deleted;
- the area of the wetlands was compared to the area of Sydney;
- reference to where waterbirds are found was deleted;
- reference to the number of stock in the wetlands, and the use of the wetlands for drought relief for stock was deleted;
- the introductory paragraph on the effect of less water on the Gwydir Wetlands was deleted;
- the sections on waterbird breeding and endangered and protected species were combined;
- the sentence describing the effect of reduced water on stock feed and the income of graziers was deleted; and
- information on when changes in the quality of the wetlands occurred was included.

Information about irrigation

Various changes were also made to the information describing irrigation to improve clarity. The main changes are as follows:

- a pie chart was used to describe the revenue from different forms of agricultural production;
- the percentage of people in the region employed in irrigation in the region was reported;
- reference to the per-unit water charge was deleted;
- respondents were told that a megalitre was about equal to the amount of water in an Olympic swimming pool; and
- the sections on water use and trading were carefully reworded.
Information about available options

Respondents were told that one option available for farmers to reduce their water needs was the use of drip irrigation. Respondents were also told that the levy on water rates would be payable only in 1998, and that the water would be purchased from farmers in the following year.\textsuperscript{55}

Information on why attributes change independently

Information about why attributes change independently had been included in the section describing the wetlands in the draft Gwydir questionnaire (see Section 6.3.2). However, respondents did not appreciate the importance of this paragraph. The information was therefore modified, and included just before respondents started to make their choices, so that it would be fresh in their minds. The paragraph was as follows:

The outcomes in each of the options have been specifically defined so that you have a broad range of choices. Within this range, some options may seem strange according to your experience, but bear in mind that there are many different ways of managing water. For example, wetland vegetation requires regular floods of various sizes to remain healthy. In contrast, waterbird breeding requires flooding of sufficient size at a suitable time of year.

Photos

Two photos were included on the cover of the questionnaires: one showing a healthy wetland, and the other a prosperous looking cotton farm at harvest. The photos were placed side by side and were the same size.

\textsuperscript{54} This format was tested in the second set of focus groups and received support.

\textsuperscript{55} A limitation of the questionnaire is that no explicit social decision making structure or rule was described, other than the increase in water rates applying to all households. Defining a decision rule is of importance as it can affect the incentives that respondents face to answer truthfully. However, defining a decision rule in the context of CM is more difficult than with the CVM. This is because respondents are presented with multiple alternatives and answer multiple choice sets. Therefore this should be a topic of further research.
Experimental design

An $L^{MN}$ experimental design selected from Hahn and Shapiro (1966) was selected for the questionnaires. It was a main effects, unprotected, design, that allowed the estimation of selected two-way interactions (Resolution 3). The design had twenty-seven alternatives, however, seven of the alternatives were dominated or implausible. These alternatives were deleted in the Gwydir Wetlands surveys, leaving twenty alternatives that were blocked into four sets using a four level blocking factor. An additional block of five dominated alternatives was included in the Macquarie Marshes survey to allow a test of the effect of deleting dominated alternatives (see Appendix 5).

This experimental design was chosen because it had few choice sets, which meant that there would be more replications per choice set. Secondly, the $L^{mn}$ was favoured over a foldover design because it is easier for respondents to answer as trade-offs do not occur on each attribute. Thirdly, this design permits the estimation of mother logit models to test for IIA violations. Finally, a design that had a higher resolution was not chosen because the results from the focus groups suggested that 2-way interactions were not likely to be significant given the difficulty respondents faced when answering choice sets.

Attributes

Five attributes were included in each choice set in the final questionnaires as it was decided to include an employment attribute to reduce bias. The attributes were:
Several of the names of the attributes were modified to reduce confusion. 'Water rates' was changed to 'Your water rates' and '(one-off increase)' was placed in brackets after the descriptor. 'Employment' was changed to 'Irrigation related employment'. 'Endangered and protected species' was changed to 'Endangered and protected species present'. Changes were also made to the descriptors used for the attribute levels. For irrigation related employment the word 'jobs' was included, and 'species' was included when describing the number of endangered and protected species present.

**Choice set design**

A modification of the combined balanced-unbalanced choice set format was used in the final questionnaires (see Appendix 4). It was decided not to use the balanced format because of (1) the implausibility of the rebate, and (2) because of concern about respondents not being discriminating in their choices. The combined balanced-unbalanced format was used because it was thought that it would be less biased than the unbalanced format. The final alternative was, however, altered to 'I would not choose any of these alternatives because I would prefer more water to be allocated for irrigation'. An example of the choice set design used in the questionnaires is show in Table 6.5:
### Table 6.5: Example of a choice set from the Macquarie Marshes questionnaire

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Macquarie Marshes</th>
<th>Option 3: Increase water to Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$20 increase</td>
<td>$50 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4350 jobs</td>
<td>4350 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>1250 km²</td>
<td>1650 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every 3 years</td>
<td>every year</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1
☐ I would choose option 2
☐ I would choose option 3
☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation

### Debrief questions

The usual socioeconomic and attitudinal debrief questions were included in the questionnaire to determine whether the survey samples were representative of the population, and for use in later modelling. Six likert scale statements were also included to provide information about what respondents thought about the adequacy of the questionnaire, given the problems experienced in the focus groups. The statements were as follows:

(1=strongly agree, 5=strongly disagree)

- I understood the information in the questionnaire
- The information in the questionnaire was biased in favour of the wetlands
- The information in the questionnaire was biased in favour of irrigation
- I needed more information than was provided
- I found questions 2 to 6 confusing
- I thought the proposal to purchase water from the farmers would work
- I don’t trust the government to make the increase in water rates one-off
Focus groups are a vital part of any CM application. They are useful in determining which attributes should be included in choice sets, what information should be included in questionnaires, for trialing alternative choice set formats and for detecting the existence of bias or other problems. This is especially the case given the limited experience of using CM to estimate non-market and, particularly, non-use values.

The importance and iterative nature of the focus group process are evident from this chapter. The questionnaire used in the Macquarie Marshes focus groups suffered from a number of significant problems including bias, confusion and implausibility. It was apparent that these problems were causing many participants to use 'indicator attributes' in making choices, which is a type of heuristic. Using the information gathered in the Macquarie Marshes focus groups, strategies for dealing with these problems were developed. These strategies appeared to be successful so that the Gwydir Wetlands questionnaire was generally less problematic. This demonstrates how careful focus grouping and careful questionnaire design can be used to minimise the use of inappropriate heuristics. Nonetheless, some respondents still experienced problems from confusion. Further changes were made to the questionnaires to address this problem.

The design of a CM questionnaire is an exercise involving many constraints. For example, the selection of relevant background information, the development of a plausible payment scenario and the selection of attributes depends not only on what is relevant to respondents, but also on the policy context and respondents' capacity to digest the information. As a result, there are many compromises involved in designing a CM questionnaire. Frequently, assessments must be made about the relevance of a particular piece of information to the majority of respondents, and whether the information is likely to be understood. Focus
groups are a useful vehicle providing answers to many of these questions and therefore for designing CM questionnaires.
Chapter 7 Survey Results

7.1 Introduction

In Chapter 5, two case study sites were described: the Gwydir Wetlands and the Macquarie Marshes. The development of CM questionnaires for use in estimating the value of improving the quality of these two wetlands was then described in Chapter 6. In this chapter the results from three CM surveys are presented and evaluated. The first two surveys were for the Gwydir Wetlands case study and were conducted in Moree and Sydney. Moree is a rural town centre about 70 kilometres from the Gwydir Wetlands, while Sydney is a major urban centre about 600 kilometres away. The third survey focused on the Macquarie Marshes and was conducted in Sydney.

The structure of the chapter is as follows. In Section 7.2 the survey logistics are described, and in Section 7.3 the survey statistics are reported. The results from MNL models estimated for each of the surveys are then presented in Section 7.4. The basic formulation of the MNL model was found to suffer from violations of the IIA property using a mother logit test for each survey. Models that allowed for heterogeneous preferences were then estimated and significant IIA violations were not found. Two hypotheses are tested in Section 7.5. The first is whether CM can be used to produce valid estimates of non-use values, and the second is whether non-use values exist for non-environmental attributes. Support is found for the first of these hypotheses, but support for the second hypothesis is mixed. Conclusions are offered in Section 7.6.
Three separate surveys were conducted. For the Gwydir Wetlands, questionnaires were distributed in Sydney and in Moree. For the Macquarie Marshes, questionnaires were distributed only in Sydney. The surveys were conducted in these locations, rather than throughout New South Wales, to permit testing of the benefit transfer hypotheses which are central to this thesis. A drop-off and pick-up survey technique was used because of concerns that low response rates would be achieved if a mail survey was used. Personal interviews were not used because of the expense involved, and because it was thought that respondents may need additional time to consider the options involved than is generally possible in a personal interview.

The questionnaires for the Gwydir Wetlands survey in Moree\textsuperscript{56} were distributed on the weekend of the 28 and 29 June 1997. Three hundred drop-off and pick-up questionnaires were distributed in Moree by five interviewers. One of these was an Aboriginal woman who delivered the questionnaires to the predominantly Aboriginal areas in Moree. The sample of respondents was selected from across the entire area of Moree.

For the Gwydir Sydney survey, 302 questionnaires were delivered between the 12-14 July, which included the weekend. Six interviewers distributed the survey, including four of the five interviewers who had distributed the survey in Moree. The surveys were delivered in 18 different areas throughout Sydney using a cluster sampling technique. Because of the minor nature of the changes made to the pretest version, the questionnaires collected in the Sydney pretest were also included in the final sample. The pretest was conducted at about the same time as the final survey.

\textsuperscript{56} This is referred to hereafter as the Gwydir Moree survey; the corresponding survey in Sydney is referred to as the Gwydir Sydney survey.
The questionnaires for the Macquarie Marshes survey were distributed on the weekend of the 11 and 12 October 1997. Four hundred and sixteen drop-off and pick-up questionnaires were distributed in Sydney by nine interviewers in 18 areas.

For each survey, interviewers were provided with an outline of an introduction to use when speaking to potential respondents. It briefly outlined what the survey was about and the time it would take to complete the questionnaire. A cluster sampling approach was used to distribute the questionnaires. Interviewers were given a set number of questionnaires to distribute in pre-allocated areas. They were asked to select randomly a household within each area. If nobody was home this was noted, and if somebody was not willing to do the questionnaire the reason why was noted. They then went to the next house. After a survey was left at a house, the interviewers were instructed to leave three houses before approaching another house. The questionnaires were picked up later that day or the next day. Three further attempts were made to collect questionnaires if they were not completed. On the third attempt a reminder slip was left to post the questionnaire back using a reply paid envelope. The interviewers were required to keep a record of each house they visited.

The response rates for the surveys are listed in Table 7.1. The response rates from the three surveys are broadly similar. With the drop-off and pick-up format a high return rate for surveys distributed was achieved (>75%). However, when rejections are included the response rate falls to 45-50%. The inclusion of respondents who are not home, which is the most exacting measure, yields a response rate of about 22%.
Table 7.1: Sampling statistics

<table>
<thead>
<tr>
<th></th>
<th>Gwydir - Moree</th>
<th>Gwydir - Sydney</th>
<th>Macquarie - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of questionnaires</td>
<td>301</td>
<td>349</td>
<td>416</td>
</tr>
<tr>
<td>distributed (1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final (useable) data set (2)</td>
<td>233</td>
<td>294</td>
<td>318</td>
</tr>
<tr>
<td>Number of houses with</td>
<td>606</td>
<td>753</td>
<td>693</td>
</tr>
<tr>
<td>nobody home (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of rejections (4)</td>
<td>162</td>
<td>259</td>
<td>233</td>
</tr>
<tr>
<td>Response rate 1</td>
<td>77.4%</td>
<td>84.2%</td>
<td>76.4%</td>
</tr>
<tr>
<td>Response rate 2</td>
<td>50.3%</td>
<td>48.4%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Response rate 3</td>
<td>21.8%</td>
<td>21.6%</td>
<td>23.7%</td>
</tr>
</tbody>
</table>

Note: response rate 1 is based on the number of surveys distributed ie \([2 - r - l]\); response rate 2 is based on the number of surveys distributed and the number of rejections ie \([2 - (1 + 4)]\); response rate 3 includes the number of surveys distributed, rejections and people not home ie \([2 - (1 + 4 + 3)]\).

Reasons given by respondents for not doing the survey are shown in Table 7.2. The most common reason for not participating was ‘lack of interest’ (50-60%), followed by being ‘too busy’ (25-30%).

Table 7.2: Reasons for non-response

<table>
<thead>
<tr>
<th></th>
<th>Gwydir - Moree</th>
<th>Gwydir - Sydney</th>
<th>Macquarie - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too busy</td>
<td>44 (27%)</td>
<td>63 (24%)</td>
<td>74 (32%)</td>
</tr>
<tr>
<td>Not interested</td>
<td>79 (49%)</td>
<td>143 (55%)</td>
<td>137 (59%)</td>
</tr>
<tr>
<td>Too old/too sick</td>
<td>15 (9%)</td>
<td>23 (9%)</td>
<td>11 (5%)</td>
</tr>
<tr>
<td>Illiterate/can’t speak English</td>
<td>10 (6%)</td>
<td>24 (9%)</td>
<td>9 (4%)</td>
</tr>
<tr>
<td>Insufficient knowledge</td>
<td>5 (3%)</td>
<td>5 (2%)</td>
<td>1 (0%)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (2%)</td>
<td>24 (9%)</td>
<td>1 (0%)</td>
</tr>
<tr>
<td>Total</td>
<td>162</td>
<td>259</td>
<td>233</td>
</tr>
</tbody>
</table>

7.3 Survey Statistics

The socio-demographics of the survey samples are presented in this section (see Table 7.3). The socio-demographics of the two Sydney samples are close to the Sydney average, except for income. In part the difference in income can be explained by the inclusion in the census of several regional centres that have lower average income levels than the Sydney average. It is possible, though, that part of the difference is due to some respondents not reporting their income, which is a common problem with SP surveys (about 25% of respondents did not report their income)^57.

^57 Assuming that people who don’t report their income are not randomly distributed.
It may also reflect sampling or non-response bias. Chi-squared tests of independence were conducted to determine whether the Gwydir Sydney and Macquarie Marshes samples had the same socio-demographics as the Sydney population. Except for income (both samples) and age (Gwydir Sydney sample only), the null hypothesis of no independence could not be rejected. Hence, apart from income and age, both Sydney samples are representative of the Sydney population.

Table 7.3: Socio-demographics of the survey respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Gwydir-Moree</th>
<th>Gwydir-Sydney</th>
<th>Macquarie-Sydney</th>
<th>Sydney average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (&gt;17 years)</td>
<td>41.7 years</td>
<td>44.1 years</td>
<td>44.3 years</td>
<td>43.9 years</td>
</tr>
<tr>
<td>Sex (% male)</td>
<td>59.4%</td>
<td>55.0%</td>
<td>55.8%</td>
<td>49.2%</td>
</tr>
<tr>
<td>Children (%)</td>
<td>74.6%</td>
<td>76.7%</td>
<td>72.1%</td>
<td>67.0%*</td>
</tr>
<tr>
<td>Own house (%)</td>
<td>54.9%</td>
<td>75.7%</td>
<td>71.3%</td>
<td>67.4%</td>
</tr>
<tr>
<td>Education (% &gt; year 12)</td>
<td>70.1%</td>
<td>74.6%</td>
<td>74.8%</td>
<td>77.4%</td>
</tr>
<tr>
<td>Income (household)</td>
<td>$48,127</td>
<td>$51,978</td>
<td>$54,680</td>
<td>$46,184</td>
</tr>
<tr>
<td>Employed full or part</td>
<td>71.8%</td>
<td>65.6%</td>
<td>65.7%</td>
<td>59.3%</td>
</tr>
<tr>
<td>time (%) (&gt; 18 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% pro-environment</td>
<td>23.9%</td>
<td>39.5%</td>
<td>35.5%</td>
<td>n.a.</td>
</tr>
<tr>
<td>% pro-development</td>
<td>26.7%</td>
<td>10.0%</td>
<td>10.3%</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

State average; n.a—information not available; own house - respondents who own or are paying off their house; pro-environment - tend to favour the environment more frequently in environment/development conflicts; pro-development - tend to favour development more frequently in environment/development conflicts.

* Source: ABS 1996 Census data

### 7.4 Multinomial Logit Models

The results from the three surveys are reported in this section. Several different MNL models are estimated for each of the surveys. The first model shows only the importance of choice set attributes in explaining respondents’ choices. The second model also includes both socioeconomic and attitudinal variables. The third model includes choice set attributes, socioeconomic and attitudinal variables, and variables based on evaluations of the questionnaires. Mother logit models are estimated

---

58 Earlier versions of the results in this section were reported in Morrison, Bennett and Blamey (1997b) and Morrison, Bennett and Blamey (1998). The latter has been accepted for publication in the journal Water Resources Research.

59 All models were estimated using LIMDEP 7.0 (Greene 1993).
for each model specification to test for the existence of violations of the IIA property. Definitions for the variables used for these models are presented in Table 7.4.

### Table 7.4: Definitions of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_2, C_3$</td>
<td>Alternative specific constants for options 2 and 3</td>
</tr>
<tr>
<td>INCOME</td>
<td>Household income</td>
</tr>
<tr>
<td>INCOME DUMMY</td>
<td>Dummy variable set to 1 if respondents did not report their income</td>
</tr>
<tr>
<td>HOUSE</td>
<td>Dummy variable set to 1 if respondents own or are paying off their home</td>
</tr>
<tr>
<td>CHILD</td>
<td>Dummy variable set to 1 if respondents have children</td>
</tr>
<tr>
<td>VISIT</td>
<td>Dummy variable set to 1 if a respondent is intending to visit the wetlands in the future</td>
</tr>
<tr>
<td>PRODEV</td>
<td>Dummy variable set to 1 if a respondent is pro-development</td>
</tr>
<tr>
<td>PROGRE</td>
<td>Dummy variable set to 1 if a respondent is pro-environment</td>
</tr>
<tr>
<td>UNDER</td>
<td>Likert scale showing whether respondents understood the information in the questionnaire (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>BIASWET</td>
<td>Likert scale showing whether respondents thought the questionnaire was biased towards the wetlands (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>BIASIRR</td>
<td>Likert scale showing whether respondents thought the questionnaire was biased towards irrigation (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>MOREINFO</td>
<td>Likert scale showing whether respondents thought that they needed more information to answer the questionnaire (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>CONFUSED</td>
<td>Likert scale showing whether respondents found answering the choice set questions confusing (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>WILLWORK</td>
<td>Likert scale showing whether respondents believed the scenario would work (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>ONE-OFF</td>
<td>Likert scale showing whether respondents believed that payment would be one-off (1=strongly agree, 5=strongly disagree)</td>
</tr>
<tr>
<td>RATES</td>
<td>Water rates</td>
</tr>
<tr>
<td>JOBS</td>
<td>Irrigation related employment</td>
</tr>
<tr>
<td>AREA</td>
<td>Wetlands area</td>
</tr>
<tr>
<td>BREED</td>
<td>Frequency of waterbird breeding</td>
</tr>
<tr>
<td>SPECIES</td>
<td>Number of endangered and protected species present</td>
</tr>
</tbody>
</table>
The first model estimated using the data from each of the three surveys contains only the choice set attributes (RATES, JOBS, AREA, BREED, SPECIES) and two alternative specific constants (C₁ and C₂). There are three indirect utility functions derived from the MNL model, as shown below. Each of these are simple additive functions. The first function represents the alternative to continue the current situation. The second and third functions represent the indirect utility associated with alternatives that yield improved wetland quality. Respondents who chose the fourth alternative, to increase water to irrigation, were recoded as being in favour of continuing the current situation⁶⁰.

\[
V_1 = \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES}
\]
\[
V_2 = C_2 + \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES}
\]
\[
V_3 = C_3 + \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES}
\]

The sign for the coefficient for RATES should be negative, as increases in water rates imply a decrease in income and hence utility. The signs of the coefficients for JOBS, AREA and SPECIES are expected \textit{a priori} to be positive as an increase in the quantity of each of these variables should be desirable to respondents. The coefficient for BREED, however, should be negative as respondents are expected to prefer more frequent waterbird breeding events.

The results for this model specification are shown in Table 7.5. The coefficients for three of the attributes (RATES, BREED and SPECIES) are

---

⁶⁰ The fourth alternative was included as a means of reducing perceptions of bias (see Appendix 4). However, the attributes for this alternative were not defined, and hence respondents who chose this alternative were recoded as being in favour of choosing the current situation. It is possible to include an alternative specific constant to represent the fourth alternative (e.g. Morrison, Bennett and Blamey 1998), although the estimation of nested logit models as described in Appendix 5 proved to be more difficult when a constant was included for the fourth alternative. Tests demonstrated that implicit prices under the 3 and 4 alternative approaches were statistically equivalent when MNL models were used.
significant at least at the 5% level and have the *a priori* expected sign in each of the models. This indicates that respondents are willing to pay to ensure that extra endangered and protected species are present at each of the wetlands, and for increased frequency of waterbird breeding. Except in the Macquarie Marshes model, the coefficients for the two other attributes (JOBS and AREA) are insignificant, and the coefficient for AREA does not have the expected sign. Hence, only for the Macquarie Marshes is there evidence that respondents are willing to pay to preserve extra wetland area and existing rural jobs.

Only one of the alternative specific constants, for the Macquarie Marshes model, is significant. This indicates that the models are likely to suffer from IIA violations. The alternative specific constants were noted to capture the mean of the error terms in Chapter 4. The mean of the errors, and hence the alternative specific constants, should be approximately equal in a generic choice set if the IIA property holds. Two-sample t-tests indicated that the means of the alternative specific constants were significantly different for each of the models presented below and later in the chapter. Hence the alternative specific constants have not been constrained to be equal.

Each of the models is significant at the 1% level, as shown by the chi-squared statistic. The explanatory power of both of the Gwydir models is relatively low, with an adjusted rho squared of about 0.05. The

---

61 A relevant question is why all of the choice set attributes are significant in the Macquarie Marshes model, but not in the Gwydir models. This may partly reflect the large sample size in the Macquarie Marshes survey and differences in preferences. Part of the cause, though, appears to be differences in experimental design. As discussed in Section 6.4, in the Macquarie Marshes survey an extra block of dominated alternatives was included. In Appendix 5, it is shown that when this extra block is deleted the coefficient for JOBS is insignificant in the Macquarie Marshes model, and the coefficient for AREA is only significant at the 10% level. However, all of the coefficients remain highly significant if any of the other blocks are deleted. One explanation for this result is that dominated alternatives are easy for respondents to answer, thereby reducing variance. As this result is only indirectly related to the topic of this thesis it will be the subject of further research. The main implication, however, is that dominated alternatives should not be deleted from experimental designs.

62 It is possible, however, that some of the difference between the alternative specific constants is due to ordering effects.
explanatory power of the Macquarie Marshes model is more satisfactory, with a rho squared of 0.11. However, the explanatory power of none of these models could be considered exceptional. As detailed in Chapter 4, Hensher and Johnson (1981) suggest that a rho-squared of between 0.2-0.4 is considered to be very good.

A mother logit model was estimated for each of the models to test for the existence of violations of the IIA property (see Appendix 5). As discussed in Section 4.6, the mother logit allows a general pattern of dependence among the unobserved attributes of alternatives, and therefore can be used to test for violations of the IIA property. The null hypothesis is that the MNL is the more accurate model, while the alternative hypothesis is that the mother logit is the more accurate model. The null hypothesis was rejected at the 5% level for the Gwydir Moree model, but could not be rejected at the 1% level. This indicates that the Gwydir Moree model has moderate but not severe violations of the IIA property. However, the null hypothesis was rejected at the 1% level for the Gwydir Sydney and Macquarie Marshes models, indicating that in these two models there are severe violations of the IIA property.
Table 7.5: Multinomial logit models with choice set attributes

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1: Gwydir</th>
<th>Model 2: Moree</th>
<th>Model 3: Sydney</th>
<th>Model 4: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>-0.294</td>
<td>0.160</td>
<td>-0.560***</td>
<td>-0.560***</td>
</tr>
<tr>
<td></td>
<td>(0.201)</td>
<td>(0.172)</td>
<td>(0.192)</td>
<td>(0.192)</td>
</tr>
<tr>
<td>C3</td>
<td>0.089</td>
<td>0.434***</td>
<td>0.247</td>
<td>0.247</td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td>(0.174)</td>
<td></td>
<td>(0.193)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.876E-2***</td>
<td>-0.112E-1***</td>
<td>-0.127E-1***</td>
<td>-0.127E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.102E-2)</td>
<td>(0.931E-3)</td>
<td>(0.824E-3)</td>
<td></td>
</tr>
<tr>
<td>JOBS</td>
<td>0.123E-3</td>
<td>0.158E-2</td>
<td>0.178E-2***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.122E-2)</td>
<td>(0.109E-2)</td>
<td>(0.663E-3)</td>
<td></td>
</tr>
<tr>
<td>AREA</td>
<td>-0.167E-3</td>
<td>-0.380E-5</td>
<td>0.538E-3***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.319E-3)</td>
<td>(0.279E-3)</td>
<td>(0.129E-3)</td>
<td></td>
</tr>
<tr>
<td>BREED</td>
<td>-0.127**</td>
<td>-0.096**</td>
<td>-0.312***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.049)</td>
<td>(0.052)</td>
<td></td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.038***</td>
<td>0.034***</td>
<td>0.048***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td></td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 1: Gwydir</th>
<th>Model 2: Moree</th>
<th>Model 3: Sydney</th>
<th>Model 4: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-1159.135</td>
<td>-1476.999</td>
<td>-1531.307</td>
<td></td>
</tr>
<tr>
<td>(\chi^2) (constants only)</td>
<td>100.600</td>
<td>182.951</td>
<td>380.620</td>
<td></td>
</tr>
<tr>
<td>(\rho^2) adjusted</td>
<td>0.046</td>
<td>0.057</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>Iterations completed</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1109</td>
<td>1429</td>
<td>1575</td>
<td></td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level
standard errors are in brackets

7.4.2 MNL models with socioeconomic and attitudinal variables

One way of removing violations of the IIA property is to include socioeconomic and attitudinal variables in the model\(^{63}\). This is likely to remove or reduce the extent of the IIA violations if it is primarily due to random taste variations (see Section 4.5). This has been the finding of other studies (eg Swait, Erdem, Louviere and Dubelaar 1992). The use of both socioeconomic and attitudinal variables is justified under the hypothesis that attitudes and socioeconomic characteristics are separate

\(^{63}\) Attempts were also made to use more complex models to deal with IIA violations. The multinomial probit was trialed, but there was difficulty with model convergence. The heteroscedastic extreme value model was more promising, as was the nested logit for the Macquarie Marshes data set (see Appendix 5). But with the Gwydir data sets there appeared to be problems due to inadequate sample size. Given that it was possible to minimise IIA violations by allowing for heterogeneity, as described shortly, following Occum’s Razor the MNL model has been used.
factors influencing behavioural intentions and behaviour (Lynne, Shonkwiler and Rola 1988)

The specification for this model is shown below. Socioeconomic and attitudinal variables (CHILD, HOUSE, INCOME, INCOME DUMMY, VISIT, PRODEV, PROGRE) have been included through interactions with the alternative specific constants. These interactions show the effect of various attitudes and socioeconomic characteristics on the probability that a respondent will choose either option 2 or 3. It is possible, however, to interact these variables with choice set attributes as has been done in some earlier modelling exercises (eg Morrison, Bennett and Blamey 1998). Attempts were made to estimate models that include polynomials for each of the variables representing the choice set attributes and age, however none of the coefficients for these additional variables proved to be significant.

\[ V_1 = \beta_1 \text{RATE} + \beta_2 \text{RATES} + \beta_3 \text{JOBS} + \beta_4 \text{AREA} + \beta_5 \text{BREED} + \beta_6 \text{SPECIES} \]
\[ V_2 = C_2 + C_{\text{CHILD}} + C_{\text{HOUSE}} + C_{\text{INCOME}} + C_{\text{INCOME DUMMY}} + C_{\text{VISIT}} + C_{\text{PRODEV}} + C_{\text{PROGRE}} + \beta_1 \text{RATE} + \beta_2 \text{RATES} + \beta_3 \text{JOBS} + \beta_4 \text{AREA} + \beta_5 \text{BREED} + \beta_6 \text{SPECIES} \]
\[ V_3 = C_3 + C_{\text{CHILD}} + C_{\text{HOUSE}} + C_{\text{INCOME}} + C_{\text{INCOME DUMMY}} + C_{\text{VISIT}} + C_{\text{PRODEV}} + C_{\text{PROGRE}} + \beta_1 \text{RATE} + \beta_2 \text{RATES} + \beta_3 \text{JOBS} + \beta_4 \text{AREA} + \beta_5 \text{BREED} + \beta_6 \text{SPECIES} \]

An issue that arises in the specification of these models is how income should be included. In the above model income is included as an interaction with the alternative specific constants. The coefficient shows the extent to which people with higher income are more likely to choose a non-base alternative. The coefficient should not be interpreted as the marginal utility of income. This approach for including income has been

64 There is some debate surrounding the use of both attitudes and socioeconomic factors as explanatory variables (see Blamey, Common and Quiggin 1995; Rolfe and Bennett 1996a). There appeared to be little evidence of multicollinearity for this model specification.

65 A dummy variable for income has been included to minimise data loss due to respondents not reporting their income. If respondents did not report their income they were recoded as having zero income. Thanks are due to Professor Jordan Louviere for suggesting this approach.
used in a number of previous CM applications (eg Eom 1994, Morley 1994, Stevens et al 1997). An alternative approach is to calculate a new variable by subtracting the change in water rates from income. This variable would be included in place of both water rates and income. An advantage of this approach is that it is more consistent with the way that compensating surplus is estimated in practice (see Morey et al 1993). However, previous studies have found this variable to be insignificant (eg Roe et al 1996) and more substantial problems with missing data may be experienced because of the large number of respondents who do not report their income.

Theory provides some guidance in terms of the expected signs of several of the above variables. PROGRE should have a positive sign as respondents with a pro-environmental orientation would be expected to choose option 2 or 3 more frequently. Similarly, PRODEV should have a negative sign. VISIT should have a positive sign if respondents intend to visit the Gwydir Wetlands in the future or have positive option value. INCOME should have a positive sign as respondents with higher income should have a greater capacity to pay. The sign for CHILD is, however, ambiguous. Bequest motives would be expected to induce higher willingness to pay, yielding a positive coefficient; however, households with children may have lower disposable income, thereby lowering willingness to pay. The sign of HOUSE is also ambiguous. The coefficient could either be positive or negative depending on whether respondents have paid off their house and their subsequent capacity to pay. It is also possible that respondents who don’t own their own house will free-ride because they do not think that increases in water rates would be passed on. This would result in a negative coefficient.

The results for this model are shown in Table 7.6. Similar to the previous model specification, three choice set attributes (RATES, BREED and

---

66 Because of the interactions with the alternative specific constants, socioeconomic variables can only be included for J-1 alternatives.
SPECIES) are significant at the 5% level or better in each of the three models. Except in the Macquarie Marshes model, neither JOBS nor AREA are significant. The insignificance of the coefficient for JOBS in the Gwydir Moree model is surprising given that this is where the impacts on employment would occur.

Both of the alternative specific constants are significant in the Macquarie Marshes model, and one alternative specific constant is significant in each of the Gwydir models. In each of the models (and especially the Macquarie Marshes model) the percentage difference in the magnitude of the constants is less, suggesting that IIA violations may have been reduced. However, as mentioned earlier, 2 sample t-tests indicated that the means of the alternative specific constants in each of the models were different, suggesting that it is not appropriate to constrain the alternative specific constants to be equal.

The coefficients for PRODEV and PROGRE are significant at the 1% level, and have the \textit{a priori} expected signs in each of the three models. The coefficient for PRODEV is larger than the coefficient for PROGRE in both of the Gwydir models. This shows that having a pro-development disposition has a much greater impact on respondents’ choices than having a pro-environmental disposition. However, the opposite is true in the Macquarie Marshes model. The coefficient for VISIT is positively signed and significant in each model. This provides evidence that respondents have both use and/or option values for improved wetland quality\textsuperscript{68}. The coefficient for CHILD is significant in each of the three models; however, the sign differs across the models. The sign for CHILD is positive in the Macquarie Marshes and Gwydir Moree models, indicating that bequest motives outweigh reduced capacity to pay. However, the sign is negative.

\textsuperscript{67} Note that having children is not the only reason that people may have bequest motives. \textsuperscript{68} As discussed in Chapter 2, from a supply-side perspective option value should be positive for the case where there is possible supply if the project doesn’t go ahead, but certain supply if it does go ahead. This is Freeman’s (1985) and Plummer’s (1986) ‘Case B’. However, if there is demand uncertainty in addition to supply uncertainty the sign of option value is indeterminate theoretically.
in the Gwydir Sydney model. It is possible that the difference between the
two Sydney models could reflect the greater environmental significance of
the Macquarie Marshes, and hence that it has larger bequest values. The
coefficient for HOUSE is significant in both Gwydir models only. The
coefficient is positive for the Gwydir Sydney survey, showing that home
ownership is associated with increased willingness to pay, but the sign is
negative for Moree, showing that home ownership is associated with
reduced willingness to pay. The coefficient for INCOME is only
significant for the two Sydney models, and is correctly signed. This
coefficient shows that respondents with higher income are more likely to
choose the option to improve wetland quality. The coefficient for the
INCOME DUMMY, however, was significant and negative in the Gwydir
Moree model, showing that respondents who didn’t report their income
were, on average, less likely to choose to improve wetland quality.

Each of the models is significant overall and has slightly higher
explanatory power than for the previous model specification. The
explanatory power of the Gwydir Moree model is still relatively low with
an adjusted rho squared of 0.07; the Gwydir Sydney model is satisfactory
with an adjusted rho squared of 0.11; and the Macquarie Marshes model is
good with an adjusted rho squared of 0.19.

Mother logit models were again estimated to test for the existence of IIA
violations (see Appendix 5). The null hypothesis that the MNL was the
more accurate model could not be rejected at the 5% level for the Gwydir
Moree and Macquarie Marshes models. However, the null hypothesis was
rejected at the 5% level, but not at the 1% level for the Gwydir Sydney
model. Hence IIA violations were not found to be significant in the
Gwydir Moree and Macquarie Marshes models after the inclusion of
socioeconomic and attitudinal variables, but moderate violations still
remain in the Gwydir Sydney model.
Table 7.6: Multinomial logit models with socioeconomic and attitudinal variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 4: Gwydir Moree</th>
<th>Model 5: Gwydir Sydney</th>
<th>Model 6: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>-0.665** (0.340)</td>
<td>-0.437* (0.260)</td>
<td>-1.657*** (0.270)</td>
</tr>
<tr>
<td>C3</td>
<td>-0.308 (0.337)</td>
<td>-0.138 (0.262)</td>
<td>-1.315*** (0.270)</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.843*** (0.179)</td>
<td>-0.741*** (0.182)</td>
<td>0.396*** (0.151)</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.461*** (0.159)</td>
<td>0.616*** (0.173)</td>
<td>-0.112 (0.163)</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.262E-5 (0.247E-5)</td>
<td>0.643E-5*** (0.225E-5)</td>
<td>0.941E-5*** (0.241E-5)</td>
</tr>
<tr>
<td>INCOMEDUMMY</td>
<td>-0.564*** (0.215)</td>
<td>0.196 (0.205)</td>
<td>-0.235 (0.208)</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.485*** (0.198)</td>
<td>0.671*** (0.138)</td>
<td>0.890*** (0.136)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.461*** (0.185)</td>
<td>0.720*** (0.147)</td>
<td>1.157*** (0.162)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-0.887*** (0.268)</td>
<td>-1.53*** (0.209)</td>
<td>-0.922*** (0.196)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.859E-2*** (0.117E-2)</td>
<td>-0.119E-1*** (0.102E-2)</td>
<td>-0.137E-1*** (0.922E-3)</td>
</tr>
<tr>
<td>JOBS</td>
<td>-0.637E-3 (0.140E-2)</td>
<td>0.177E-2 (0.120E-2)</td>
<td>0.191E-2*** (0.729E-3)</td>
</tr>
<tr>
<td>AREA</td>
<td>-0.537E-3 (0.363E-3)</td>
<td>0.249E-3 (0.301E-3)</td>
<td>0.547E-3*** (0.140E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.128** (0.064)</td>
<td>-0.109** (0.052)</td>
<td>-0.321*** (0.055)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.032*** (0.013)</td>
<td>0.036*** (0.010)</td>
<td>0.050*** (0.011)</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>$\chi^2$ (constants only)</th>
<th>$\rho^2$ adjusted</th>
<th>Iterations completed</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-883.438</td>
<td>138.511</td>
<td>0.070</td>
<td>4</td>
<td>872 (237 skipped*)</td>
</tr>
<tr>
<td></td>
<td>-1233.092</td>
<td>318.666</td>
<td>0.111</td>
<td>4</td>
<td>1269 (160 skipped)</td>
</tr>
<tr>
<td></td>
<td>-1244.856</td>
<td>564.255</td>
<td>0.185</td>
<td>5</td>
<td>1397 (178 skipped)</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level standard errors are in brackets; * the observations are skipped because of missing socioeconomic data.
One further model was estimated using the data from each of the three surveys. The model specification was identical to the previous model except that seven variables relating to questionnaire evaluation were included (UNDER, BIASWET, BIASIRR, MOREINFO, CONFUSED, WILLWORK and ONEOFF). These variables are included to show the effect of respondents' perceptions about the questionnaire on their choices. For instance, if respondents act conservatively, it would be expected that doubts about the questionnaire would result in a reduced likelihood to choose options 2 or 3. Hence the sign for UNDER and WILLWORK should be negative, and the sign for BIASWET, BIASIRR, MOREINFO, CONFUSED and ONEOFF should be positive.

As expected, the coefficients for UNDER and WILLWORK is negative and significant, and for ONE-OFF is positive and significant in each of the three surveys. This shows that problems with understanding the survey and doubts about payment being one-off had a consistent effect on the likelihood that respondents would choose to improve wetland quality. The significant coefficient for ONE-OFF also suggests the existence of payment vehicle bias (Morrison, Blamey and Bennett forthcoming). Hence, these are two areas that should be of concern during the survey design process. Two other variables also had reasonably consistent effects on respondents' choices. The coefficient for WILLWORK, as expected, was negatively signed in each model, but was significant in only two models. This suggests that doubts about the plausibility of the scenario can affect willingness to pay estimates. The coefficient for BIASWET was also significant and negative in two models. The coefficients for the remaining variables (BIASIRR, MOREINFO and CONFUSED) were either insignificant or inconsistently signed.
The inclusion of these variables has had several effects on the models. The coefficients for the AREA and JOBS variables are now significant at the 5% level or better in the Gwydir Sydney model. The alternative specific constants are much closer, suggesting that IIA violations may have been removed. However, the coefficient for INCOME is no longer significant, suggesting the existence of multicollinearity.

The inclusion of these variables has also had a relatively large effect on the explanatory power of the models. The adjusted rho squared increased from 0.07 to 0.17 in the Gwydir Moree model. This compares to an increase of 0.02 from the inclusion of the socioeconomic and attitudinal variables. The explanatory power of the Gwydir Sydney model is also better at 0.16. The explanatory power of the Macquarie Marshes model is in the ‘very good’ range, with a rho squared of 0.22. The increase in explanatory power has occurred despite the loss of a relatively large amount of power from skipped observations due to respondents not answering the questionnaire evaluation questions.

A mother logit model was estimated to determine whether IIA violations still existed with the Gwydir Sydney model (see Appendix 5). The null hypothesis that the MNL is the more accurate model could not be rejected, even at the 1% level. Therefore, including these extra questionnaire evaluation variables has been successful in removing IIA violations due to random taste variations. The necessity of using these extra variables emphasises the importance of including extra debrief questions in CM questionnaires in addition to the usual socioeconomic and attitudinal variables.
Table 7.7: Multinomial logit models with socioeconomic, attitudinal and questionnaire evaluation variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 7: Gwydir Moree</th>
<th>Model 8: Gwydir Sydney</th>
<th>Model 9: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>-0.205</td>
<td>2.378***</td>
<td>-3.295***</td>
</tr>
<tr>
<td></td>
<td>(0.813)</td>
<td>(0.775)</td>
<td>(0.614)</td>
</tr>
<tr>
<td>C3</td>
<td>0.124</td>
<td>2.671***</td>
<td>-2.961***</td>
</tr>
<tr>
<td></td>
<td>(0.812)</td>
<td>(0.776)</td>
<td>(0.615)</td>
</tr>
<tr>
<td>CHILD</td>
<td>1.145***</td>
<td>-3.733***</td>
<td>0.423***</td>
</tr>
<tr>
<td></td>
<td>(0.224)</td>
<td>(1.528)</td>
<td>(0.169)</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.239</td>
<td>0.422*</td>
<td>-0.515E-2</td>
</tr>
<tr>
<td></td>
<td>(0.209)</td>
<td>(0.217)</td>
<td>(0.189)</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.200E-5</td>
<td>0.229E-5</td>
<td>0.711E-5***</td>
</tr>
<tr>
<td></td>
<td>(0.311E-5)</td>
<td>(0.270E-5)</td>
<td>(0.270E-5)</td>
</tr>
<tr>
<td>INCOMEDUMMY</td>
<td>-0.866***</td>
<td>0.434*</td>
<td>-0.248***</td>
</tr>
<tr>
<td></td>
<td>(0.299)</td>
<td>(0.263)</td>
<td>(0.249)</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.524**</td>
<td>0.596***</td>
<td>0.822***</td>
</tr>
<tr>
<td></td>
<td>(0.256)</td>
<td>(0.171)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>1.048***</td>
<td>0.713***</td>
<td>1.217***</td>
</tr>
<tr>
<td></td>
<td>(0.257)</td>
<td>(0.181)</td>
<td>(0.183)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-1.588***</td>
<td>-1.17***</td>
<td>-0.736***</td>
</tr>
<tr>
<td></td>
<td>(0.357)</td>
<td>(0.248)</td>
<td>(0.224)</td>
</tr>
<tr>
<td>UNDER</td>
<td>-0.569***</td>
<td>-0.508***</td>
<td>-0.246***</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.110)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>BIASWET</td>
<td>0.360***</td>
<td>-0.023</td>
<td>0.420***</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.085)</td>
<td>(0.083)</td>
</tr>
<tr>
<td>BIASIRR</td>
<td>0.166</td>
<td>-0.167</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.107)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>MOREINFO</td>
<td>0.279***</td>
<td>-0.031</td>
<td>-0.218***</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.085)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>CONFUSED</td>
<td>-0.371***</td>
<td>-0.024</td>
<td>0.285***</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.084)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>WILLWORK</td>
<td>-0.591***</td>
<td>-0.365***</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.088)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>ONEOFF</td>
<td>0.392***</td>
<td>0.249***</td>
<td>0.338***</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.075)</td>
<td>(0.071)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.109E-1***</td>
<td>-0.139E-1***</td>
<td>-0.148E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.830E-1)</td>
<td>(0.115E-2)</td>
<td>(0.103E-2)</td>
</tr>
<tr>
<td>JOBS</td>
<td>-0.675E-3</td>
<td>0.272E-2**</td>
<td>0.284E-2***</td>
</tr>
<tr>
<td></td>
<td>(0.165E-2)</td>
<td>(0.134E-2)</td>
<td>(0.804E-3)</td>
</tr>
<tr>
<td>AREA</td>
<td>-0.129E-3</td>
<td>0.595E-3*</td>
<td>0.657E-3***</td>
</tr>
<tr>
<td></td>
<td>(0.427E-3)</td>
<td>(0.329E-3)</td>
<td>(0.151E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.196***</td>
<td>-0.140***</td>
<td>-0.345***</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.057)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.046***</td>
<td>0.045***</td>
<td>0.055***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 7: Gwydir Moree</th>
<th>Model 8: Gwydir Sydney</th>
<th>Model 9: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-638.119</td>
<td>-994.351</td>
<td>-1037.286</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>272.222</td>
<td>364.968</td>
<td>601.550</td>
</tr>
<tr>
<td>( \rho^2 )</td>
<td>0.167</td>
<td>0.155</td>
<td>0.224</td>
</tr>
<tr>
<td>Iterations</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Observations</td>
<td>707 (402 skipped)</td>
<td>1081 (348 skipped)</td>
<td>1227 (338 skipped)</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level
standard errors are in brackets
7.5 Hypothesis Tests

Several hypotheses were outlined in Chapter 1. The first of these was that CM produces valid estimates of non-use values. While this hypothesis has previously been tested for the CVM (Bennett 1981), it has not been formally tested for CM. The second hypothesis tested in this chapter is that non-use values exist for non-environmental outcomes.

Hypothesis 1: choice modelling produces valid estimates of non-use values

The validity of non-use values estimated using CM can be demonstrated using several different measures. Mitchell and Carson (1989) outline three main measures of validity. The first is content validity. In the context of CM, content validity reflects whether the information in the questionnaire is accurate, understandable, meaningful, unbiased, adequate and plausible. The second measure is criterion or predictive validity. Predictive validity is demonstrated when the choices or estimates predicted using CM are equal to choices or estimates revealed in actual market situations. The final measure is construct or theoretical validity. Theoretical validity is demonstrated when a priori relationships between the dependent and certain independent variables hold. It indicates that systematic forms of bias may be limited, as systematic forms of bias are likely to confound theoretical relationships.

Predictive validity has been tested in a number of CM studies which have estimated use values or market share (Louviere and Swait 1996). However, for non-use values it is difficult to demonstrate predictive validity because there are limited market or related market data available. The only opportunity to test predictive validity would be if the government decided to institute environmental referenda\textsuperscript{60}, as are used in some states.

\textsuperscript{60} This would require correspondence between the survey instrument and the referendum instrument.
of the USA (Bennett and Carter 1993) or possibly through experimental or field comparisons (e.g. Sinden 1988). The focus of this thesis is therefore on the remaining two measures of validity: content and theoretical validity.

Content validity

Some evidence of content validity is provided during the questionnaire design process. Responses during focus groups and pretesting indicate in part whether a questionnaire is meaningful. However, such responses do not provide sufficient information with which to conduct statistical tests of a hypothesis. Use is made of the Likert scale evaluation questions included in each of the three questionnaires to test content validity. As detailed in Chapter 6, these questions asked whether respondents (1) understood the information in the questionnaire; (2) thought the information in the questionnaire was biased towards the wetlands; (3) thought the information in the questionnaire was biased towards irrigation; (4) needed more information than was provided; (5) found answering the choice sets confusing; (6) thought the water purchasing scheme would work; and (7) thought the government would make the increase in water rates one-off.

T-tests are conducted to determine whether, on average\(^{70}\), problems with content exist. This involves testing the null hypothesis of whether the mean is less than (or greater than) or equal to three, depending on a priori expectations. Three corresponds to ‘neither agree nor disagree’ on the Likert scale; hence a mean rating of ‘three’ indicates that, on average, respondents did not perceive there to be a specific problem with the questionnaire. If the questionnaires are valid, UNDER and WILL WORK should be less than or equal to three, and the remaining means should be

---

\(^{70}\) These tests show whether, on average, problems were identified with the questionnaires. Averages, however, can be misleading. If the variance is particularly large, one might conclude that a particular issue is not a problem in a questionnaire when many respondents are actually concerned, and vice versa. In general, however, the standard errors are small.
greater than or equal to three. The mean values, standard errors, and t-statistics\(^{71}\) for these tests are reported in Table 7.8.

**Table 7.8: T-statistics for tests of content validity**

<table>
<thead>
<tr>
<th></th>
<th>UNDER</th>
<th>BIASWET</th>
<th>BIASIRR</th>
<th>MORE</th>
<th>CONFUSED</th>
<th>WILLWORK</th>
<th>ONE-OFF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gwydir Moree</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.937</td>
<td>2.704</td>
<td>3.429</td>
<td>2.852</td>
<td>3.171</td>
<td>3.106</td>
<td>1.979</td>
</tr>
<tr>
<td>SE</td>
<td>0.061</td>
<td>0.077</td>
<td>0.067</td>
<td>0.080</td>
<td>0.081</td>
<td>0.090</td>
<td>0.079</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-17.410</td>
<td>-3.823</td>
<td>6.421</td>
<td>-1.845</td>
<td>2.102</td>
<td>1.184</td>
<td>-12.902</td>
</tr>
<tr>
<td><strong>Gwydir Sydney</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.961</td>
<td>2.786</td>
<td>3.496</td>
<td>3.048</td>
<td>3.290</td>
<td>2.820</td>
<td>2.122</td>
</tr>
<tr>
<td>SE</td>
<td>0.049</td>
<td>0.063</td>
<td>0.054</td>
<td>0.065</td>
<td>0.069</td>
<td>0.058</td>
<td>0.072</td>
</tr>
<tr>
<td><strong>Macquarie Marshes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.949</td>
<td>2.778</td>
<td>3.437</td>
<td>3.118</td>
<td>3.417</td>
<td>2.890</td>
<td>2.193</td>
</tr>
<tr>
<td>SE</td>
<td>0.047</td>
<td>0.057</td>
<td>0.045</td>
<td>0.060</td>
<td>0.063</td>
<td>0.057</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Note: 1-strongly agree, 2-agree, 3-neither agree nor disagree, 4-disagree, 5-strongly disagree

Each respondent was asked whether they understood the information in the questionnaire in the first question. This question is important because if respondents didn’t understand the information, their responses are not likely to be meaningful. The mean values should be less than three if respondents, on average, understood the information. In all three surveys the mean values are less than three. The t-statistics shown in Table 7.8 for each questionnaire are greater than the critical value at the 1% level (-2.58). Therefore it can be concluded that for each of the three questionnaires respondents, on average, understood the information presented to them in each questionnaire.

Respondents were asked in the second question whether they thought information in the questionnaire was biased towards the wetlands. As noted in Chapter 6, bias was a fairly significant problem with the draft questionnaires. If respondents thought the questionnaire was unbiased then the mean value should be at least three. The mean value for each

---

\(^{71}\) The use of t-tests requires that the data are drawn from a normally distributed population. Examination of kurtosis and skewness for each of the samples represented in Table 7.8 indicates that departures from normality are likely. Hence the accuracy of these tests is limited. However, it is not possible to use non-parametric tests because of the limited number of points on the Likert scales.
survey is less than three, indicating that the questionnaires may be biased. The t-statistics for each survey are also greater than the critical value at the 1% level. Therefore it is concluded that respondents found, on average, the information biased towards the wetlands.

The third question also focused on bias, but bias towards irrigation. In contrast to the previous question, the mean values are greater than three for each of the questionnaires, and the t-statistics are greater than the critical value at the 1% level. Therefore it is concluded that respondents, on average, did not find the information biased towards irrigation.

The objective of the fourth question was to determine whether respondents felt they needed more information than was presented to them in the questionnaire. Large amounts of information had been deleted from the draft questionnaire to reduce confusion, therefore there was concern that too much information may have been deleted. The mean value should be at least three for the questionnaires to be valid. For the two Sydney surveys the mean value is greater than three, but by a small amount. For the Moree survey, the mean value is less than three. Only the t-statistic for the Macquarie Marshes survey is greater than the critical value at the 5% level. Therefore it is concluded that, on average, respondents from all three surveys felt that they had sufficient information to answer the questionnaires.

The fifth question asked whether respondents found answering the choice sets confusing. As noted in Chapter 6, confusion created by the choice sets was found to be a substantial problem in the focus groups. Validity is demonstrated here by a mean value of at least three. For all three surveys the mean value is greater than three. The t-statistics all show that the null-hypothesis that the mean values are greater than three cannot be rejected at the 5% level. Therefore it is also concluded that, on average, respondents did not find answering the choice sets confusing.
The sixth question focused on whether respondents thought that the water purchasing scheme would work. Respondents in the Moree focus groups expressed concerns about the scheme working. Respondents in the Sydney focus groups, however, thought it would work. Validity is demonstrated for this question by mean values of less than three. For both Sydney questionnaires the mean value is less than three, and for the Gwydir Sydney questionnaire the t-test shows it is significantly less than three. In contrast, the mean value for the Gwydir Moree questionnaire is greater than three. However, the t-statistic is less than the critical value at the 5% level, so it is concluded that respondents from all three surveys, on average, thought that the water purchasing scheme would work.72

The final question focused on whether respondents believed that payment for the scheme would only be one-off. Validity is demonstrated here by a mean value of at least three. For each of the questionnaires the mean values is significantly less than three. Therefore, respondents, on average, did not believe that payment would be one-off.

In summary, for five of the seven questions evidence was found to support the hypothesis of content validity. Hence, most of the evidence favours acceptance of the null hypothesis. A relevant question is what effect have doubts about payment being one-off and the existence of bias had on estimates of non-use values? In Models 7-9 reported in the previous section, the sign for ONEOFF was positively signed and significant at the 1% level in each of the models. This indicates that doubts about payment being one-off were correlated with lower estimates of willingness to pay. The sign for BIASWET was positively signed and significant in the Gwydir Moree and Macquarie Marshes models, but insignificant in the Gwydir Sydney model. This indicates that perceptions of bias were also correlated with lower estimates of willingness to pay in the former two

---

72 For the fourth, fifth and sixth questions, paired t-tests showed that there were significant differences between the responses for the Sydney and Moree surveys. This would indicate that there may be a need to develop separate questionnaires for different
surveys. Hence in these surveys, doubts about payment being one-off and
the existence of bias were not only noted to be a problem, but were found
to have significant effects on value estimates.

Theoretical validity

Theoretical validity is tested using the MNL models reported in Section 7.4. First, the overall models are tested for significance to determine whether the models are explaining the behaviour of respondents. Second, the significance and sign of different independent variables are examined to see if they concur with theory. These include socioeconomic and attitudinal variables, as well as variables that reflect environmental quality and cost to respondents. The significance of the overall models, and the existence of the a priori expected relationships provides evidence of theoretical validity.

The first test is that the MNL model is significant overall. Stated formally, the hypothesis for this test is:

\[ H_0: \beta_1 = \ldots = \beta_n = 0 \]
\[ H_1: \beta_1 = \ldots = \beta_n \text{ are not all } 0 \]

(where \( \beta_1 = \ldots = \beta_n \) are model coefficients other than for the alternative specific constants)

The MNL models that included socioeconomic, attitudinal and questionnaire evaluation variables (models 7, 8 and 9) are used for this test because they were all found to be free of IIA violations. Overall significance can be determined using a likelihood ratio test. As discussed in Chapter 4, Ben-Akiva and Lerman (1985) recommended including the constants for this test. The test statistic is equal to minus two multiplied by the difference between the log-likelihood of the restricted and unrestricted models \([-2(L_R-L_U)]\). This statistic is chi-squared distributed,
with degrees of freedom equal to the number of restrictions on the model. For models 7, 8 and 9 the chi-squared statistics are:

\[
\chi^2(\text{model 7}) = 249.766 \quad \chi^2(\text{model 8}) = 354.324 \quad \chi^2(\text{model 9}) = 601.241
\]

The \( \chi^2_{\alpha=0.01} \) value with 19 degrees of freedom is equal to 36.191. Hence, because the chi-squared values for each of the models exceeds the critical value at the 1% significance level, the null hypothesis is rejected and it is concluded that each of the models is significant overall.

Theoretical validity is also demonstrated by showing that the coefficients for the independent variables have \textit{a priori} expected signs and are significant. The coefficients of most concern are for the choice set attributes. They are important because they are used to generate estimates of non-use values. As discussed in Section 7.4, in each of the three models the coefficients for RATES, BREED and SPECIES were found to be correctly signed and significant at the 5% level or better using t-tests. In both of the Sydney models the coefficients for JOBS and AREA were also found to be correctly signed and significant at the 10% level or better. However, in the Gwydir Moree model the coefficients for JOBS and AREA were insignificant.

Several other independent variables of theoretical importance were also included in the models. These include INCOME, PRODEV and PROGRE. INCOME was only found to be correctly signed and significant in one of the models (model 9). This is of concern, as income is usually considered to be one of the most important variables for demonstrating theoretical validity (Flores and Carson 1995). However, it should be noted that the insignificance of INCOME in model 8 appears to be primarily a result of multicollinearity (see Section 7.4). The coefficients for the remaining variables are correctly signed and significant at the 1% level.
There is some support, therefore, for the null hypothesis of theoretical validity. All of the models were significant overall, and the majority of the independent variables were, as expected, correctly signed and significant. However, there were a few variables, such as INCOME in the Gwydir Moree model, that were not significant. Hence, in terms of both content and theoretical validity there is some support for the hypothesis that CM produces valid estimates of non-use values, however some concerns have been noted which could be further explored in the future.

**Hypothesis 2: non-use values exist for non-environmental outcomes**

The second hypothesis is that non-use values exist for non-environmental outcomes. As discussed in Chapter 1, this hypothesis was suggested by Portney (1994). This hypothesis can be stated formally as follows, where $B_j$ is the coefficient for the JOBS attribute:

$$H_0: \beta_j \text{ is not greater than 0}$$

$$H_1: \beta_j > 0$$

This hypothesis can be tested by using a student’s $t$-test. The $t$-statistics for models 7, 8 and 9 are as follows:

$t$-statistic (model 7) = -0.664  
$t$-statistic (model 8) = 1.986  
$t$-statistic (model 9) = 3.521

The $t_{0.01}$ (one-tailed) value is 2.326. Only for model 9 is the $t$-statistic greater than the critical value at the 1% level. The $t_{0.05}$ (one-tailed) value is 1.645. The $t$-statistic is greater than this critical value for model 8. Therefore, the null hypothesis for model 9 is rejected at the 1% significance level, for model 8 at the 5% significance level, and accepted for model 7.

It is concluded that non-use values can exist for non-environmental attributes, but that they need not exist in all cases.
The results from the three CM surveys were presented in this chapter. The response rate was acceptable in all of the surveys and, for the Sydney samples where comparisons were feasible, the socio-demographics were close to those of the population.

Three models were estimated using each data set. The first was a basic MNL model that included only choice set attributes. Two of the coefficients for the choice sets attributes were significant only in the Macquarie Marshes model. All three basic models had only modest explanatory power and were found to suffer from violations of the property of independence of irrelevant alternatives.

Models that included socioeconomic and attitudinal variables were then estimated to allow for violations of the IIA property. The inclusion of these variables was found to increase substantially the explanatory power of all models; however moderate violations of the IIA property were still found with the Gwydir Sydney model. Models were then estimated that included questionnaire evaluation variables, in addition to socioeconomic and attitudinal variables. The questionnaire evaluation variables provided some indications about the factors related to questionnaire design that affect valuation estimates, and hence should be of concern during the focus group process. The inclusion of these extra variables was successful in eliminating IIA violations in the Gwydir Sydney models. This shows the difficulty of dealing with IIA violations in practice. Particularly well specified models that allow for heterogeneity may be needed if a MNL model is to be used. Alternatively, it may be possible to use a more complex model to deal with IIA violations. However, as was found in Appendix 5, these other models can be problematic.

Finally, two hypotheses were tested in this chapter. The first was that CM can be used to provide valid estimates of non-use values, and the second
was that non-use values exist for non-environmental attributes. Relatively strong support was found for the first of these hypotheses, although some problems with content validity were noted. Hence it appears that there is justification for the wider use of CM in environmental decision making. Support for the second of these hypotheses was, however, mixed. It was concluded that non-use values can exist for non-environmental attributes, but that they need not exist in all cases. Therefore, it would appear necessary to include non-environmental attributes in primary valuation studies as they may be relevant. However, considerable caution should be exercised in transferring benefit estimates for non-environmental attributes because they may not exist in all cases.
8.1 Introduction

Most choice models are individual level models. Individual level models were estimated in virtually all of the conjoint studies focusing on non-market values that were reported in Chapter 3 (eg Adamowicz et al 1995, Roe et al 1996). Moreover, individual level models have been derived in many marketing studies (eg Blamey, Bennett, Morrison and Louviere 1998). However, the purpose of most decision makers when funding non-market valuation studies is not to derive individual level estimates, such as the sample mean. Rather, they usually require the extrapolation of sample estimates to aggregate benefit estimates so that they can use the results in cost-benefit analyses.

Probably the most basic way to derive an aggregate estimate is to substitute the mean socioeconomic values from the sample into the appropriate variables in a regression equation that shows, for example, how willingness to pay is a function of respondents' socioeconomic characteristics. Aggregate benefits can then be calculated by multiplying the mean sample estimate by the relevant population. Alternatively, compensating surplus can be calculated separately for each individual in the sample, averaged for the sample and expanded for the population (eg Peters, Adamowicz and Boxall 1995).

However, the expansion of sample estimates is not as straightforward as it may appear. There are three reasons for this. First, as discussed in Chapter 2, there are theoretical limitations regarding the assumptions that
must be made about the nature of individual level preferences if they are to be aggregated consistently. As demonstrated by Johansson (1987), there are only two possible ways to aggregate individual level estimates such that the sum of individual preferences will be proportional to social welfare. The first is to assume that the product of each individual's marginal utility of income and welfare weight is constant. However, as described in Chapter 2, this has the unreasonable implication that the welfare weight of people with higher income exceeds that of people with lower income. The only other way to derive a consistent aggregation is to assume that the marginal utility of income and welfare weight of each individual is constant across the population. While this assumption is also unrealistic, it is less unreasonable than the alternative, and is commonly used in practice. The implication of this assumption for empirical applications is that individual level model must be specified so that the marginal utility of income is constant if preferences are to be aggregated.

The second reason why extrapolation is not straightforward is that it is difficult to know the relevant population. Is it solely the area that was sampled, or can the mean sample estimate be extrapolated across a wider population? This problem has been described as that of knowing 'the geographical extent of the market' of the good in question (Smith 1993, p.687). Two broad approaches have been utilised to deal with this problem. The first involves the estimation of distance decay functions that show how distance from the non-market good of interest affects willingness to pay. The idea behind this approach is that, by estimating this function, the exact limits of the market can be defined. For example, Sutherland and Walsh (1985) estimated functions showing how willingness to pay for improved water quality was a function of distance. Similarly, Pate and Loomis (1997) examined how willingness to pay for (1) a wetland improvement program and (2) a river and salmon improvement program was affected by distance. A limitation of this

---

73 Individual level models result when the data are analysed using the observations from each respondent, rather than frequency or proportions data.
approach is that significant relationships between willingness to pay and
distance need not exist in all cases. Pate and Loomis (1997) only found a
significant relationship for the wetland improvement program, but not for
river and salmon improvement. They conjecture that this may have
resulted because the salmon improvement program primarily involved use
values. However, examples can also be cited of where distance decay
functions have not existed for non-use values. For example, Imber et al
(1991) found that people from throughout Australia were willing to pay
more to preserve the Kakadu Conservation Zone from mining than people
in the Northern Territory who were closer to the site. It is possible that
many factors apart from distance—such as effects on employment
opportunities and environmental preferences in general—could affect
peoples' willingness to pay. Hence the estimation of distance decay
functions is likely to be useful only in certain cases for defining the
geographical extent of the market. The second approach is to attempt to
transfer value estimates from the population where the survey was
conducted to another population, given assumptions about the equality of
the two populations. For example, the results of a survey conducted in
Sydney might be extrapolated across the entire population of New South
Wales. This procedure is known as benefit transfer and is considered in
detail in Part 3 of this thesis.

The third reason for complexity is that there may be divergences between
the characteristics of the sample and the actual population. If the sample is
unrepresentative of the population, extrapolating using the sample mean
may lead to biased aggregate estimates. Some method is, therefore, needed
to correct for any divergences. Broadly, two types of approaches have been
used in the literature. In the first type, sample means are adjusted to allow
for differences between sample and population characteristics. Examples
include using independent variables in regression analysis, weighted
regression analysis and weighted averaging. In the second type,
assumptions are made about the preferences of non-respondents\textsuperscript{74}, the
existence of whom is presumably the main cause of divergences, assuming
random sampling techniques were used. In Section 8.2, existing
approaches for aggregation given divergences between sample and
populations characteristics are considered, and a new approach is
suggested. The results from using these approaches are presented and
compared in Section 8.3. Conclusions are offered in Section 8.4.

8.2 Dealing with Divergences
between Sample and Population
Characteristics

Despite the use of random sampling procedures, sample characteristics
may diverge from population characteristics. This could be due to non-
random non-responses, other limitations associated with the sampling
method (eg out of date mailing list, survey administrators deliberately
choosing houses that are likely to agree to do the survey), or simply
chance. Several researchers have demonstrated that ignoring these
divergences when aggregating benefit estimates, especially when caused
by the existence of non-respondents, can cause bias (Edwards and
Anderson 1987; Whitehead, Groothuis and Blomquist 1993). Researchers
have several options available to them for correcting sampling differences.
As suggested in Section 8.1, they can either adjust mean values, or make
assumptions about the preferences of non-respondents\textsuperscript{75}.

\textsuperscript{74} Non-respondents are of two types. There are item non-responses where the
questionnaire is returned but answers are not reported for all questions, such as income.
Unit non-response occurs when the questionnaire is not returned or someone declines to
do an interview (Mitchell and Carson 1989). The latter type is most critical during
aggregation, but the former can also cause problems if it leads to missing observations in
an estimated model. For parsimony, only the effect of the latter has been considered in
this chapter.

\textsuperscript{75} It should be noted that these two approaches are mutually exclusive. This is because
one of the most likely sources of divergences is the existence of non-random non-
respondents. If assumptions are made about the preferences of non-respondents,
corrections should not also be made to the mean value estimate of survey respondents.
For example, the sample mean should not also be adjusted to allow for differences
between sample and population characteristics if non-respondents are assumed to have
One of the most common methods of dealing with divergences between sample and population characteristics is to adjust the sample mean. This has been done in several different ways (Loomis 1987b). In some studies, differences in socioeconomic characteristics such as income, age, sex or education have been corrected by substituting population averages into the independent variables of a regression equation (e.g., Schulze, Brookshire, Walther, MacFarland, Thayer, Whitworth, Ben-David, Malm, and Molenar 1983; Bennett et al. 1997). For example, if the mean income for the sample is higher than that of the population, the population mean could be used when calculating willingness to pay. An advantage of this approach is the capacity to allow for differences in multiple socioeconomic characteristics. However, its ability to deal with unrepresentative samples is partly limited because preferences are usually imperfectly related to socioeconomic characteristics. Also, there are concerns that if sample proportions do not match population proportions for relevant socioeconomic characteristics, ordinary least squares regression estimates may be inconsistent (Du Mouchel and Duncan 1983).

Another option is to use weighted regression analysis. Unlike ordinary least squares, this technique has the advantage of producing consistent results. Weighted regression analysis, which is possible with both discrete choice and ordinary regression analysis, involves the use of weights to correct for differences in the proportion of people in, say, each income class. Each observation in a stratum of the sample is weighted according to the proportion of people in the sample stratum and population stratum:

\[ \text{zero willingness to pay. This is because the sample no longer represents the population; it is only representative of the part of the population who are assumed to be respondents.} \]

\[ \text{It is possible that this may also be a problem with discrete choice regression analysis, although no information on this issue was identified.} \]
\[ W_i = \frac{N_i}{S_i} \]

where \( W_i \) = weight for observations occupying the ith stratum
\( N_i \) = population proportion of the ith stratum
\( S_i \) = sample proportion of the ith stratum

A weakness of this approach is that it is only possible to allow for divergences in one variable unless detailed information is available about the combinations of socioeconomic characteristics held by people in the population. For example, it is not possible to allow for differences in both income and education unless there is information available about the income distribution of people in each education category. It is not possible to create weights across more than one variable if this information is unavailable. A second weakness, as demonstrated in Section 8.3, is that the standard errors of the model coefficients increase substantially when weighted discrete choice models are used. A third weakness is that where there are missing data for the variable used for the weighting, observations have to be deleted from the sample. It is not possible to include these observations by using dummy variables, as was done in Chapter 7. This is because their weighting would be zero. Weighted regression analysis requires that the weight of each observation is greater than zero. Hence, the use of weighted regression analysis implies a smaller sample size.

A third option is the weighted average approach. In this approach, a weighted average is computed using population proportions for a given socioeconomic characteristic (Mitchell and Carson 1989). Mean willingness to pay is first calculated for different segments of the sample (eg different income groups). The population weighted average is then calculated by multiplying mean willingness to pay for each segment by the proportion of people in the population in each respective segment. More specifically, the weighted average is as follows:
\[ \sum_{i=1}^{n} (N_i \times WTP_i) \]

where \( N_i \) is the population proportion for the \( i \)th stratum \( WTP_i \) is the average willingness to pay of respondents occupying the \( i \)th stratum.

This approach is similar to the previous one except that divergences are corrected after the regression analysis. Inconsistent results may still be produced, but it does not have the associated problems with weighted regression analysis. The advantage of this approach is that the adjusted mean estimate more accurately reflects the distribution of a given critical variable, such as income, within society.

### 8.2.2 Dealing with non-responses

Instead of adjusting mean value to allow for divergences between sample and population characteristics, an alternative approach is to make assumptions about the preferences of non-respondents. This approach may be preferred because people’s preferences are generally imperfectly related to their socioeconomic characteristics. In other words, even though peoples’ preferences may be a function of their socioeconomic characteristics, they may also depend on unobserved forms of heterogeneity. Ignoring these unobserved forms of heterogeneity may lead to sample selection bias (Whitehead et al 1993).

One approach to dealing with non-random non-respondents is to assume that their willingness to pay is equal to zero. Some researchers (eg Boyle

---

77 Non-responses may not always cause sample characteristics to diverge from the population. Specifically, if non-responses are random they may not be problematic. Loomis (1987b, p.396), for example, contends that this may occur where researchers are focusing on a ‘highly select population’, such as hunters or fisherman, whose members are all fairly homogenous.
and Bishop 1985, Bennett et al 1997) assert that, in the context of non-use values, it is not appropriate to assume that non-respondents have the same preferences as respondents. Hence they have, conservatively, assigned non-respondents zero compensating surplus. This approach is based on the assumption that people are implying something about their preferences by not responding. That is, the benefits of completing the survey are less than the costs (Dillman 1978)\(^79\). It is likely that, for most non-respondents, the benefits of completing the survey depend to some extent on their perceived value for the good involved. Hence not completing the questionnaire implies, all else being equal\(^80\), a low valuation for the non-market good of interest. Moreover, substantial research in social psychology and marketing has demonstrated that non-respondents often differ from respondents in terms of interest in the subject of the survey, as well as in socioeconomic characteristics (Edwards and Anderson 1987). These features of non-respondents suggest that willingness to pay may be lower than for respondents. Yet while it is possible that the average compensating surplus of non-respondents may be lower (eg because of lower income or just disinterest), it does not follow that it will be equal to zero. Moreover, if a large proportion of non-respondents are those who are protesting against some aspect of the questionnaire, and have pro-environmental preferences, it is conceivable that the willingness to pay of non-respondents could be higher than that of the sample.

Another approach to deriving estimates of the preferences of non-respondents used with mail surveys involves determining the relationship between willingness to pay and the time responses are received (Hockstimm 1967, Filion 1976, Blamey 1995). Under this approach it is hypothesised

---

\(^78\) According to Mitchell and Carson (1989, p.263), sample selection bias occurs 'where the probability of obtaining a valid WTP response among sample elements having a particular set of observed characteristics is related to their value for the good.'

\(^79\) Even with personal interviews or drop-off and pick-up surveys, the cover and title of a questionnaire give non-respondents an idea of the content and likely benefits from completion. Non-respondents are able to do a quick personal cost-benefit analysis of their perceived benefits and costs of completing the survey.

\(^80\) Non-response may also reflect respondents' beliefs about their capacity to influence government policy, or that the costs of completing the survey are very high.
that willingness to pay is an inverse function of the time that a questionnaire was received. In other words, willingness to pay is lower with later responses. The willingness to pay of non-respondents is then found by extrapolating up to 100% return rate using this estimated relationship, and assuming that the entire sample would have responded given enough time and enough reminders. One weakness of this approach is that it involves extrapolating out of the range of the existing data. This is of particular concern as the relationship between willingness to pay and the time responses were received may be non-linear or non-decreasing. Secondly, this approach can only be used for mail surveys.

A third approach for allowing for the existence of non-respondents was developed by Cameron, Shaw and Ragland (forthcoming) for use with mail surveys. They use a two stage procedure. The propensity of people to return a completed questionnaire is modelled in the first stage. Data on non-respondents in any particular area is included in the model by using census zip (post) code information. In the second stage, the propensity of people to take a recreational trip, or agree to pay for some other non-market good, is modelled. This two stage process is modelled using a Tobit model with sample selection (Greene 1995). The capacity of this two stage process to identify and correct non-response bias results from allowing correlations between the error terms of the two parts of the model. The errors should be uncorrelated if there is no non-response bias, implying that unobserved factors (ie preferences) causing people to respond are not also making them more likely to agree to payment. By not allowing for these correlations, over or under predictions of the populations' propensity to visit a site or agree to pay for an environmental

---

81 In principle it could also be used with other types of surveys.
improvement may result. This method offers a novel way of dealing with non-response bias. However, it requires detailed information about the socioeconomic characteristics of non-respondents which may not be available at the zip code level in countries other than the USA.

A fourth approach involves attempting to identify whether non-respondents are likely to have similar preferences to respondents in the sample. As described above, for most non-respondents it can be assumed that the marginal benefits of completing a questionnaire are less than the marginal costs. This is likely to reflect that their value for the good of interest is low, and hence it is reasonable to assume that their willingness to pay is negligible. However, it is possible that the willingness to pay of a proportion of non-respondents is equal to that of the sample. For these respondents the survey may simply have come at a bad time, when they were too busy to complete it or faced with some other difficult situation. Given that respondents to personal interviews or drop-off and pick-up surveys are often required to complete the questionnaire immediately or within a fairly short amount of time, some people may be unable to participate even though they want to. One way to identify these respondents is to ask questions to find out why they are not able to respond. If they answer that they are too busy but nonetheless interested in the survey, then it may be reasonable to assume that their mean willingness to pay is the same as that of the sample.

This approach offers potential to be a relatively simple means of identifying the proportion of non-respondents whose preferences are similar to those of respondents. However, it does have limitations. What

---

82 Cameron et al (forthcoming, p.11), in the context of water based recreation, found that survey saliency (ie peoples preferences for a good) was outweighed by survey complexity. They write, 'our models suggest that unobserved factors which make targeted households less likely to respond to our questionnaire than our response model predicts also make them more likely to take water-based recreational trips than our trips models would predict. In this case, the endogenous survey complexity effect appears to dominate: infrequent participants and non-users of these waters may have found it far easier to fill in our questionnaire...'. The implication is that non-response bias need not always lead to overestimation of aggregate benefits.
some respondents say may not be equivalent to what they mean. For example, when a respondent says they’re not interested they may mean that they’re too busy, or vice versa. This, however, could be remedied by using closed ended rather than open ended questions. For example, people could be asked which of the following reasons was why they chose not to participate:

- not interested in the survey;
- not interested in the survey and too busy;
- interested in the survey, but too busy;
- too sick;
- too old;
- can’t speak English; or
- can’t read.

Another limitation is that people might not be interested in doing the survey, but might be interested in the issue. To determine whether this is true of a non-respondent, an additional statement could be included which asks whether they are interested in the issue. Provided non-respondents indicate that they are interested in the survey, even if they fit into more than one category, then it reasonable to assume that they have similar preferences to respondents.

Despite the limitations, this approach provides some information about the preferences of non-respondents. This technique is compared with existing approaches in the next section.

8.3 A Comparison of Approaches for Aggregating Sample Estimates

The models presented in Chapter 7 that included choice set attributes and socioeconomic characteristics are used to calculate aggregate willingness
to pay in this section. For each data set it is assumed that the frequency of breeding has increased by two years, that endangered and protected species present has increased by eight species, that wetland area has increased by 400 km$^2$ and that there has been no effect on irrigation related employment.

All of the approaches described in the previous section are applied except those approaches involving mail surveys. It was not possible to use these approaches as a drop-off and pick-up survey was used for the collection of data.

### 8.3.1 Use of independent variables

The first approach trialed involves substituting the Sydney population characteristics for income, children and house ownership in the appropriate independent variables in the MNL equations for the Gwydir Sydney and Macquarie Marshes surveys$^83$. For the remaining independent variables, where no information is available about population characteristics (eg environmental attitudes, future visitation), sample means have been used.

The adjusted sample mean estimates are shown in Table 8.3 (at the end of this section) and can be compared to the sample mean where no adjustment has been made. Mean willingness to pay is different in the Macquarie Marshes survey, having fallen from about $86 to $79. However, it is almost the same for the Gwydir Wetlands survey, having fallen from about $89 to $88. Particularly in the Gwydir survey, the relatively small magnitude of the changes in willingness to pay reflects the offsetting effects of changes in the variables between the Sydney population and the survey sample. Decreasing income reduced willingness to pay, while decreasing children increased willingness to pay. Estimates
of aggregate values for the Sydney and Moree populations derived using this approach are shown in Table 8.4.

### 8.3.2 Weighted regression analysis

The next approach used was weighted regression analysis. Income was used as the weighting variable in this regression because in Chapter 7 this was shown to be one the main characteristics where the Gwydir Sydney and Macquarie Marshes samples diverged from the Sydney population. As mentioned earlier, the standard errors of the coefficients increased substantially when this approach was used (see Table 8.1). Note also the relatively smaller sample size compared to the models presented in Chapter 7\(^4\).

The results from the use of weighted regression analysis are also reported in Tables 8.3 and 8.4. Mean willingness to pay for the Macquarie Marshes sample is about 15% larger using this approach, having increased from about $86 to $97. However, mean willingness to pay is again almost identical to the sample mean for the Gwydir Wetlands sample.

\(^3\)The estimates reported below were calculated using equation 3.10 reported in Section 3.3.2.

\(^4\)In Chapter 7 the equivalent sample size for the Macquarie Marshes model was 1397 observations (178 skipped) and for the Gwydir Sydney model was 1269 observations (160 skipped).
Table 8.1: Multinomial logit models estimated using weights

<table>
<thead>
<tr>
<th>Variables</th>
<th>Macquarie Marshes</th>
<th>Gwydir Wetlands - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>-1.884** (0.819)</td>
<td>-0.565 (0.766)</td>
</tr>
<tr>
<td>C3</td>
<td>-1.586** (0.819)</td>
<td>-0.286 (0.769)</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.575 (0.464)</td>
<td>-0.892 (0.591)</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.289 (0.500)</td>
<td>0.702 (0.543)</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.124E-4 (0.795E-5)</td>
<td>0.640E-5 (0.708E-5)</td>
</tr>
<tr>
<td>VISIT</td>
<td>1.131*** (0.420)</td>
<td>0.661* (0.407)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>1.419*** (0.508)</td>
<td>0.855** (0.438)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-0.919 (0.585)</td>
<td>-1.030* (0.615)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.139E-1*** (0.268E-2)</td>
<td>-0.119E-1*** (0.300E-2)</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.255E-2 (0.213E-2)</td>
<td>0.157E-2 (0.355E-2)</td>
</tr>
<tr>
<td>AREA</td>
<td>0.570E-3 (0.400E-3)</td>
<td>0.441E-3 (0.865E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.297* (0.157)</td>
<td>-0.125 (0.150)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>-0.548E-1** (0.306E-1)</td>
<td>0.450E-1 (0.230E-1)</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Macquarie Marshes</th>
<th>Gwydir Wetlands - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-144.484</td>
<td>-144.618</td>
</tr>
<tr>
<td>( \chi^2 ) (constants only)</td>
<td>71.843</td>
<td>39.566</td>
</tr>
<tr>
<td>( \rho^2 ) adjusted</td>
<td>0.201</td>
<td>0.116</td>
</tr>
<tr>
<td>Iterations completed</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Observations</td>
<td>1272 (109 skipped)</td>
<td>1157 (90 skipped)</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level. Standard errors are in brackets.

8.3.3 Weighted averaging

The third approach used is that of weighted averaging. Similar to the use of weighted regression analysis, income was used as the weighting variable. The MNL equations reported in Chapter 7 were used to calculate the average willingness to pay of people in each income stratum. These
values were then multiplied by the proportion of the population in each income stratum and summed to calculate the population average, which is reported in Table 8.3. The weighted averaging approach produced mean estimates about 5% lower than the sample mean for the Macquarie Marshes survey, decreasing from about $86 to $81. The mean estimate for the Gwydir Sydney survey decreased from about $89 to $87, a minor adjustment.

### 8.3.4 Assuming non-respondents have zero willingness to pay

The results from assuming non-respondents have zero willingness to pay are also reported in Table 8.3. The household mean is calculated by multiplying the sample mean by the response rate. This approach produces the most conservative estimate of household willingness to pay in each of the surveys. The mean estimate fell from about $89 to $43 for the Gwydir Sydney survey; for the Gwydir Moree survey it fell from about $42 to $21; and for the Macquarie Marshes from about $86 to $38. These are decreases in mean willingness to pay of 50% or more.

### 8.3.5 Classifying non-respondents

As described in Chapter 7, the three surveys conducted in this thesis were administered using a drop-off and pick-up format. An advantage of this format is that contact is made with both respondents and non-respondents. While the contact with non-respondents is usually brief, it is possible to ask people why they were unwilling to do the survey. The results of this questioning are shown in Table 8.2. The most common reason given for not participating was lack of interest (about 50-60%), followed by being too busy (about 25-30%).
Table 8.2: Reasons for non-response

<table>
<thead>
<tr>
<th>Reason</th>
<th>Gwydir - Moree</th>
<th>Gwydir - Sydney</th>
<th>Macquarie - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too busy</td>
<td>44 (27%)</td>
<td>63 (24%)</td>
<td>74 (32%)</td>
</tr>
<tr>
<td>Not interested</td>
<td>79 (49%)</td>
<td>143 (55%)</td>
<td>137 (59%)</td>
</tr>
<tr>
<td>Too old/too sick</td>
<td>15 (9%)</td>
<td>23 (9%)</td>
<td>11 (5%)</td>
</tr>
<tr>
<td>Illiterate/can’t speak English</td>
<td>10 (6%)</td>
<td>24 (9%)</td>
<td>9 (4%)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (5%)</td>
<td>29 (11%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Total</td>
<td>162</td>
<td>259</td>
<td>233</td>
</tr>
</tbody>
</table>

As discussed in the previous section, it is assumed that the respondents who state they are too busy to do the survey have willingness to pay that is equal to the sample mean. Therefore it is assumed that at least 25-30% of non-respondents are likely to have similar preferences to the survey sample. Willingness to pay is assumed to be equal to zero for the remaining respondents. Admittedly this procedure requires further refinement, such as through the use of closed ended rather than open ended questions. Moreover, the credibility of this procedure could be enhanced if surveys of non-respondents who state that they are ‘too busy but interested’ demonstrate that their willingness to pay is similar to that of the sample and different from that of other non-respondents. Hence these results are presented here primarily to illustrate this approach rather than provide definitive estimates of aggregate benefits.

The results of using this assumption when calculating mean household willingness to pay are shown in Table 8.3. Willingness to pay is greater than when assuming that all non-respondents have zero willingness to pay, but less than with the other approaches.

Table 8.3: Mean household willingness to pay

<table>
<thead>
<tr>
<th>Method</th>
<th>Gwydir-Sydney</th>
<th>Gwydir-Moree</th>
<th>Macquarie - Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean</td>
<td>$89.23</td>
<td>$41.90</td>
<td>$85.77</td>
</tr>
<tr>
<td>Use of independent variables</td>
<td>$87.84</td>
<td>-</td>
<td>$78.81</td>
</tr>
<tr>
<td>Weighted regression analysis</td>
<td>$89.08</td>
<td>-</td>
<td>$97.45</td>
</tr>
<tr>
<td>Weighted averaging</td>
<td>$87.04</td>
<td>-</td>
<td>$81.10</td>
</tr>
<tr>
<td>Non-responses recoded to zero</td>
<td>$43.19</td>
<td>$21.08</td>
<td>$37.91</td>
</tr>
<tr>
<td>Classifying non-responses</td>
<td>$52.29</td>
<td>$25.06</td>
<td>$51.98</td>
</tr>
</tbody>
</table>

# Not all techniques have been applied to the Gwydir-Moree survey because the population data were not available.
Aggregate estimates of willingness to pay are presented in Table 8.4. These figures have been calculated using a population in Sydney of 3,741,290 and in Moree of 15,517, and given 2.6 people per household (Australian Bureau of Statistics 1996 Census data). Overall, it appears that the use of the first three approaches has only marginally altered estimates of aggregate willingness to pay for these case studies. However, making assumptions about the nature of non-respondents' preferences has had substantial effects. Recoding non-respondents as having zero willingness to pay has more than halved aggregate benefits, while classifying non-respondents resulted in estimates that are about 40% lower.

Table 8.4: Aggregate willingness to pay

<table>
<thead>
<tr>
<th></th>
<th>Gwydir-Sydney</th>
<th>Gwydir-Moree</th>
<th>Macquarie – Sydney</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean</td>
<td>$128,398,222</td>
<td>$250,059</td>
<td>$123,419,428</td>
</tr>
<tr>
<td>Use of independent variables</td>
<td>$126,398,071</td>
<td></td>
<td>$113,404,280</td>
</tr>
<tr>
<td>Weighted regression analysis</td>
<td>$128,182,379</td>
<td></td>
<td>$140,226,457</td>
</tr>
<tr>
<td>Weighted averaging</td>
<td>$125,246,904</td>
<td></td>
<td>$116,699,494</td>
</tr>
<tr>
<td>Non-responses recoded to zero</td>
<td>$62,148,596</td>
<td>$125,805</td>
<td>$54,550,898</td>
</tr>
<tr>
<td>Classifying non-responses</td>
<td>$75,243,114</td>
<td>$149,558</td>
<td>$74,797,037</td>
</tr>
</tbody>
</table>

8.4 Conclusion

Issues associated with the aggregation of individual estimates have been explored in this chapter. In particular, attention has been given to methods of allowing for divergences between sample and population characteristics. The first three approaches considered are characterised by attempts to adjust mean sample estimates. They include the use of independent variables, and two different weighting approaches. The use of independent variables was argued to have the advantage of enabling correction of divergences in multiple variables, however there was concern that it might produce inconsistent results. The use of weighted regression analysis has the potential to correct this problem, but it also has several limitations resulting from the use of weights. The third approach considered was the weighted averaging approach. While still based on normal regression
analysis, the advantage of this approach is that the adjusted mean estimate more accurately reflects the distribution of a given critical variable, such as income, within society. A limitation of both weighting approaches is that it is generally only possible to use weights for a single variable, even though their might be differences in multiple characteristics. In sum, none of these approaches appear to be ideal.

Several other approaches that involve making assumptions about the preferences of non-respondents were considered. The first approach is to assume that the willingness to pay of non-respondents is equal to zero. However, this is likely to be an overly conservative assumption. The second approach involves extrapolating willingness to pay for non-respondents using time of receipt of the questionnaire. Apart from being limited to mail surveys, this requires extrapolation out of the range of the data. The third approach involves modelling, in a two stage process, whether people will respond to a survey and their willingness to pay. While this is a sophisticated method of dealing with non-respondents, it requires socioeconomic data about all non-respondents, which may not always be available in a sufficiently decentralised form. The final approach described involves reclassifying non-respondents into those who are likely to have the same willingness to pay as respondents, and those who are likely to have zero willingness to pay. This approach has the benefit of providing some information about the preferences of non-respondents; however further research is necessary before it could be used in practice.

Five of these approaches have been trialed, and compared to the alternative of aggregating using the sample mean. Generally the three approaches that attempted to adjust the sample mean all produced results that were close to the sample mean. Willingness to pay using these three approaches changed by no more than 15%. However, making assumptions about the preferences of non-respondents was found to have substantial effects. Assuming that all non-respondents had zero willingness to pay reduced
aggregate benefits by about half, thus demonstrating the conservative nature of this assumption. Classifying non-respondents also led to much lower aggregate estimates, but not as low as assuming that all non-respondents had zero willingness to pay.
Part 3  Benefit Transfer

The focus of this section of the thesis is the issue of benefit transfer.

There are two chapters in this part of the thesis. In Chapter 9 the literature pertaining to benefit transfer is reviewed to provide the context for the benefit transfer tests that are conducted in the next chapter. The use and development of benefit transfer are detailed in this chapter, and various benefit transfer tests are reviewed. In Chapter 10 the results from two benefit transfer tests are reported. The first test involves a test of the transferability of value estimates across sites, and the second involves a test of the transferability of value estimates across populations.
Chapter 9  Benefit Transfer

9.1  Introduction

In many situations, because of time and monetary constraints, those tasked with making decisions regarding the allocation of natural resources are required to extrapolate from existing data that were collected for a different purpose. The use of existing studies in project evaluations and policy analyses is known in the resource economics literature as ‘benefit transfer’. As stated by Boyle and Bergstrom (1992, p.657):

Benefit transfer is defined as the transfer of existing estimates of nonmarket values to a new study which is different from the study for which the values were originally estimated. In essence, this is simply the application of secondary data to a new policy issue.

The use of existing data is not something new to economics, or indeed many other disciplines. The novelty of ‘benefit transfer’ is that data that are believed to be sensitive to changes in the context in which they were collected, and subject to various uncertainties, are being reused.

Resource economists have been somewhat divided in their view about the validity of using benefit transfer. Boyle and Bergstrom (1992) suggested that resource economists tend to have one of three different philosophical orientations about benefit transfer. The pragmatists believe that benefit transfer is valid and should be expanded; proponents of the impossibility myth believe that differences between sites makes the transfer of value estimates impossible; and idealists believe that benefit transfer is possible, but that there must be strict standards. The idealist also believes that further research into the validity of benefit transfer is required if the transfers are to be undertaken with any certainty.
The first objective pursued in this chapter is to review the benefit transfer literature to determine what is known about the validity and reliability of benefit transfer. The second objective is to identify areas in need of further research. In order to achieve these objectives, the next section provides a brief review of the use and development of benefit transfer. In Section 9.3 information is provided about the determinants of value estimates and hence some of the factors that could be expected to affect the validity of benefit transfer. Studies that have provided tests of benefit transfer are reviewed in Section 9.4, and conclusions are offered in Section 9.5.

9.2 The Use and Development of Benefit Transfer

Benefit transfer has a relatively long history. It has been used in a range of applications, including for recreation values, the value of improved air and water quality, the value of noise reductions, and even non-use values. The earliest benefit transfers were of valuations made using expert opinions. In the USA expert opinions were used in 1962 to derive a table of administratively approved 'unit day values' of recreation for use in evaluations of water resource developments. According to Loomis (1992), the main reasons that expert opinions were used was that there was a dearth of primary studies. The use of expert opinions was, however, seen to be more prone to difficulties than the extrapolation of primary data (Loomis 1992). Consequently, later benefit transfers generally involved the extrapolation of primary estimates, particularly those made using the travel cost method. For example, Cichetti, Fisher and Smith (1976) derived a system of demand equations for six different sites using the travel cost method. Using this system of demand equations they were able to infer the expected benefits of a proposed ski resort in California. In a somewhat different study, Smith, Desvousges and McGivney (1983)
developed a travel cost model that could be used to show the effect of a site's characteristics on recreation demand based on data from 22 different recreation sites. The results from this study were then used to infer the value of water quality improvements in the Monongahela River in Pennsylvania. Estimates made using the hedonic price model have also been used for benefit transfer. For example, the Roads and Traffic Authority (1995) used summary estimates of the effect of highway noise on property values of noise reported in Nelson (1980) to evaluate proposed road construction projects.

There have been attempts in more recent years to use SP studies for benefit transfer. For example, Desvousges, Naughton and Parsons (1992) and Luken, Johnson and Kibler (1992) used several existing CVM studies to infer the benefits attached to water quality regulations for the pulp and paper industry in the USA. In Australia Dumsday, Jakobsson and Ransome (1992) used the results from a number of SP studies to infer the value of protecting river segments in Victoria. The NSW EPA ENVALUE database contains a number of CVM studies that are reported to facilitate use in benefit transfer (Morrison, Groenhout and Moore 1995).

There have also been attempts in recent years to develop more rigorous approaches to benefit transfer. Various 'protocols' have been developed to assist in selecting appropriate studies for use in benefit transfer (Desvousges, Naughton and Parsons 1992; Boyle and Bergstrom 1992; Smith 1992; and Kask and Shogren 1994). Protocols have been developed to provide guidance for analysts about factors that can significantly affect value estimates. One limitation of these protocols is that often it is impossible to find studies that satisfy all of the suggested criteria for study selection. An analyst has three choices in such a circumstance: (1) to not use benefit transfer; (2) ask if the bias is small enough to be acceptable; and (3) determine if value estimates can be altered systematically to remove any bias (Boyle and Bergstrom 1992). Analysts need to know the factors that affect willingness to pay and the likely direction and
Boyle and Bergstrom (1992) suggested that two main lines of research are required to achieve this goal. The first involves various tests of convergent validity. The objective of these tests is to determine whether benefit transfer is statistically valid, what biases might be expected, their extent, and whether they can be corrected:

We could... compare the benefit transfer values for the policy site... with the value estimates for the policy site from primary data... If benefit transfer estimates are not statistically different from the primary data value estimates developed at the policy site, convergent validity is established. When benefit transfer estimates are biased, these concurrent evaluations can examine the size of the bias, direction of the bias and adjustments that might be made in study site estimates to mitigate the bias. Validity investigations ultimately will identify conditions where benefit transfer works and procedures necessary to make benefit transfer operational (p.661).

Boyle and Bergstrom (1992) also suggested that studies based on primary data be undertaken to determine which variables have a significant effect on value estimates. This could involve separate studies or ‘meta-analyses’ which attempt to model the results of multiple studies.

In the following two sections of this chapter the literature pertaining to both of these lines of investigation is reviewed.

9.3 The Determinants of Value Estimates

Broad indications of the types of factors that are likely to affect valuation estimates are provided in the various benefit transfer protocols referred to in the previous section. The three main factors appear to be:
(1) the characteristics of the nonmarket commodity being valued;
(2) the characteristics of the populations at the study and the policy sites; and
(3) the questionnaire design and delivery, and the method of estimation.\textsuperscript{78}

Information about the importance of these factors on value estimates, and the nature of any effects, has been found using 'meta analyses'. Meta analysis is an 'econometric review' of a number of different studies, with the goal of finding systematic factors that affect value estimates. An increasing number of meta-analyses have been published in recent years, including Smith and Kaoru (1990), Walsh, Johnson and McKean (1992), Smith and Huang (1993, 1995), Boyle, Poe and Bergstrom (1994) and Smith and Osborne (1996). Three of these meta-analyses include CVM studies in their analyses, and are therefore relevant to this thesis and so are reviewed here. These are the studies by Walsh et al (1992), Boyle et al (1994) and Smith and Osborne (1996).

Walsh et al (1992) assembled 287 different estimates of the value of a recreation day. One hundred and twenty nine of these were made using the CVM. They regressed a variety of different independent variables against the consumer surplus estimated for each study site. They did this separately for the estimates made using the CVM and travel cost methods, as well as in a model that used the entire data set. For the CVM model they found nine significant dummy variables. Six of these variables represented site characteristics. They were site quality (+), two regional variables, a fishing variable (+) and two hunting variables (+). Of these variables, the largest effects on consumer surplus were from site quality and whether salt water and anadromous fishing were available. Of the remaining significant variables, two were related to survey design. The first represented whether the survey was conducted on site (-), and the second was whether an open-ended question format was used instead of either dichotomous choice or iterative bidding (-). The actual location of

\textsuperscript{78} The effect of different methods of modelling data on value estimates generated using the travel cost model is considered by McConnell (1992).
where the survey was collected (on or off site) had a much larger influence of value estimates than the choice of survey format. Insignificant variables included the use of the dichotomous choice format, whether the survey was administered by the government, and whether out of state residents were included in the sample.

The focus of Boyle et al.'s (1994) meta analyses was on the estimation of groundwater values made using the CVM. They assembled 52 different estimates made in eight different studies. The dependent variable in this study was the log of mean willingness to pay. The majority of the independent variables in their regression related to site and socioeconomic characteristics (nine out of 12). These variables included whether nitrates were mentioned as the source of contamination (+), whether the cost of substitute water sources was mentioned (-), respondent income (+), whether policy aims to contain contamination (-), and whether a change in the supply of water was expected (+). Insignificant variables included whether substitute sources of water were mentioned, the proportion of respondents who derived water from a public source, and whether cancer was mentioned. It is interesting that some of the factors that one would expect to have an effect were not significant (e.g., substitute water sources and cancer). This could be because respondents anticipated these factors even though they were not explicitly mentioned. Three other variables were included to show the effect of questionnaire design and delivery. They were for the use of the dichotomous choice format (rather than the open ended or iterative bidding formats), estimation of use values only (rather than both use and non-use values), and mail response rate. The dummy variable for the dichotomous choice format was, contrary to the finding of Walsh et al (1992), positive and significant. This suggests that the use of the dichotomous choice format can have a positive effect on willingness to pay. As expected, the estimation solely of use values was found to have a significant and negative effect on willingness to pay. The variable for response rate was negative and significant. Boyle et al (1994)

The sign in brackets indicates the direction of the effect.
suggest that this variable may reflect the effect of the quality of the study, and hence reduced hypothetical bias.

The meta-analysis by Smith and Osborne (1996) aimed to determine the factors which explain variations in CVM estimates of improved visibility in National Parks in the USA. Of particular interest in this paper was the question of whether CVM estimates pass a 'scope test'; that is, whether the value estimates are sensitive to changes in the characteristics of the non-market commodity. Smith and Osborne's (1996) data set included 116 observations from five different studies. Smith and Osborne (1996) estimated a number of different equations, each of which had the log of mean willingness to pay as the dependent variable. Nine different independent variables were included in the regressions, including five that related to the characteristics of the non-market good and four that related to questionnaire design or delivery. Smith and Osborne (1996) included two different scope variables: a continuous variable that showed willingness to pay for improvements in visibility and a dummy variable to indicate studies that estimated willingness to pay for preventing declines in visibility. The willingness to pay was positive for both variables, however it was only consistently significant for the continuous variable. This raises the question of whether people are willing to pay different amounts for gains and losses. Independent variables that were relevant to questionnaire design were the use of the iterative bidding format (+), use of the open ended format (-), use of on-site interviews (-) and inclusion of residents of the state where the park was located (+). Only the variables for the open-ended format and on-site interviews were consistently significant, confirming the results found by Walsh et al (1992) and Boyle et al (1994). Also, similar to the finding of Walsh et al (1992), the variable representing whether the sample only included residents of the state where the park was located was insignificant.

Overall, these meta-analyses provide a number of indications about the determinants of value estimates, and hence the factors that need to be
considered when undertaking benefit transfer. The characteristics of non-market goods have significant effects on value estimates. This is consistent with Lancaster's (1966) theory of consumer demand, and with expectations about estimates being 'contingent' on the information provided. The factors that have been found to effect value estimates include recreation site quality, the type of recreation activity, the scope of an environmental improvement, and the cause of a decline in environmental quality and its consequences. The meta-analyses have produced less information about the effect of respondent characteristics of value estimates. The analyses have, however, shown that income is a significant variable, and that the location of the respondents is not necessarily significant. Finally, the analyses have provided information about the effect of questionnaire and survey design. Elicitation format has been shown to affect value estimates, with open-ended questions producing lower value estimates. The estimation of use values only, as would be expected, has been found to produce lower estimates of value compared to the dichotomous choice format. The location of where the survey was administered (ie on site) can effect willingness to pay, but the organisation who administered the survey (ie government or non-government) was not found to have an effect.

This type of information is likely to be useful to analysts when attempting benefit transfer. Knowing what are the key determinants of value estimates is useful for determining whether a study is suited to benefit transfer, and for adjusting for any bias. Tests designed to determine the effect of changes in the characteristics of non-market goods and populations on the equality of value estimates are reported in the following section.
9.4  Tests of Benefit Transfer

Three main types of tests have been conducted to determine the validity and reliability of benefit transfer. These are tests of the transferability of results (1) over time, (2) across different sites, and (3) across different populations. The first two aim to determine the effect of temporal and physical characteristics of the non-market good, while the third considers the effect of characteristics of the population on preferences. Studies that have incorporated these tests are reviewed in the following sub-sections.

9.4.1  Reliability over time

Tests of reliability over time aim to determine whether value estimates generated at one point of time would be equivalent to those used at a later date. A large number of these tests have been undertaken using the CVM. Most studies have found that value estimates are stable over at least a few years. Overall these results are encouraging for the use of benefit transfer over time.\(^{80}\)

Kealy, Montgomery and Dovidio (1990) examined stability of models analysing the choice of both private and public goods using panel and student data. The private good was a Cadbury chocolate bar and the public good was the prevention of damage to Adirondack lakes due to acid rain. They found no statistical difference between models using a likelihood ratio test. However, their study was conducted over only a two week period.

Loomis (1990) used a much longer time period of nine months and found significant differences between models. Loomis (1990) examined
willingness to pay for two different improvements in environmental quality at Mono Lake in California using both open-ended and dichotomous choice elicitation formats. While test-retest correlations were significantly different from zero, indicating a degree of reliability, significant differences between models were found using likelihood ratio tests for one of the improvements under each elicitation format.

Reiling, Boyle, Phillips and Anderson (1990) compared the willingness to pay of one set of respondents to control black flies (a type of mosquito) at the peak of the black fly season with that of a second set of respondents at the end of the season (3 months later). Using a paired t-test they concluded that the observed mean values were equivalent.

Stevens, Moore and Glass (1994) compared respondents' willingness to pay to preserve the bald eagle in 1989 and 1992 using both panel data and independent samples from each year. Using various statistical tests including a paired t-test and a chow test, they concluded that the observed mean values in the two time periods were equivalent.

Lastly, Teisl, Boyle, McCollum and Reiling (1995) compared respondents' willingness to pay for moose hunting permits across a five month period in 1989 and 1990. They conducted a pretest for one sub-sample, a test-retest for another sub-sample, and a post-test for a third sub-sample in 1990 which had not been surveyed in 1989. Likelihood ratio tests indicated that the regression equations for each of the groups were equivalent.

Hence, estimated models tend to be relatively stable over several years. However, the results from other types of benefit transfer tests have been less positive.

80 It should be noted that tests of reliability do not provide any indication of the validity of results. Indeed it could be argued that the lack of change in preferences over time is indicative of perfect embedding.
A common type of benefit transfer is the extrapolation of benefit estimates across sites. The inability to reject the null hypothesis of equivalence of results across sites implies that respondents' underlying preferences for the two goods are identical. It also implies that respondents' valuations are independent of the context of the valuation exercise. The latter may occur because respondents' preferences are similar; but also may imply that there may be problems associated with yea-saying or symbolic responses (Andreoni 1990; Blamey, Bennett and Morrison forthcoming). A few tests of this hypothesis have been conducted in both the transport (Watson and Westin 1975; Atherton and Ben-Akiva 1976) and non-market valuation literature (Loomis 1992; Bergland, Magnussen and Navrud 1995).

The first published study involving a test of transferability across sites appeared in the transportation literature. Watson and Westin (1975) gathered SP data showing the conditions under which respondents would choose to use the Edinburgh-Glasgow railway. They split their sample into six sub-samples based on the origin and destination of their trip. The authors then estimated binary logit models for each sub-sample. They conducted a number of different tests to determine whether the models were statistically equivalent. This included tests of differences in the overall model structures using likelihood ratio tests. Watson and Westin (1975) found that the models estimated for respondents who take trips at central locations were, for the most part, statistically equivalent. However, the models based on data from respondents who travel to or from non-central locations were frequently different from the other models. This indicates that travel choice models may in some circumstances be statistically equivalent, depending on the similarity of destinations and origins.

Less ambiguous findings for the transferability of travel choice models were reported by Atherton and Ben-Akiva (1976). They estimated three
separate MNL models for data from Washington DC collected in 1968, data from New Bedford collected in 1963, and data from Los Angeles collected in 1967. They report the results of a modified likelihood ratio test which showed that the Washington and New Bedford models were statistically equivalent. They also tested the equivalence of model coefficients using paired t-tests and found that all but two coefficients were the same. No other test statistics for the remaining sites were reported in the paper.

Another test of this transferability hypothesis was conducted by Loomis (1992) using recreation fishing data and travel cost models. Loomis (1992) compared four different data sets: ocean sport salmon fishing (Oregon and Washington), and freshwater steelhead fishing (Oregon and Idaho). He first compared the two ocean sport fishing data sets using a chow type test and found that they were statistically different. He then compared the two steelhead data sets and found that they were equivalent. Loomis (1992) suggested that the difference in the ocean sport fishing data sets in part resulted from the different time periods of data collection: the two data sets were collected six years apart, and over this time angler use levels in both states had decreased by 75%. The results from this study suggest that model transferability across time and sites are likely to be more inaccurate, especially if there have been temporal changes in natural resources.

Bergland, Magnussen and Navrud (1995) conducted a transferability test of non-use values across sites. They estimated the value of improved water quality at two different watercourses in Norway using the CVM. Each survey was conducted in the area around the respective watercourse, hence the test also involves different populations. Bergland et al (1995) tested if models estimated using the data from each survey were equivalent using a likelihood ratio test, and found that the models differed statistically. They also tested if estimates of willingness to pay were equivalent, but found that they differed. Two different willingness to pay
estimates were generated, which differed by 25% and 31% across the two sites, which indicates that the differences are not large, at least from a policy perspective. Bergland et al (1995) suggested two reasons why all hypotheses were rejected: (1) estimated models did not explicitly consider differences in environmental quality between sites; and (2) the influence of socioeconomic factors was not adequately considered. The authors suggest that 'a potential alternative is to explore benefit functions which take changing environmental quality and differing socio-economics explicitly into account' (p.21). This comment is supportive of the use of CM for benefit transfer because of its ability to allow for both differences in changes in environmental quality and in socioeconomic characteristics.

9.4.3 Transfers across populations

The next type of test involves determining whether different population groups have equivalent preferences for the same non-market good. For example, whether residents in a small country town near a National Park have the same willingness to pay to preserve the park as residents in a large urban city some distance way, after allowing for differences in socioeconomic characteristics such as income. Two tests of this type involving non-market goods have been reported in the literature, by Parsons and Kealy (1994) and Swallow, Weaver, Opaluch and Michelman (1994).

Parsons and Kealy (1994) estimated MNL models using travel cost data to value improvements in water quality at lakes in Wisconsin. They separated their sample into two different groups: respondents living in Milwaukee (an urban centre) and respondents living in rural areas or small towns. Parsons and Kealy (1994) estimated separate models for each data set and tested for overall differences between the models and individual differences in model coefficients. Using a likelihood ratio test, Parsons
and Kealy (1994) rejected the hypothesis that the two models were the same. The results from paired t-tests of the equality of individual coefficients were mixed, with several variables not being significantly different. Parsons and Kealy (1994) also tested for differences in implicit prices. They did this by constructing standard errors using the Krinsky and Robb (1986) procedure. They found only minor differences in implicit prices derived from the true model and a transfer model that allowed for differences in individual characteristics and recreational opportunities. However, the authors attribute this similarity more to offsetting variables rather than identical behavioural models.

Swallow, Weaver, Opaluch and Michelman (1994) conducted a second test of this type of benefit transfer. The context for their study was the avoidance of impacts on natural and social resources due to landfill siting. They estimated separate MNL models for urban and non-urban residents using SP data. While it was somewhat incidental to their study, they compared the equality of individual model coefficients using paired t-tests. Seven out of ten of the coefficients were statistically different. A likelihood ratio test was not conducted to determine whether the models were different.

9.5 Conclusion

The validity of benefit transfer continues to be questioned by some resource economists who concur with the 'impossibility myth' (eg Krupnick 1993). And while information has been collected over the last decade which may provide some guidance and comfort to the 'idealists', research on this topic still appears to be in an early stage. Precisely what causes value estimates to differ, and by how much, is in many cases still unknown.

81 See Section 10.3.
The meta-analyses reported in Section 9.2 provided some useful indications of the factors that affect value estimates. However, except for the study by Walsh et al (1992), these analyses have been based on relatively small data sets. Hence the conclusions from these studies are, to a certain extent, speculative. More research of this sort would appear to be required before firm conclusions can be drawn.

Further evidence about the factors that affect value estimates has been provided by studies that have attempted to conduct tests of transferability. These studies have demonstrated that value estimates are reliable over time periods of up to a few years. The reliability over longer time periods is, however, unclear. The existing evidence regarding the validity of transfers across either sites or populations is mixed. Generally it appears that combining transfers (eg across time and across sites, or across sites and populations) is likely to be more problematic than just one type of transfer. Transfers involving non-use values appear to be more difficult than those involving use values. However the evidence available is limited and it would appear that further research in this area is warranted. Two tests of the validity of benefit transfer are conducted in the next chapter. The first is a test of the transferability solely across sites, and the second is of transferability solely across populations. Consistent with the suggestion of Bergland et al (1995), a technique which allows for differences in environmental improvements—choice modelling—is used for this testing.
Chapter 10 Benefit Transfer Tests

10.1 Introduction

The final objective of this thesis is to test the transferability of non-use value estimates generated using CM. As discussed in the previous chapter, tests of this type have not previously been undertaken using CM. An advantage of using CM for these tests is that it is possible to allow for different changes in environmental quality as well as differences in socio-demographics when transferring value estimates.

The models used for the benefit transfer tests are initially presented in Section 10.2. These models are different from the models presented in Chapter 7. Several insignificant variables have been deleted from the models previously presented. This was done to replicate what analysts attempting benefit transfer are likely to do in practice. Also, to allow full comparability between models, data generated from the fifth block of dominated alternatives has been deleted from the Macquarie Marshes data set (see Appendix 5).

Two hypotheses are tested in the chapter. The first is that benefit transfer across sites is valid. This hypothesis is tested by comparing the results for the Gwydir Sydney and Macquarie Marshes surveys and is reported in Section 10.3. The second hypothesis is that transfers across populations are valid. This is tested using the Gwydir Sydney and Gwydir Moree models, and thus involves a comparison between the preferences of a

---

82 An earlier version of the results presented in this chapter was reported in Morrison, Bennett, Blamey and Louviere (1998b).
population from a rural town close to the natural resource, and of an urban population some distance away. Tests of this hypothesis are reported in Section 10.4. Conclusions are offered in Section 10.5.

10.2 MNL Models Used for Benefit Transfer Tests

The MNL models used for the benefit transfer tests are presented in Table 10.1. Similar to the results presented in Chapter 7, the coefficients for RATES, BREED and SPECIES are significant at the 5% level or higher and have \textit{a priori} expected signs in each model. The coefficient for AREA has the correct sign and is significant at the 10% level or higher in both Sydney models. The coefficient for JOBS is only significant in the Gwydir Sydney model. However, the coefficient is significant in the Macquarie Marshes model if a one-tailed t-test is used (given expectations about the sign of the coefficient for JOBS). Neither the coefficients for JOBS nor AREA are significant in the Gwydir Moree model.

The coefficients for VISIT, PROGRE, PRODEV, UNDER, WILLWORK and ONE-OFF are all significant at the 1% level and have \textit{a priori} expected signs. The sign of the coefficient for CHILD is positive in the Macquarie Marshes and Gwydir Moree models, but negative in the Gwydir Sydney model. The sign of INCOME also differs across the first three surveys. In the two Sydney surveys INCOME, in accordance with theory, has a positive sign. It had a negative sign in the Moree survey\textsuperscript{83}. The explanatory power of the models is satisfactory, with adjusted rho-squared between 0.13 and 0.17. The chi-squared statistics indicate that each model is significant overall.

\textsuperscript{83} To some extent the negative sign could reflect the fact that many people with higher incomes in Moree are involved in the cotton industry, and therefore have a vested interest in seeing less water going to the Gwydir Wetlands.
### Table 10.1: MNL models used for benefit transfer tests

<table>
<thead>
<tr>
<th>Variables</th>
<th>Gwydir</th>
<th>Moree</th>
<th>Gwydir Sydney</th>
<th>Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>0.878*</td>
<td></td>
<td>1.606***</td>
<td>-1.001**</td>
</tr>
<tr>
<td></td>
<td>(0.484)</td>
<td></td>
<td>(0.454)</td>
<td>(0.444)</td>
</tr>
<tr>
<td>C3</td>
<td>1.180***</td>
<td></td>
<td>1.897***</td>
<td>-0.757*</td>
</tr>
<tr>
<td></td>
<td>(0.483)</td>
<td></td>
<td>(0.457)</td>
<td>(0.444)</td>
</tr>
<tr>
<td>CHILD</td>
<td>1.157***</td>
<td></td>
<td>-0.447**</td>
<td>0.395**</td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td></td>
<td>(0.185)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.492E-5*</td>
<td></td>
<td>0.545E-5**</td>
<td>0.529E-5**</td>
</tr>
<tr>
<td></td>
<td>(0.276E-5)</td>
<td></td>
<td>(0.243E-5)</td>
<td>(0.264E-5)</td>
</tr>
<tr>
<td>INCOMEDUMMY</td>
<td>-0.766***</td>
<td></td>
<td>0.573**</td>
<td>-0.334</td>
</tr>
<tr>
<td></td>
<td>(0.267)</td>
<td></td>
<td>(0.241)</td>
<td>(0.259)</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.614***</td>
<td></td>
<td>0.558***</td>
<td>0.552***</td>
</tr>
<tr>
<td></td>
<td>(0.229)</td>
<td></td>
<td>(0.163)</td>
<td>(0.163)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.856***</td>
<td></td>
<td>0.576***</td>
<td>1.149***</td>
</tr>
<tr>
<td></td>
<td>(0.237)</td>
<td></td>
<td>(0.172)</td>
<td>(0.188)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-1.356***</td>
<td></td>
<td>-1.103***</td>
<td>-1.120***</td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
<td></td>
<td>(0.243)</td>
<td>(0.245)</td>
</tr>
<tr>
<td>UNDER</td>
<td>-0.404***</td>
<td></td>
<td>-0.521***</td>
<td>-0.284***</td>
</tr>
<tr>
<td></td>
<td>(0.110)</td>
<td></td>
<td>(0.101)</td>
<td>(0.984E-1)</td>
</tr>
<tr>
<td>WORK</td>
<td>-0.546***</td>
<td></td>
<td>-0.364***</td>
<td>-0.225***</td>
</tr>
<tr>
<td></td>
<td>(0.772E-1)</td>
<td></td>
<td>(0.844E-1)</td>
<td>(0.830E-1)</td>
</tr>
<tr>
<td>ONE-OFF</td>
<td>0.258***</td>
<td></td>
<td>0.287***</td>
<td>0.343***</td>
</tr>
<tr>
<td></td>
<td>(0.889E-1)</td>
<td></td>
<td>(0.716E-1)</td>
<td>(0.741E-1)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.106E-1***</td>
<td></td>
<td>-0.136E-1***</td>
<td>-0.145E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.135E-2)</td>
<td></td>
<td>(0.112E-2)</td>
<td>(0.120E-2)</td>
</tr>
<tr>
<td>JOBS</td>
<td>-0.700E-3</td>
<td></td>
<td>0.291E-2**</td>
<td>0.155E-2</td>
</tr>
<tr>
<td></td>
<td>(0.157E-2)</td>
<td></td>
<td>(0.132E-2)</td>
<td>(0.974E-3)</td>
</tr>
<tr>
<td>AREA</td>
<td>-0.480E-3</td>
<td></td>
<td>0.527E-3*</td>
<td>0.484E-3***</td>
</tr>
<tr>
<td></td>
<td>(0.409E-3)</td>
<td></td>
<td>(0.321E-3)</td>
<td>(0.165E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.163**</td>
<td></td>
<td>-0.128**</td>
<td>-0.350***</td>
</tr>
<tr>
<td></td>
<td>(0.704E-1)</td>
<td></td>
<td>(0.535E-1)</td>
<td>(0.605E-1)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.410E-1***</td>
<td></td>
<td>0.437E-1***</td>
<td>0.614E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.141E-1)</td>
<td></td>
<td>(0.109E-1)</td>
<td>(0.119E-1)</td>
</tr>
</tbody>
</table>

**Summary statistics**

- Log-likelihood: -703.209, -1041.689, -944.204
- Chi-squared (constants only): 219.618, 361.570, 406.128
- $\hat{R}^2$ adjusted: 0.128, 0.148, 0.171
- Number of observations: 742 (367 skipped), 1121 (308 skipped), 1045 (178 skipped)

Note: ***significant at 1%, **significant at 5%, *significant at 10%

standard errors are in brackets
10.3 Transferability Across Sites

The third hypothesis of this thesis is tested in this section. This hypothesis is that benefit transfer across sites is valid. Consistent with previous tests in the benefit transfer literature, it is tested by determining whether there is convergent validity. Four different tests are conducted to determine whether this hypothesis should be accepted or rejected. These tests focus on the equality of (1) overall models, (2) overall models after allowing for differences in variance, (3) implicit prices and (4) estimates of compensating surplus.

Test 1: the equality of overall models

The first test of this hypothesis is whether the overall models are equivalent. Stated formally, the null and alternative hypotheses for this test are:

\[ H_0: \beta_{MM} = \beta_{GS} \]
\[ H_1: \beta_{MM} \neq \beta_{GS} \]

where \( \beta_{MM} \) and \( \beta_{GS} \) are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney data sets.

Before conducting a formal test of the equality of taste parameter vectors it is useful to conduct a visual examination of the similarity of the vectors. Figure 10.1 plots the parameter vectors shown in Table 10.1. Excluding the alternative specific constants, there appears to be some indication of proportionality, but still a relatively large amount of variance. This suggests that the hypothesis of equal models will probably be rejected. The slope of the line of best fit (excluding the constants) appears to be close to one, suggesting that rescaling will have limited effect on the next hypothesis test.
A likelihood ratio test is used to determine whether $H_0$ should be rejected. The test statistic is:

$$-2[L_{MM+GS} - (L_{MM} - L_{GS})] \sim \chi^2 \text{ with } r \text{ degrees of freedom}$$

where $L_{MM+GS}$ is the log-likelihood calculated using both data sets, $L_{MM}$ and $L_{GS}$ are the log-likelihoods calculated using the Macquarie Marshes and Gwydir Sydney data sets, and $r$ is the number of parameters in each of the models.

The log-likelihoods for the Gwydir Sydney and Macquarie Marshes model are reported in Table 10.1. The log-likelihood of the combined model without rescaling, which is reported in Table 10.2, is 2012.05. The test statistic is therefore:

$$-2 [-2012.05 - (-1041.69 + -944.20)] = 52.32$$

The critical value for this test given 16 degrees of freedom is 26.30 at the 5% level. Hence we reject the null hypothesis and conclude that the two models are not equivalent overall.
Table 10.2: MNL models estimated using the combined Gwydir Sydney and Macquarie Marshes data sets

<table>
<thead>
<tr>
<th>Variables</th>
<th>Combined model without rescaling</th>
<th>Combined model with rescaling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>0.328</td>
<td>0.587**</td>
</tr>
<tr>
<td></td>
<td>(0.306)</td>
<td>(0.252)</td>
</tr>
<tr>
<td>C3</td>
<td>0.586**</td>
<td>0.845***</td>
</tr>
<tr>
<td></td>
<td>(0.308)</td>
<td>(0.253)</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.245E-2</td>
<td>-0.884E-2</td>
</tr>
<tr>
<td></td>
<td>(0.122)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.526E-5***</td>
<td>0.530E-5***</td>
</tr>
<tr>
<td></td>
<td>(0.173E-5)</td>
<td>(0.194E-5)</td>
</tr>
<tr>
<td>INCOMEDUMMY</td>
<td>0.162</td>
<td>0.772E-1</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.194)</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.544***</td>
<td>0.591***</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.854***</td>
<td>0.971***</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.140)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-1.095***</td>
<td>-1.266***</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(0.187)</td>
</tr>
<tr>
<td>UNDER</td>
<td>-0.354***</td>
<td>-1.446***</td>
</tr>
<tr>
<td></td>
<td>(0.684E-1)</td>
<td>(0.752E-1)</td>
</tr>
<tr>
<td>WORK</td>
<td>-0.302***</td>
<td>-0.374***</td>
</tr>
<tr>
<td></td>
<td>(0.572E-1)</td>
<td>(0.613E-1)</td>
</tr>
<tr>
<td>ONE-OFF</td>
<td>0.315***</td>
<td>0.336***</td>
</tr>
<tr>
<td></td>
<td>(0.496E-1)</td>
<td>(0.548E-1)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.139E-1***</td>
<td>-0.158E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.813E-3)</td>
<td>(0.914E-3)</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.223E-2***</td>
<td>0.245E-2***</td>
</tr>
<tr>
<td></td>
<td>(0.766E-3)</td>
<td>(0.825E-3)</td>
</tr>
<tr>
<td>AREA</td>
<td>0.313E-3**</td>
<td>0.390E-3***</td>
</tr>
<tr>
<td></td>
<td>(0.136E-3)</td>
<td>(0.144E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.223***</td>
<td>-0.249***</td>
</tr>
<tr>
<td></td>
<td>(0.404E-1)</td>
<td>(0.449E-1)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.512E-1***</td>
<td>0.565E-1***</td>
</tr>
<tr>
<td></td>
<td>(0.797E-2)</td>
<td>(0.897E-2)</td>
</tr>
</tbody>
</table>

Summary statistics

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>-2012.048</td>
</tr>
<tr>
<td></td>
<td>Chi-squared (constants only)</td>
<td>719.888</td>
</tr>
<tr>
<td></td>
<td>$p^2$ adjusted</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>Number of observations</td>
<td>2166 (486 skipped)</td>
</tr>
</tbody>
</table>

Note: ***significant at 1%, **significant at 5%, *significant at 10%

standard errors are in brackets
Test 2: the equality of overall models after allowing for differences in variance

The next test involves determining whether the models are equivalent after allowing for differences in variance. As discussed in Swait and Louviere (1993), MNL models have a scale parameter ($\lambda$) which is inversely proportional to variance, but is confounded with the $\beta$ vector. While it is not possible to estimate the scale parameter directly, it is possible to identify the ratio of scale parameters between two data sets. This ratio can be estimated by stacking the two data sets and conducting a one-dimensional grid search using different values of the scale parameter. The correct value of the scale parameter is found when the log-likelihood of the MNL model estimated using the stacked data set is maximised. Using this log-likelihood value it is then possible to test the hypothesis that two data sets are equal, except for differences in variance (Swait and Louviere 1993)\textsuperscript{84}. This is the objective of the second test:

\[ H_0: \beta_{\text{MM}} = \lambda \beta_{\text{GS}} \]
\[ H_1: \beta_{\text{MM}} \neq \lambda \beta_{\text{GS}} \]

where $\beta_{\text{MM}}$ and $\beta_{\text{GS}}$ are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney data sets, and $\lambda$ represents the ratio of scale factors.

Following Swait and Louviere (1993), a likelihood ratio test is used to determine whether $H_0$ should be rejected. The test statistic is:
-2[L_\lambda-(L_{MM}+L_{GS})] \sim \chi^2 \text{ with } r+1 \text{ degrees of freedom}

where $L_\lambda$ is the log-likelihood calculated using the combined data set that has been rescaled, $L_{MM}$ is the log-likelihood using the Macquarie Marshes data set, $L_{GS}$ is the log-likelihood calculated using the Gwydir Sydney data set and $r$ is the number of parameters in each of the models.

The log-likelihood for the combined and rescaled model is 2007.50 (see Table 10.2)\(^85\). The test statistic is therefore:

\[-2 [-2007.50 - (-1041.69 + -944.20)] = 43.22\]

There is now an extra degree of freedom for this test because the scale parameter is allowed to vary between the models. There are therefore 17 degrees of freedom and the critical value at the 5% level is 27.59. The null hypothesis is again rejected and it can be concluded that the two models are different, even after rescaling\(^86\).

**Test 3: the equivalence of implicit prices across sites**

The third test focuses on the equality of implicit prices. Implicit prices, which are also known as part-worths, are point estimates of the value of a unit change in a non-monetary attribute. As discussed in Chapter 3, while implicit prices are not measures of changes in welfare, they provide information about the value respondents place on a particular aspect of a resource. The null and alternative hypotheses for this test are as follows:

\[^{84}\text{Swait, Louviere and Williams (1994) and Louviere and Swait (1996) recommended that alternative specific constants not be rescaled. The alternative specific constants capture the mean of the unobserved variables, which are likely to vary across sites, and may be subject to a different level of variance to the observed attributes.}\]

\[^{85}\text{The ratio of the scale parameters was 0.78. As the variance of the Gwydir Sydney model was set equal to one, this indicates that the variance of the Macquarie Marshes data is greater than that of the Gwydir Sydney data.}\]

\[^{86}\text{It should, however, be noted that this result is sensitive to the specification of the models. The different sign of the coefficient for CHILD would be expected to affect the equality of the models. When this variable is deleted from the two models, the null hypothesis that the models are equivalent after allowing for differences in variance is rejected at the 5\% level, but not at the 1\% level.}\]
H_0: \beta_{(MM)i}/\beta_{(MM)M} = \beta_{(GS)i}/\beta_{(GS)M}

H_1: \beta_{(MM)i}/\beta_{(MM)M} \neq \beta_{(GS)i}/\beta_{(GS)M}

where \beta_i is a non-monetary choice set attribute (i= 1,..., n-1), \beta_M is the monetary attribute, MM represents the Macquarie Marshes model, and GS represents the Gwydir Sydney model.

Testing this hypothesis is less straightforward than the previous hypotheses. This is because standard errors for implicit prices are not directly calculated in the MNL model. Alternative procedures must be used to derive confidence intervals. The Krinsky and Robb (1986) procedure is used in this thesis. This procedure involves randomly drawing a large number of parameter vectors from a multivariate normal distribution with mean equal to the \beta vector and variance equal to the variance-covariance matrix from the estimated MNL model. Implicit prices are then estimated using each of the parameter vectors that are drawn from the normal distribution and confidence intervals can be calculated. Confidence intervals were estimated using 500 draws. The \((100 - \alpha)\%\) confidence interval is then found by ranking the 500 resulting implicit prices and removing the top and bottom \(\alpha/2\%) of the implicit prices from the final ranking (Park, Loomis and Creel 1991).

Implicit prices and 95% confidence intervals for the two models are shown in Table 10.3. Confidence intervals overlap for each of the four implicit prices, indicating a degree of similarity.

<table>
<thead>
<tr>
<th>Jobs</th>
<th>Area</th>
<th>Breed</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macquarie</td>
<td>10.7 cents</td>
<td>3.4 cents</td>
<td>$24.15 ($15.83, $33.72)</td>
</tr>
<tr>
<td>Marshes</td>
<td>(-2.7, 23.8)</td>
<td>(1.1, 5.7)</td>
<td></td>
</tr>
<tr>
<td>Gwydir</td>
<td>21.1 cents</td>
<td>3.9 cents</td>
<td>$9.81 ($2.40, $17.42)</td>
</tr>
<tr>
<td>Sydney</td>
<td>(5.1, 40.1)</td>
<td>(0.9, 8.4)</td>
<td></td>
</tr>
<tr>
<td>P-values</td>
<td>0.142</td>
<td>0.400</td>
<td>0.006</td>
</tr>
<tr>
<td>(Poe et al test)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Poe, Severance-Lossin and Welsh (1994) have, however, demonstrated that overlapping confidence intervals, generated using the Krinsky-Robb (1986) procedure, provide an inaccurate test of the equality of mean estimates. They show that the actual significance given by overlapping confidence intervals does not correspond to the stated level of significance implied, and is actually more conservative. Hence, type two errors are more likely to occur.

One alternative that Poe et al (1994) propose for testing the equality of means is to calculate differences between the two random distributions developed using the Krinsky-Robb (1986) procedure. A one-sided approximate significance level is calculated by the proportion of negative values in the distribution of differences, depending on which mean is thought to be greater (Poe, Welsh and Champ 1997; Foster and Mourato 1998). The results from this test are reported in the last row of Table 10.3. The implicit prices for JOBS, AREA and SPECIES are shown to still be equivalent. However, the implicit price for BREED is shown to be significantly different at the 1% level.

To illustrate their argument, Poe et al (1994) use as an example a 90% confidence interval for the mean of an estimate ie $X \pm 1.6450 \times \text{S.E.}$, where $X$ is the mean. Assuming two normal distributions with equal variances and sample sizes, the critical difference of means $(X-Y)$ associated with non-overlapping 90% confidence intervals would have to be at least 3.290 standard errors apart before they would be considered statistically different. However, if this value is substituted into a Z test of differences, the implied significance level is 2% rather than 10%. Hence the significance level is overstated when confidence intervals are used to test the equality of means.

The equality of mean estimates were also tested using non-parametric Mann-Whitney tests (see Morrison, Bennett, Blamey and Louviere 1998b). The Mann-Whitney test involves examining a ranking of the two sets of observations to determine if they are randomly scattered throughout the set of pooled data. The null hypothesis of equality of implicit prices was consistently rejected using these tests. However, the Mann-Whitney test requires the assumption that both samples were drawn from the same population. This is an unlikely assumption in this case, and hence the rejection of the null hypothesis may simply reflect different shapes of the distribution rather than differences in mean estimates. Therefore this test is likely to be less accurate than the Poe et al (1994) test.
Test 4: the equivalence of estimates of compensating surplus across sites

This test is the most important of the four, because deriving welfare estimates is the primary objective of benefit transfer. Moreover, the acceptance of the previous three hypotheses need not imply that this hypothesis will be accepted. Even if the overall models and implicit prices are equal, it is possible that estimates of compensating surplus will be different, or vice versa. Stated formally, the hypothesis for this test is:

\[ H_0: CS_{MMi} = CS_{GSi} \]
\[ H_1: CS_{MMi} \neq CS_{GSi} \]

where \( CS \) represents compensating surplus, MM represents the Macquarie Marshes model, GS represents the Gwydir Sydney model, and there are \( i=1,\ldots,n \) alternatives.

This hypothesis is difficult to test with CM because it is possible to derive numerous compensating surplus estimates from the models, depending on the levels of the attributes selected. One way of performing this test is to specify one or more policy relevant alternatives and compare estimates of compensating surplus. A limitation of this approach is that the magnitude of difference may diverge depending on the improvement chosen. Therefore limited information is provided about the transferability of welfare estimates. Some systematic way is needed to sample from the myriad of possible environmental improvements that can be valued using a choice model.

For this thesis, an experimental design was used to sample from the set of possible environmental improvements. Three levels were selected for each of the four non-monetary attributes (JOBS, AREA, BREED and SPECIES), based on the changes defined in the Gwydir Wetlands questionnaire. A one-ninth fraction of the \( 3^4 \) full factorial was taken, resulting in the selection of nine representative alternatives (see Table
Estimates of compensating surplus and confidence intervals are reported for each model. Confidence intervals were again calculated using the Krinsky and Robb (1986) procedure. Poe et al (1997) tests were also conducted to determine whether mean estimates were statistically equivalent. The socioeconomic characteristics were set at the population mean levels when estimating compensating surplus.

The results in Table 10.4 are mixed. The confidence intervals overlap for five out of nine alternatives, but the Poe et al (1994) test shows that only two of the mean estimates are equivalent at the 5% significance level, and three at the 1% level. This suggests that benefit transfer is valid for only some estimates of compensating surplus. Information about the magnitude of errors likely to be experienced when using benefit transfer is provided by determining the percentage mean difference in the estimates of compensating surplus. For these nine alternatives the differences in mean estimates range from 4% to 66%, with a mean of 32%. While the lower end of the reported range would be acceptable to most policy makers, the upper end of the range could cause concern.

A relevant question is, therefore, when are the welfare estimates likely to converge? The answer to this question will provide indications about when benefit transfer will be most valid. The estimates of compensating surplus tend to be closest when the changes in the attributes are largest. This is
because the alternative specific constants\textsuperscript{89} and the implicit price for JOBS were much larger in the Gwydir Sydney model, hence smaller changes were valued more highly than in the Macquarie Marshes model. In contrast, the environmental attributes were valued more highly in the Macquarie Marshes model, hence the estimates of compensating surplus converge with larger changes.

Table 10.4: Estimates of compensating surplus

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Change in attributes</th>
<th>Macquarie Marshes</th>
<th>Gwydir Sydney</th>
<th>P-value (Poe et al)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jobs-no change; Area +150 km(^2); Breed +1 year; Species +4</td>
<td>$27.83</td>
<td>$80.96</td>
<td>$6.64, $48.55 ($63.27, $101.81)</td>
</tr>
<tr>
<td>2</td>
<td>Jobs-no change; Area +350 km(^2); Breed +2 years; Species +13</td>
<td>$97.20</td>
<td>$127.57</td>
<td>$79.60, $113.64 ($111.08, $146.06)</td>
</tr>
<tr>
<td>3</td>
<td>Jobs-no change; Area +500 km(^2); Breed +3 years; Species +8</td>
<td>$105.13</td>
<td>$127.25</td>
<td>$90.27, $121.76 ($106.27, $147.57)</td>
</tr>
<tr>
<td>4</td>
<td>Jobs -20; Area +150 km(^2); Breed +2 years; Species +8</td>
<td>$66.90</td>
<td>$99.25</td>
<td>$51.62, $82.79 ($85.28, $116.62)</td>
</tr>
<tr>
<td>5</td>
<td>Jobs -20; Area +350 km(^2); Breed +3 years; Species +4</td>
<td>$80.81</td>
<td>$104.11</td>
<td>$65.22, $98.40 ($87.21, $120.00)</td>
</tr>
<tr>
<td>6</td>
<td>Jobs -20; Area +500 km(^2); Breed +1 year; Species +13</td>
<td>$76.03</td>
<td>$119.31</td>
<td>$59.07, $92.77 ($100.69, $139.18)</td>
</tr>
<tr>
<td>7</td>
<td>Jobs -100; Area +150 km(^2); Breed +3 years; Species +13</td>
<td>$103.80</td>
<td>$107.63</td>
<td>$85.98, $123.02 ($89.10, $128.82)</td>
</tr>
<tr>
<td>8</td>
<td>Jobs -100; Area +350 km(^2); Breed +1 year; Species +8</td>
<td>$41.00</td>
<td>$79.86</td>
<td>$26.22, $58.01 ($62.67, $97.42)</td>
</tr>
<tr>
<td>9</td>
<td>Jobs -100; Area +500 km(^2); Breed +2 years; Species +4</td>
<td>$53.20</td>
<td>$82.75</td>
<td>$38.63, $67.07 ($64.84, $102.15)</td>
</tr>
</tbody>
</table>

\textsuperscript{89} The alternative specific constants have not been included when estimating compensating surplus in some studies (eg Adamowicz, Boxall, Williams and Louviere 1998; Hanley, MacMillan, Wright, Bullock, Simpson, Parssson and Crabtree 1998; Hanley, Wright and Adamowicz 1998). However, as the alternative specific constants are equal to the mean of the error terms for each alternative, and the error term is assumed (under the RUM) to represent unobserved utility, there is reason to include them when estimating welfare changes. In other words, because the alternative specific constants reflect part of the reason respondents chose to improve environmental quality they should be included. The argument against including alternative specific constants is that they may reflect confusion, represent symbolic (or yea-saying) responses, or even nea-saying. This is a critical issue, as the results reported above to some extent depend on the decision to include the alternative specific constants. Further research to determine what effects the magnitude and sign of the alternative specific constants would appear to be particularly important. A sensitivity analysis was conducted to determine the effect of excluding alternative specific constants on this hypothesis test. When the constants were excluded, the mean difference in willingness to pay increased to 45% and all nine of the estimates of compensating surplus were found to be statistically different using a Poe et al test.
In conclusion, although the results from these four tests are somewhat mixed, the weight of evidence is against the equivalence of value estimates. The results about the equality of models and implicit prices are important, because they provide information about the structure of people’s preferences. However, the remaining hypothesis about the equality of estimates of compensating surplus is more important. This is because compensating surplus is what is used in cost-benefit analyses, and hence is the primary focus for benefit transfer. The tests of equality of compensating surplus estimates provides little support for the hypothesis of convergent validity. However, for the two models reported here, the value estimates are closer for larger improvements. But for other models the results may converge for different improvements. Unfortunately, insufficient information is available from this study to answer this question adequately. More research is needed to understand the factors that affect valuation estimates in different contexts. Further research is also needed to clarify the role of alternative specific constants in welfare estimation.

10.4 Transferability Across Populations

The fourth and final hypothesis of this thesis is tested in this section. This hypothesis is that benefit transfer across populations is valid. This is tested by comparing the results from the Gwydir Wetlands surveys in Moree and Sydney. The same tests that were used in the previous section are relevant for this section. However, because only three of the five coefficients for the choice set attributes are correctly signed and significant in the Gwydir Moree model, the results from the fourth test are self-evident. Hence only the first three tests are repeated.
Test 1: the equality of overall models

As with the transfer across sites, the null and alternative hypotheses are as follows, with the test statistic the same as specified in the previous section:

\[ H_0: \beta_{GM} = \beta_{GS} \]
\[ H_1: \beta_{GM} \neq \beta_{GS} \]

where \( \beta_{GM} \) and \( \beta_{GS} \) are taste parameter vectors corresponding to the Gwydir Moree and Gwydir Sydney data sets.

Similar to the previous section, a visual examination of the equality of taste parameter vectors is conducted. Figure 10.2 plots the appropriate parameter vectors shown in Table 10.1. There appears to be slightly more proportionality than with the previous graph, with the exception of CHILD and INCDUM. This contrasts with the previous graph where the alternative specific constants were most different. The line of best fit appears to have a slope of less than one, indicating that the ratio of the Gwydir Sydney and Gwydir Moree parameter vectors is also less than one. This implies that the variance of the Gwydir Sydney data set is lower and that there may be some scope for rescaling the data.

Figure 10.2: Plot of Gwydir Moree and Gwydir Sydney parameter vectors
The log-likelihoods for the Gwydir Wetlands Sydney and Moree models are reported in Table 10.1. The log-likelihood of the combined model, which is reported in Table 10.5, is 1780.09. The test statistic is therefore:

\[-2 \left[-1780.09 - (-703.21 + -1041.69)\right] = 70.38\]

The critical value for this test at the 5% level given 16 degrees of freedom is 26.30. Hence the null hypothesis is again rejected and it is concluded that the two models are not equivalent overall. Moreover, it appears that the overall difference between the models is greater than for the transfer across sites.
Table 10.5: MNL models estimated using the combined Gwydir
Sydney and Gwydir Moree data sets

<table>
<thead>
<tr>
<th>Variables</th>
<th>Combined model without rescaling</th>
<th>Combined model with rescaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.268*** (0.319)</td>
<td>1.381*** (0.290)</td>
</tr>
<tr>
<td>C2</td>
<td>1.558*** (0.320)</td>
<td>1.665*** (0.291)</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.274** (0.126)</td>
<td>0.358*** (0.137)</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.124E-7 (0.72E-5)</td>
<td>-0.106E-5 (0.190E-5)</td>
</tr>
<tr>
<td>INCOMEDUMMY</td>
<td>0.517E-1 (0.173)</td>
<td>-0.497E-1 (0.189)</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.295*** (0.115)</td>
<td>0.448*** (0.125)</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.776*** (0.133)</td>
<td>0.823*** (0.150)</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-1.062*** (0.183)</td>
<td>-1.211*** (0.203)</td>
</tr>
<tr>
<td>UNDER</td>
<td>-0.379*** (0.708E-1)</td>
<td>-0.448*** (0.770E-1)</td>
</tr>
<tr>
<td>WORK</td>
<td>-0.455*** (0.542E-1)</td>
<td>-0.509*** (0.571E-1)</td>
</tr>
<tr>
<td>ONE-OFF</td>
<td>0.270*** (0.534E-1)</td>
<td>0.278*** (0.593E-1)</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.122E-1*** (0.852E-3)</td>
<td>-0.133E-1*** (0.943E-3)</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.134E-2 (0.995E-3)</td>
<td>0.131E-2 (0.110E-2)</td>
</tr>
<tr>
<td>AREA</td>
<td>0.105E-3 (0.250E-3)</td>
<td>0.352E-4 (0.276E-3)</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.140*** (0.433E-1)</td>
<td>-0.148*** (0.476E-1)</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.413E-1*** (0.855E-2)</td>
<td>0.448E-1*** (0.944E-2)</td>
</tr>
</tbody>
</table>

Summary statistics

- Log-likelihood: -1780.090, -1777.093
- Chi-squared (constants only): 522.386, 528.380
- $\rho^2$ adjusted: 0.127, 0.128
- Number of observations: 1863 (675 skipped), 1863 (675 skipped)

Note: ***significant at 1%, **significant at 5%, *significant at 10%
standard errors are in brackets
Test 2: the equality of overall models after allowing for differences in variance

The next test is similar to the one conducted for the transfer across sites. The hypothesis is as follows:

\[
H_0: \beta_{GM} = \lambda \beta_{GS} \\
H_1: \beta_{GM} \neq \lambda \beta_{GS}
\]

where \( \beta_{MM} \) and \( \beta_{GS} \) are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney data sets, and \( \lambda \) represents the ratio of scale factors.

The log-likelihood for the combined and rescaled model is 1777.09\(^{90}\). The test statistic is therefore:

\[-2 \left[ -1777.09 - (-703.21 + -1041.69) \right] = 64.38\]

The test statistic exceeds the critical value at the 5% level of 27.59. Therefore the null hypothesis is again rejected and it is concluded that the two models are not equivalent, even after rescaling to allow for differences in variance.

Test 3: the equivalence of implicit prices across populations

The third test focuses on the equality of implicit prices. The null and alternative hypotheses for this test are:

\(^{90}\) The ratio of scale parameters was 0.84, indicating that the Gwydir Moree data set has larger variance than the Gwydir Sydney data set.
\[ H_0: \beta_{(GM)i}/\beta_{(GM)M} = \beta_{(GS)i}/\beta_{(GS)M} \]
\[ H_1: \beta_{(GM)i}/\beta_{(GM)M} \neq \beta_{(GS)i}/\beta_{(GS)M} \]

where \( \beta_i \) is a non-monetary choice set attribute \((i=1, \ldots, n-1)\), \( \beta_M \) is the monetary attribute, GM represents the Gwydir Moree model and GS represents the Gwydir Sydney model.

Implicit prices and 95% confidence intervals for the two models are shown in Table 10.6. The confidence intervals for each of the four implicit prices again overlap. However, the Poe et al (1994) tests show that only the implicit prices for BREED and SPECIES are equivalent.

| Table 10.6: Implicit prices and confidence intervals for the Gwydir Sydney and Gwydir Moree models |
|-------------------------------------------------|-----------------|----------------|-----------------|-----------------|
| Jobs Area Breed Species Gwydir Moree | Jobs Area Breed Species Gwydir Moree | Jobs Area Breed Species Gwydir Moree | Jobs Area Breed Species Gwydir Moree |
| Gwydir | -7.7 cents* | -4.7 cents | $15.18 | $3.86 |
| Moree | (-42.9, 21.7) | (-12.6, 3.5) | ($2.00, $29.95) | ($1.15, $6.73) |
| Gwydir | 21.8 cents | 3.9 cents | $9.81 | $3.21 |
| Sydney | (5.1, 40.1) | (0.9, 8.4) | ($2.40, $17.42) | ($1.50, $4.71) |
| P-value | 0.056 | 0.030 | 0.260 | 0.348 |
| (Poe et al) | insignificant coefficients in the Gwydir Moree model |

The results from these three tests are generally less positive than for the previous set. The overall models were found to be statistically different with and without rescaling. While this is the same result as for the previous set of tests, the differences tend to be greater. Perhaps of most concern for transfers across populations is the insignificance and incorrect sign of the coefficients of two of the choice set attributes. The incorrect signs would be expected to cause estimates of compensating surplus to diverge substantially, and hence benefit transfer would be subject to considerable error. The implication is that transfers between rural and urban populations should be treated with considerable caution. This also may explain why some previous CVM studies which combined transfers across both sites and populations have rejected the hypothesis of convergent validity (eg Bergland et al 1995).
The difficulty in transferring benefit estimates between rural and urban sites reflects heterogeneity of preferences across the populations. Attempts were made to explain this heterogeneity by including socioeconomic indicators as explanatory variables in the MNL models. However, they were not sufficient for explaining the heterogeneity of respondents' preferences. One way of improving the accuracy of benefit transfer between urban and rural areas would be to identify the underlying factors that drive peoples' preferences, and to include these within any models that are estimated.

10.5 Conclusion

The objective of this chapter has been to test two hypotheses regarding the validity of benefit transfer. The results from the tests of these hypotheses are not particularly positive regarding the validity of benefit transfer, although some of the results were mixed. Generally the weight of evidence appears to be against the convergent validity of both transfers across sites and populations. However, the transfers across sites tend to be less problematic than across population transfers, with estimates of compensating surplus being, at times, statistical equivalent. This suggests that transfers between urban and rural populations may be subject to greater error than transfers across sites for a given urban population. However, further research would be needed to confirm this finding before firm conclusions can be drawn.

As reported in Chapter 9, Bergland et al (1995) suggested that techniques such as CM, that allow for both different changes in environmental quality and differences in socioeconomic characteristics, are likely to be more suitable for benefit transfer. While the ability to allow for differences in environmental improvements is an obvious advantage of CM, the values associated with implicit prices (and alternative specific constants) need not
be equal. Hence there still may be divergence between estimates of compensating surplus.

One further issue that has been hinted at in this chapter concerns the way validity is generally measured. Most researchers have tested whether there is convergent validity; that is, whether estimates are statistically equivalent. As shown by the tests presented in this chapter, statistical equivalence can be a very exacting measure. This notion of validity, however, may not win much support from policy makers who could be content if estimates of compensating surplus differ by only 20-30%. For most policy makers, while it is useful to know when value estimates will differ, an equally important question is when are these differences likely to fall within an acceptable range? Policy makers often approach non-market valuation from a threshold value perspective. Errors of 30-50% may be considered trivial if the non-market benefits clearly outweigh any costs (eg benefits three times the costs). Therefore an understanding of the errors likely to be experienced when using benefit transfer is useful for determining whether this source of error could influence the results of a cost-benefit analysis. If the error is great enough to have policy implications, then there may be a rationale for undertaking a primary study. Hence of more importance than statistical significance may be questions of economic significance (McCloskey and Ziliak 1996).

Some evidence to answer this question has been reported in this thesis. The results from this thesis suggest that larger errors are likely to be found for transfers across sites when smaller environmental improvements are being valued. However, different results may be found in other contexts. Further research is needed to answer this question accurately.

Finally, it should be noted that the mixed findings of this thesis in terms of the transferability of value estimates either across sites or populations does have a positive side. If value estimates were found to be equivalent across sites and populations, it is possible that some researchers would argue that
the results were subject to perfect-embedding (Bennett, Morrison and Blamey 1997). In other words, respondents were insensitive to changes in the nature or scale of the good. These results, by showing that respondents are willing to pay different amounts for different wetlands, and from different locations, enhance the validity of CM.
Chapter 11 Conclusion

The first purpose of this chapter is to summarise briefly the method and conclusions of this thesis. No attempt is made to detail exhaustively all of the conclusions as these have been reported at the end of each chapter. Rather, the broad conclusions from throughout the thesis are recalled. The second purpose is to enumerate the most significant findings in this thesis, and their implications in terms of policy and the need for further research. The third purpose is to list the caveats associated with the research presented in this thesis.

11.1 Summary

This thesis has involved an investigation of the validity of using CM to estimate non-use values, and the validity of reusing these estimates in different contexts, which is known as benefit transfer.

The literature relating to the estimation of non-use values was reviewed in the first part of the thesis. At the beginning of this review the economic theory relevant for estimating non-use values was examined to provide a basis for analysing how resource use changes affect community welfare. Several components of non-use value were examined, including option value, quasi-option value and existence value. It was noted that in many cases it is not possible to know \textit{a priori} the sign of option value, and hence researchers should focus on estimating option price, which is the sum of use value and option value. Quasi-option value was shown to reflect the value of obtaining further information. While it is not possible to know \textit{a priori} the magnitude of quasi-option value, it was argued that respondents
might form subjective opinions about its magnitude, and this might comprise part of their willingness to pay to preserve a natural resource. Thirdly, existence value, which reflects peoples' willingness to pay to preserve a natural resource apart from any in situ use, was shown to be a function of respondents' motives. Motives were also shown to effect whether existence value can be estimated. The sum of use, option, quasi-option and existence value is equal to total economic value. Total economic value can be measured for each individual by using Hicksian surplus measures. Theoretical issues associated with aggregation of individual level estimates were also considered in Chapter 2.

Five different SP techniques, all of which are potential candidates for estimating non-use values, were then assessed. Several themes emerged from this assessment. The first was that the four conjoint techniques appear to have several advantages over the contingent valuation method, including more discriminating responses and the capacity to value multiple alternatives. In terms of the conjoint techniques it was found that: (1) not all techniques necessarily produce theoretically correct welfare estimates; (2) not all techniques have well defined behavioural bases; (3) the data gathered using each of the techniques is not of the same quality; and (4) differences could be expected in the cognitive burden experienced by respondents across the conjoint techniques. Of the five techniques evaluated, it was concluded that CM had the best potential for producing valid welfare estimates and was therefore selected for use later in the thesis.

In the fourth chapter the thesis, methodological issues related to applying CM were considered. These issues included the design of choice sets, experimental design, model selection and model evaluation. The main conclusions from this chapter were (1) that many of the decisions regarding choice set design are context specific, (2) that there is often need for compromise when selecting an experimental design, (3) that problems with computational burden and in model selection may be experienced
when using models that relax the property of independence of irrelevant alternatives; and (4) that tests for violations of the property of independence of irrelevant alternatives are powerful specification tests.

The three CM surveys were described in the second part of thesis. First, the two case study sites—the Macquarie Marshes and Gwydir Wetlands—were detailed. It was reported that the quality of both of these wetlands had been substantially affected by the development of extensive areas of irrigated agriculture. These two wetlands were chosen as case study sites as the information on the magnitude of non-use benefits from improving their quality would assist decision makers seeking socially optimal trade-offs involving the allocation of water between irrigation and wetland uses.

The results from the focus groups used to test draft CM questionnaires were then described. The draft questionnaire initially used in the focus groups was found to suffer from a number of significant problems including bias, confusion, implausibility and the use of inappropriate heuristics. While these problems were found to be less severe during latter focus groups, where revised versions of the draft questionnaire were used, confusion was a persistent problem. Further changes were thus made to reduce confusion in the final questionnaires. A pretest conducted in Sydney indicated that these measures were effective in reducing confusion.

The results from the three CM surveys were then presented in Chapter 7 (and Appendix 5). Three different model specifications were detailed. These were models that only had choice set attributes, those that included socioeconomic variables, and those that included socioeconomic and questionnaire evaluation variables. Each of the basic models had three or more significant coefficients, which was promising, but generally the explanatory power was low and IIA violations were found using mother logit tests. MNL models that allowed for heterogeneous preferences were more robust. The coefficients for more of the choice set attributes were
significant in these models, as well as for other independent variables. Significant IIA violations were either very small or non-existent, and the explanatory power was better. As reported in Appendix 5, other more complex models that relax the IIA property were also trialed; however, problems were found with these models. The first two hypotheses in the thesis were tested at the end of Chapter 7. The first was that CM can be used to provide valid estimates of non-use values, and the second was that non-use values exist for non-environmental attributes. Relatively strong support was found for the first of these hypotheses, however support for the second hypothesis was mixed.

Empirical issues associated with aggregation were then considered in Chapter 8 to show how policy relevant results might be derived, and the factors that have the most effect on results when aggregating individual level estimates. The main question raised in this chapter is how to allow for divergences between sample and population characteristics when aggregating individual level sample estimates. Two types of approaches were considered and trialed. The first involved adjusting mean estimates based on population characteristics. The second type involved making assumptions about the preferences of non-respondents. It was found that using the first type of approach had minor effects on aggregate value, while there were sizeable effects when the second type was used. The main implication is that assumptions about the preferences of non-respondents can substantially alter aggregate estimates.

The third part of the thesis focused on benefit transfer. The literature relating to benefit transfer was first reviewed in Chapter 8. Following the research agenda proposed by Boyle and Bergstrom (1993), studies were reviewed that had analysed the determinants of value estimates and tested the validity of benefit transfer. Various determinants of value were identified, including the characteristics of the non-market good, the cause of a change in environmental quality, respondents' socioeconomic characteristics and elicitation format. In terms of tests of validity, there
was evidence that value estimates are relatively stable over time, but only a few studies had examined benefit transfer across different sites or populations. These later studies showed that combining transfers (e.g., across both time and sites, or across both sites and populations) is likely to be more problematic than just one type of transfer. Also, transfers involving non-use values appear to be more problematic than transfers involving use values. The use of techniques that allow for different changes in environmental quality, as well as differences in socioeconomic characteristics, was recommended for tests of validity.

Two hypotheses regarding the validity of benefit transfer were tested in Chapter 10. The first was the benefit transfer across sites is valid, and the second was that benefit transfer across populations is valid. The weight of evidence appeared to be against the validity of both types of transfer. However, the transfers across sites tended to be less problematic than across population transfers.

11.2 Main findings and their Implications for Policy and Research

A number of significant findings have been made in this thesis. Some of these are related to the hypotheses that were tested. Others, however, were detailed in the literature reviewed or were discovered during the modelling process. The main findings from the thesis are enumerated, together with their implications for policy and further research, in this section.

1. Not all conjoint techniques produce valid estimates of non-use values

Increasingly resource economists are resorting to the use of conjoint techniques to derive estimates of non-market values. Many of these
applications have been reported in notable journals. It is therefore likely that they will spawn many more applications in the future. As described above, and principally in Chapter 3, there are serious concerns regarding the validity of using several of the techniques to estimate non-market and non-use values, especially if the estimates are to be used in cost-benefit analyses. Further research in this area is warranted to understand more fully the accuracy of welfare estimates generated using each of the techniques. Possible research topics include:

- a comparison of the results from modelling rankings data using (1) a traditional rankings model and (2) a binary choice model given that each of the choices contains the base alternative. This exercise would demonstrate whether sampling across attribute space affects the validity of welfare estimates derived using contingent ranking.

- a split sample comparison between applications of the ratings difference model and a binary choice model. This comparison would indicate whether the limitations of the ratings difference method (ie not allowing for differences in origin and no information about ‘cut-points’) affects the validity of welfare estimates.

2. The importance of using focus groups to design choice modelling questionnaires

Focus groups were used in this thesis to select attributes to include in choice sets, to identify relevant information to include in questionnaires, and to test several draft questionnaires. The first set of four focus groups demonstrated that a draft questionnaire designed by a researcher in isolation from respondents may have many flaws. The use of focus groups proved to be one of the most challenging but vital parts of the research presented in this thesis. Without carefully testing a questionnaire, it is possible to include inappropriate attributes in choice sets, and to have a questionnaire that is biased, confusing and contains inadequate or inappropriate information. A finding of this thesis is that applications that
have given little effort to focus grouping and pretesting should be interpreted with caution.

3. The validity of using choice modelling to produce estimates of non-use values

This was the first of the hypotheses in the thesis and was tested in Chapter 7. Some support was found for this hypothesis, however further research into the validity of using CM to estimate non-use values is warranted. This hypothesis was tested using two different measures of validity: theoretical and content validity. It should be noted that content validity is not usually tested in SP surveys. Nonetheless it is an important concern as it provides an indication of whether the results from a survey are meaningful. It is recommended that this type of validity is tested in future SP surveys. No tests of predictive validity were conducted in this thesis. These tests provide strong indications of validity and further research in this area, involving either experimental or field tests, would complement the tests that have been conducted in this thesis.

4. The existence of non-use values for non-environmental outcomes

This was the second hypothesis tested in this thesis and was originally proposed by Portney (1994). The results from the hypothesis tests indicated that non-use values for non-environmental outcomes can exist, but need not exist in all cases. The implication is that it may be advisable to include an attribute, such as jobs or regional income, in choice sets.

Two existing studies in the literature by Lockwood et al (1994) and Adamowicz et al (1998) found that non-use values for non-environmental outcomes were very small relative to non-use values for environmental outcomes, or non-existent. However, the research presented in this thesis
indicates that the inclusion of non-environmental outcomes can, in some contexts, reduce welfare estimates by a sizeable amount.

This finding may explain some other findings that have caused concern in the CVM literature. Two studies in particular (Imber et al. 1991 and RAC 1991) found that people were willing to pay less for large environmental improvements than smaller improvements. While this seems to be contrary to utility theory, it can be explained by the utility that people derive from knowing that others will retain their jobs. In each of the case studies mentioned above, there was a trade-off between improving environmental quality and reductions in mining or forestry. Hence when people were stating that they were willing to pay more for a smaller environmental improvement, it is likely that they were actually saying that they would be willing to pay more for a bundle that contains some environmental gains but few jobs lost. This implies that scope testing for validity of CVM studies (Bennett, Morrison and Blamey 1997) may be affected by the presence of non-use values for non-environmental outcomes.

5. The benefit of leaving dominated alternatives in experimental designs

Conventional wisdom suggests that it is preferable to generate experimental designs that have few dominated alternatives, and that it may be desirable to delete any remaining dominated alternatives. The rationale for this is the belief that dominated alternatives provide little information about respondents’ preferences. However, an alternative view is that dominated alternatives provide relatively more information about respondents’ preferences because (1) they establish that more is preferred to less, and (2) they are much simpler choices, and hence are likely to be subject to less variance. The results shown in Appendix 5 provide some evidence in support of this alternative view, suggesting that dominated alternatives should be left in experimental designs. However, more careful
modelling in which variance is made a function of whether a choice set contains a dominated alternative would provide a more accurate test of what is causing this phenomenon. Moreover, it is possible that dominance is just a special case of what constitutes 'complexity' in an experimental design. Further research may demonstrate the causes of complexity and which designs are likely to produce the simplest choice sets for respondents to answer.

6. The difficulty of using more complex models to deal with IIA violations

Several different ways of dealing with violations of the property of independence of irrelevant alternatives have been suggested in the literature. One way is to use more complex models that relax all or part of the assumption of IID errors, such as the multinomial probit, nested logit or heteroscedastic extreme value models.

However, researchers face a number of problems in using these models. First, as was found in this thesis, is the problem of computational burden. Models with very generalised error structures—such as the multinomial probit—are difficult to estimate using common computer hardware. Second, it is difficult to know which is the most appropriate model to use. Each of the models makes different assumptions about the error structure; hence it would be desirable to match a model to the problem being experienced. The problem is that the diagnostic tools only indicate that there is a specification problem; they do not indicate the cause of the problem. Third, it appears that these more complex models require large sample sizes if socioeconomic and attitudinal variables are to be included. Therefore, it may not be possible to use these models in some applications, except for simple specifications. It appears that the more complex models, while useful in some situations, can create additional problems and should not be relied upon as being a straightforward way of solving IIA
violations. Also, there is a need to develop new specification tests that can diagnose the cause of IIA violations.

7. The sensitivity of aggregate estimates to assumptions about the preferences of non-respondents

Various approaches have been used in the literature to allow for divergences between sample and population characteristics when aggregating. The approaches that attempt to adjust mean sample estimates using population characteristics—such as the use of independent variables in regression equations, weighted regression analysis and weighted averaging—were found to have minor effects on aggregate value estimates. However, assumptions about the preferences of non-respondents were found to have substantial impacts. At one extreme, assuming that non-respondents have zero willingness to pay results in substantially lower estimates, depending on the proportion of non-respondents. At the other extreme, assuming that they have the same preferences as respondents results in considerably larger aggregate estimates. Neither of these assumptions is completely correct, yet they can have significant policy implications by virtue of the impact they have on aggregate values. The degree of bias introduced during aggregation may well outweigh biases caused by strategising, yea-saying, IIA violations or scenario rejection. Hence some better way of dealing with non-respondents is desirable. Several methods of determining systematically the preferences of non-respondents have been trialed in the literature. These either involve extrapolation based on mail back survey returns, or using non-respondents’ socioeconomic data. The problem with both of these approaches is that the data used for extrapolation are only related indirectly to non-respondents preferences, and for many surveys socioeconomic data about non-respondents will be unavailable. The latter approach also requires more complex modelling. Another approach for determining the preferences of non-respondents was trialed in this thesis.
Non-respondents were asked why they were not prepared to respond, and their answers were used to infer willingness to pay. This approach has the potential to provide a simple way for gathering information about the preferences of non-responses that, in many surveys, would otherwise be unknown. However, further research is required to refine the questioning of non-respondents, and to develop understanding of the link between reasons given for non-response and preferences.

8. The validity of using choice modelling for benefit transfer

The final two hypotheses tested in this thesis were concerned with the validity of using CM for benefit transfer. Benefit transfer has increasingly been used in recent years, even though there is little understanding about whether it is valid and the circumstances under which it is likely to be problematic. The results from the tests reported in this thesis are generally not supportive of the convergent validity of benefit transfer, although further research is warranted, especially regarding the role of alternative specific constants.

Techniques such as CM have been touted as a solution for benefit transfer, as, unlike the CVM, they allow for different changes in environmental quality. This thesis has provided some justification for this belief. The equality of many implicit prices indicates that changes in attributes are often valued similarly. However, it need not be the case that the overall value of changes in the quality or quantity of a natural resource will be the same either across sites or populations.

It should be noted, though, that tests of convergent validity can be very exacting. Many policy makers are unlikely to be persuaded by the argument that benefit transfer is only valid if benefit estimates and models are statistically equivalent. Errors of between 20-30% may be quite acceptable for many policy makers, depending on the issue being
considered. In other words, economic significance is potentially of more importance than statistical significance. Decision makers frequently use threshold value models when making decisions involving natural resources (e.g., Saddler et al. 1980). If the benefits are shown to outweigh the costs substantially, then errors of this magnitude or even larger for estimates of non-use values may be considered unimportant.

The critical question, then, is knowing the likely extent of any errors, so it can be determined whether the error could affect a policy decision. The question of validity could, perhaps, be phrased somewhat differently. Perhaps the question should be ‘When are benefit estimates likely to be different, and, if so, by how much?’ The answer from the case studies considered in this thesis is that benefit estimates are likely to be different for transfers across either sites or populations, but that the differences are likely to be smaller in the former case, ranging from 4% to 66%. Whether these differences are acceptable to decision makers will depend on the issue being considered and the magnitude of costs and benefits. However, it should be noted that in other contexts the errors may be larger or smaller than what was found in this thesis. Further research is needed to determine when benefit transfer is, more generally, likely to be defensible.

11.3 Caveats

The results of any research project are subject to caveats given the constraints faced and the assumptions made. In this sections an attempt is made to acknowledge the main caveats for the research results presented in this thesis.

1. The Policy Context

As discussed in Section 4.2.1, the policy context for a valuation exercise can affect the magnitude of the benefit estimates. For example, if
respondents are facing the prospect of paying towards improving the quality of two separate wetlands, their willingness to pay for one of the wetlands may be less than if this wetland was the only one on the policy agenda. This issue is known as regular embedding. Its main implication, as derived from the contingent valuation literature, is that the current policy context should be used for any valuation exercise. In other words, if the government is considering improving the quality of multiple natural resources, then all those natural resources should be valued simultaneously in the questionnaire.

Therein lies the first caveat for this thesis. Separate surveys were conducted for both the Macquarie Marshes and the Gwydir Wetlands; they were not conducted simultaneously. The reason for this was that one of the main objectives of the study was to test the validity of benefit transfer, and thus separate surveys were needed. However, it should be recognised that if the government was considering improving the quality of both wetlands simultaneously then the benefit estimates presented in this thesis are likely to be overestimates.

2. Heterogeneity

A second caveat relates to the heterogeneity of preferences. The benefit transfer tests showed that there were significant differences between the preferences of respondents in Moree and Sydney for improvements in the quality of the Gwydir Wetlands. Hence the null hypothesis of convergent validity was rejected.

Respondents' preferences were shown to be a function of socioeconomic characteristics, such as income, and several attitudinal variables. However, further research might demonstrate the existence of more robust indicators of peoples' preferences. The inclusion of these variables might improve the accuracy of benefit transfer across different populations.
3. *Correlation between Choice Sets*

In Section 3.3 it was noted that the responses of each individual to multiple choice set questions are generally assumed to be independent in most CM applications. This assumption was used in this thesis. However, it is possible that responses will be correlated. While this has been shown not to cause parameter estimates to be biased, it can affect their efficiency (Adamowicz, Swait and Louviere 1997). As no attempt has been made to allow for correlated responses in this thesis, inaccurate estimates of standard errors may have been reported.

4. *Predictive Validity*

One of the objectives of this thesis was to test the validity of using CM to estimate non-use values. Validity was tested using two criteria: theoretical and content validity. While these two measures provide indications of the presence and effect of bias on value estimates, predictive validity provides a stronger test. However, as described in Chapter 7, the capacity of researchers to test predictive validity is limited. Nonetheless, it is possible to some extent to test predictive validity through the use of field experiments. For example, Blamey, Bennett, Morrison and Louviere (1998) report comparisons of actual and stated market share in the case of green marketable products. Further research in this area appears warranted to increase understanding of the validity of using CM to estimate non-use values.

5. *Strategic Behaviour*

The fifth caveat concerns the role of strategic behaviour in CM applications. Strategic behaviour has been a longstanding concern in the CVM literature (see Morrison et al 1996). A finding in the CVM literature
is that different elicitation questions and decision rules have different incentive compatibility properties; that is, different incentives for strategic behaviour. In CM applications it is possible to modify both the form of questioning, and the decision rule used to minimise the possibility of strategic bias. For example, a binary choice in the form of a referendum could be presented to respondents, similar to what might be done in a CVM application. Some preliminary and unpublished qualitative research on this topic has been presented by Carson, Groves and Machina (1997). However, there is no available quantitative evidence about the effect of different decision rules and choice set designs on value estimates in the context of CM.

As discussed in Chapter 6, no explicit social decision making structure or rule was described in the questionnaires used in this thesis, other than the increase in water rates applying to all households. It is possible that this may have affected the benefit estimates presented in this thesis.

6. Non-response Bias

Achieving low non-response rates is important for ensuring the reliability of survey results (Arrow et al 1993). Unless low non-response rates are achieved, extrapolation of survey results across the population may be subject to error. The response rates achieved in the surveys presented in this thesis ranged from approximately 45% to 50%, which are comparable to other stated preference surveys (eg Bennett et al 1997). However, there is still potential for non-response bias that may affect the validity of the results.

Arrow et al (1993) comment that in very high quality surveys non-response rates of below 20% are achieved. While this is a formidable target from a survey research perspective, it shows that there is opportunity for improvement in response rates. One possibility might be to send
reminder cards to households who have not returned their questionnaire a week after the survey was conducted.

7. IIA violations

A final caveat relates to the technique used to minimise IIA violations. As discussed in Chapter 4, there are several methods available to researchers to minimise IIA violations. The approach used in this thesis was to include socioeconomic and attitudinal interactions in the MNL models to allow for random taste variations. The advantage of this approach is that it is straightforward. It was possible to estimate similar models for each of the surveys. This contrasts with using nested logit models, which could only be estimated for the Sydney data sets, as discussed in Appendix 5.

The results presented in this thesis regarding the validity of using CM to estimate non-use values and for benefit transfer are based on this modelling approach. It is possible that different results may have been generated if alternative modelling procedures were used. The effect of different modelling procedures on the above mentioned areas is recommended as a topic of further research.
Bibliography


Louviere (pers.comm.). Professor Jordan Louviere, Department of Marketing, University of Sydney.


298


*Psychological Review*, 79: 281-299.

Company, New York.

Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-
offs for Chronic Bronchitis. *Journal of Environmental Economics and 

Research*, 28(3): 707-713.

14-29.

Mode Choice Models. *Regional Science and Urban Economics*, 5: 227-
249.

Weisbrod, B.A. (1967). Collective-Consumption Services of Individual-

Non-response and Sample Selection Bias in Contingent Valuation. 

of Community Preferences for Water Quality-Cost Alternatives. *Decision 
Sciences*, 5:614-631.


Appendix 1  The Macquarie Marshes draft questionnaire (excluding choice sets and debrief questions)
Management of the Macquarie Marshes in North-Western New South Wales

A Community Survey 1996

Map: The Macquarie Marshes

Funded by:
Land and Water Resources Research and Development Corporation
Your views on the management of the Macquarie Marshes

An environmental issue facing people in New South Wales is the deteriorating size and quality of the Macquarie Marshes. This deterioration has occurred because the amount of water in the Macquarie River that flows downstream into the Macquarie Marshes has fallen substantially. Instead, most of the water in the Macquarie River is now collected in Burrendong Dam and is used to irrigate crops, such as cotton and wheat.

It has been proposed to reduce the amount of water used for irrigation and increase the amount of water reaching the Macquarie Marshes by setting up a scheme to purchase water from farmers.

We would like to ask you a number of questions about these proposed changes.

Where is the Macquarie Marshes?

The Macquarie Marshes is located in the Macquarie Valley and are about 500 km north-west of Sydney (see Map).

What is the Macquarie Marshes?

The Macquarie Marshes is the largest wetland in NSW. It contains the largest area of common reeds in NSW and the largest area of river redgums in northern NSW.

The Macquarie Marshes is well known for its waterbirds. It is the largest breeding area in NSW for waterbirds such as ibis, egrets, spoonbills and herons, which only breed in large groups or ‘colonies’. These colonial waterbirds use few other wetlands in NSW for breeding. The Macquarie Marshes is used by 15 birds that are listed as endangered.

The Macquarie Marshes acts as a filter that improves water quality in the Macquarie River. It does this by slowing down the flow of water which causes sediment to settle out, and by trapping other impurities.

Within the Macquarie Marshes is a Nature Reserve covering approximately 18,000 ha which is administered by the National Parks and Wildlife Service. The Nature Reserve is listed as a wetland of international importance.

Parts of the Macquarie Marshes have been used for grazing sheep and cattle since the 1830s. The sheep and cattle primarily graze on ‘water couch’, which is a high quality feed. Grazing is not permitted within the Nature Reserve.

Effect of reduced water on the Macquarie Marshes

The Macquarie Marshes require regular periods of flooding to maintain wetland quality. The number and size of floods that occur in the Marshes have fallen because water is held upstream in Burrendong Dam for use in irrigating crops.

The best available scientific evidence suggests that reduced flooding has had the following effects on the Macquarie Marshes since irrigation began:

<table>
<thead>
<tr>
<th>Wetland area</th>
<th>The total area of the Macquarie Marshes during flood periods has fallen from more than 250,000 ha to 130,000 ha.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation health</td>
<td>The health of the vegetation in the Macquarie Marshes has declined. The long term health of vegetation is indicated by the area of healthy river redgums. It is estimated that the area of healthy river redgums has declined from 55,000 ha to 30,000 ha.</td>
</tr>
</tbody>
</table>

The area of water couch has also decreased and been replaced by weeds. This has affected the earnings of graziers operating within the Marshes.
What environmental issues are important?

Before considering what should be done about the Macquarie Marshes, we would first like you to think about a range of environmental issues.

Please rate the importance of improving wetland quality at the Macquarie Marshes relative to the importance of each of the following issues.

(eg circle 5 if you think that improving wetland quality at the Macquarie Marshes is much more important that improving water quality at Sydney beaches)

<table>
<thead>
<tr>
<th></th>
<th>Much less important</th>
<th>Slightly less important</th>
<th>Same Importance</th>
<th>Slightly more important</th>
<th>Much more important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving water quality at Sydney beaches</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Controlling land degradation in farmland in NSW</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Preventing logging of native forests in NSW</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Improving air quality in Sydney</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

The Proposed Water Purchasing Scheme

Extra water is needed to improve wetland quality in the Macquarie Marshes. Yet reducing the amount of water used for irrigation would be costly for farmers who have purchased irrigation equipment because they expected this water.

Currently, irrigators have an unreliable supply of water and only receive the water that they are entitled to in six out of every ten years. Increasing the amount of water going to the Macquarie Marshes would reduce the reliability of their water supply.

Irrigators have already had their water supply reduced. In 1995 the NSW Government reduced the average amount of water available for irrigation in the Macquarie Valley by 50,000 ML or 8%. Another reduction in water for irrigation is expected to put some irrigators out of business.

A possible way of increasing the amount of water reaching the Marshes while compensating farmers for receiving less water, is for the Government to set up a scheme to purchase water from farmers. Water is currently bought and sold in the Macquarie Valley in a water trading market. The Government could buy water from farmers in this market. The amount of water purchased would depend on how much the community wants to improve wetland quality at the Macquarie Marshes.

How would the Water Purchasing Scheme be funded?

The purchasing scheme is expected to be costly and the NSW Government does not have the necessary money available to pay for the scheme from existing taxation revenue.

If the purchasing scheme was to go ahead it would be necessary to charge households in NSW a once-off levy on their water rates to pay for the scheme. The revenue would go into a special fund that would only be used to pay for the scheme.

Options for the Macquarie Marshes

To help determine how the community would like to see the Macquarie Marshes managed we have prepared 16 sets of management options for the Macquarie Marshes. We would like to know which management option you prefer the most in each set of options. You'll find some of the options more attractive than others. You
may choose to give up attractive features in one option (eg larger area), to get some other features in a second option (eg more waterbirds breeding).

We are asking your opinion about these options to help the government determine which characteristics or features of the Macquarie Marshes matter to people.

To keep matters simple, we do not describe how each management option would come about. For this reason, some options appear a little unusual, or even unrealistic. Bear in mind that there are a lot of ways that water can be managed, so that options which seem odd are actually quite possible.

When thinking about which option you prefer, keep in mind your available income and all the other things that you have to spend money on. It is possible that in the future other environmental projects may cost you additional money.

EXAMPLE

Here is an example of how to answer questions 1 to 16.

Suppose options A and B are the only available options. If you preferred option A, you would tick the box labelled ‘I would choose A’ as shown below:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in wetland area</td>
<td>40,000 ha</td>
<td>80,000 ha</td>
</tr>
<tr>
<td>Increase in area of healthy river redgums</td>
<td>5000 ha</td>
<td>10,000 ha</td>
</tr>
<tr>
<td>Increase in number of waterbirds</td>
<td>80,000</td>
<td>160,000</td>
</tr>
<tr>
<td>Increase in waterbirds breeding</td>
<td>20,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Increase in waterbird species</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Water quality</td>
<td>Good</td>
<td>Fair</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
</tr>
</tbody>
</table>

I would choose A [✓] I would choose B

I would choose neither A or B
Appendix 2 The Gwydir Wetlands balanced choice set questionnaire (excluding choice sets and debrief questions)
Management of the Gwydir Wetlands in Northern New South Wales

A Community Survey 1997

How To Fill Out This Questionnaire

The questions take different forms. In most cases you only need to tick the box which is closest to your view. Here is an example:

EXAMPLE

Do you think the government should spend more or less on education?

Spend more on education

Spend less on education

Don't know

If you think the government should spend more on education, you would tick the box as shown.

Sometimes you need to write an answer—in these cases, simply write your answer in the space provided. If you have any questions regarding this survey, please contact Professor Jeff Bennett of The University of New South Wales on 06 268 8833.

Completion of this survey is voluntary

All your answers will be kept strictly confidential

Your completed questionnaire will be collected by a surveyor at an agreed time. If you are unable to give the questionnaire to the surveyor at this time, please use the reply paid envelope to mail it to:

Professor Jeff Bennett
School of Economics and Management
University College
The University of New South Wales
Canberra ACT 2600

We hope you enjoy completing this questionnaire and thank you very much for taking part in this survey.
Your views on the management of the Gwydir Wetlands

An issue facing people in NSW is the reduction in the size and quality of the Gwydir Wetlands. This has occurred because the amount of water that flows into the Gwydir Wetlands has fallen substantially. Instead, much of the water in the Gwydir River is now collected in Copeton Dam and used to irrigate crops.

There are several options available to the government in deciding how much water to allocate for irrigation and to the Gwydir Wetlands. We would like to ask you a number of questions about these options.

What issues are important?

Question 1

Before considering what should be done about the Gwydir Wetlands, we would first like you to think about a range of issues that Australians face today. Below is a list of areas where the government has been considering reductions in funding. Please rank these five areas by placing the numbers 1 (most deserving of continued funding) to 5 (least deserving) in the following boxes:

- Education
- The Environment
- Crime Prevention
- Hospitals
- Unemployment

Question 2

One of the areas just mentioned was ‘The Environment’. Please rank the following five environmental goals by placing the numbers 1 (most deserving of continued government funding) to 5 (least deserving) in the following boxes:

- Improving water quality in rivers in NSW
- Conserving wetlands in NSW
- Improving land quality in rural areas in NSW
- Preventing logging of native forests in NSW
- Improving air quality in urban centres in NSW

THE GWYDIR WETLANDS

Where are the Gwydir Wetlands?

The Gwydir Wetlands are in the Gwydir Valley, about 600 kilometres north of Sydney and about 70 kilometres west of Moree (see map).

What are the Gwydir Wetlands?

Before water from the Gwydir River was diverted for irrigation, the Gwydir Wetlands were the third largest wetlands in NSW. They had an area of about 2000 square kilometres. This included 200 square kilometres containing reeds, rushes and other aquatic plants and 1800 square kilometres of woodlands containing acacias, coolabahs and lignum bushes.

The Gwydir Wetlands were an important habitat for birds. Before irrigation, some 225 bird species were found in the area, with 125 species breeding there. The wetlands are one of the largest breeding areas in NSW for waterbirds such as ibis, egrets, spoonbills and herons, species which breed only in large groups. These waterbirds breed in only a few other wetlands in NSW. Before irrigation the number of waterbirds in a single breeding event stimulated by flooding ranged from 5000 pairs to well over 200,000 pairs. Many of these waterbirds are found throughout NSW, including urban areas.

Before irrigation, the Gwydir Wetlands were used as habitat by 19 bird species listed as endangered by the National Parks and Wildlife Service and 18 migratory waterbirds protected under international agreements.

All of the Gwydir Wetlands are owned by private landholders. Several whole properties have been declared wildlife sanctuaries.

The Gwydir Wetlands have been used for grazing sheep and cattle since the 1800s. They have been an important drought refuge for cattle from throughout NSW and parts of Victoria and Queensland. About 260,000 sheep and cattle were held on properties surrounding the Gwydir Wetlands in 1991.
How has less water affected the Gwydir Wetlands?

The removal of water for irrigation has altered the natural pattern of flooding. This has had a different effect on each aspect of the Gwydir Wetlands. Wetland vegetation requires regular floods of various sizes to remain healthy. Waterbird breeding requires flooding of sufficient size at a suitable time of year. Endangered species require the preservation of certain habitats through regular flooding. These different flooding needs mean that water managers must choose which aspect of the Wetlands to preserve if extra water is available.

**Wetland area**
The area of the Gwydir Wetlands had fallen from 2000 to 400 square kilometres by 1996 (400 square kilometres is about equal to the area of a large country town such as Dubbo or Goulburn). The area of weeds—such as lippia, roly-poly and scotch thistle—had also increased significantly.

A consequence of the reduction in area is that the amount of stock feed has fallen by 50%. This has reduced the income of graziers.

**Waterbird breeding**
Waterbird breeding used to occur every second year but now occurs every five years.

**Endangered and protected species**
The number of endangered bird species using the Gwydir Wetlands had fallen from 19 to 5 species by 1996. The number of visiting migratory species protected under international agreements has fallen from 18 to 7.

---

**IRRIGATION IN THE GWYDIR VALLEY**

**Production**
Irrigation in the Gwydir Valley began after the construction of Copeton Dam in 1976. The most widespread irrigated crop is cotton. The Gwydir Valley is the largest cotton producing area in Australia, with 500-900 square kilometres harvested each year. The estimated annual revenue from cotton production between 1989 and 1993 was $110 million. Other irrigated crops include wheat, soybeans, sorghum, vegetables and pecan nuts. The total annual revenue from these other crops is $25 million. Irrigated agriculture makes up 58% of total agricultural production in the Gwydir Valley.

**Employment**
About 1600 people are employed on irrigated farms in the Gwydir Valley. Two towns in the valley—Moree and Mungindi—also rely on the revenue from irrigated agricultural production. It is estimated that irrigation creates about 1200 jobs in these towns.

**Water use**
Most of the water in Copeton Dam is used for irrigation. About 517,000 megalitres is allocated for irrigation in the Gwydir Valley each year. Farmers pay about $5 per megalitre for the water they use each year.

Farmers received licenses to irrigate by applying to the Department of Water Resources. These licenses specify an amount of water that will be allocated to farmers for irrigation, provided that there is enough water available. In five out of every ten years there is enough water available for farmers to receive all of the water specified in their licenses. In the remaining years they receive less than 100% of their allocation. However, this is supplemented by water they collect from tributaries to the Gwydir River and keep in on-farm storages.

**Water trading**
Water is currently bought and sold in a water trading market which covers the entire Gwydir Valley. Farmers can buy or sell the rights to receive water to other farmers permanently. The price of water rights in the Gwydir Valley has been about $600 to $700 per megalitre in recent years.

**OPTIONS FOR MANAGING THE GWYDIR WETLANDS**
The three broad options for managing the Gwydir Wetlands are discussed below.

**Option 1: Continue current situation**
Under this option the amount of water allocated for both irrigation and the Gwydir Wetlands would remain constant. Wetland size and quality would also remain constant.
Option 2: Increase water allocated to the Gwydir Wetlands

This option would increase wetlands size and improve wetlands quality. One way of increasing water for the Gwydir Wetlands is for the NSW Government to purchase water from farmers using the existing water trading market. This would be voluntary.

Farmers would be expected to sell water if the price offered was high enough. The money received from selling their water would enable farmers to purchase more efficient irrigation equipment that would reduce their water needs. It would also compensate other farmers who decide to stop irrigating and return to dryland farming. The sale of water would cause a decline in employment on farms and in Moree and Mungindi.

Purchasing water from farmers would be costly and the NSW Government does not have money to pay for the scheme from existing taxation revenue. If the purchasing scheme were to go ahead it would be necessary to charge households in NSW a one-off levy on water rates, payable only in 1998. The revenue would go into a special fund that would only be used to pay for the scheme.

Option 3: Reduce water allocated to the Gwydir Wetlands

This option would further reduce the quality of the Gwydir Wetlands, but would mean that there would be more water available for irrigation. As a result the income of farmers and the number of jobs would increase.

This extra water could be allocated to farmers by selling it to them on the water trading market. This would provide extra revenue for the NSW Government which could be used to fund a once-off rebate on water rates. This would be paid to all households in NSW in 1998.

WHICH OPTION DO YOU PREFER?

To help find out how the community would like the Gwydir Wetlands managed, we have prepared eight sets of potential management options. These options are defined in terms of five different outcomes: wetland area, waterbirds breeding, endangered and protected species using the wetland as habitat, employment, and household cost or benefit. We would like to know which management option you prefer in each set of options.

The outcomes in each of the options have been specifically defined so that you have a broad range of choices. Within this range some options may seem strange according to your experience, but bear in mind that there are many ways of managing water.

When thinking about which option you prefer, keep in mind your available income and all the other things that you spend money on. It is possible that in the future other environmental projects may cost you additional money.

EXAMPLE

Here is an example of how to answer questions 3 to 10. Suppose Options 1, 2 and 3 are the only available management options for the Gwydir Valley. If you preferred Option 1, you would tick the box under Option 1 as shown below:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Reduce water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water rates</td>
<td>no change</td>
<td>$30 increase</td>
<td>$30 rebate</td>
</tr>
<tr>
<td>Employment</td>
<td>2800</td>
<td>2780</td>
<td>2880</td>
</tr>
<tr>
<td>Wetland area</td>
<td>400 km²</td>
<td>800 km²</td>
<td>100 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 9 years</td>
</tr>
<tr>
<td>Endangered and protected species</td>
<td>12</td>
<td>20</td>
<td>4</td>
</tr>
</tbody>
</table>

Choice (tick a box) [ ] [ ] [ ]

317
Appendix 3  Questionnaires used in the Macquarie Marshes and Gwydir Wetlands surveys
See inside the back cover of the thesis for copies of the questionnaires
Appendix 4: Alternative choice set formats

A number of different choice set formats were trialed in both the Macquarie Marshes and Gwydir Wetlands focus groups. The reactions of respondents to these different formats are detailed in this appendix.

A4.1 Alternative Choice Set Formats
Trialed in the Macquarie Marshes Focus Groups

The choice set format used in the draft Macquarie Marshes questionnaire is presented in Table A4.1. The attributes were expressed as increments to the current situation in this choice set format.

Table A4.1: Draft Macquarie Marshes questionnaire choice set format

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in wetland area</td>
<td>40,000 ha</td>
<td>80,000 ha</td>
</tr>
<tr>
<td>Increase in area of healthy river redgums</td>
<td>5000 ha</td>
<td>10,000 ha</td>
</tr>
<tr>
<td>Increase in number of waterbirds</td>
<td>80,000</td>
<td>160,000</td>
</tr>
<tr>
<td>Increase in waterbirds breeding</td>
<td>20,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Increase in waterbird species</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
</tr>
</tbody>
</table>

I would choose A ☐  I would choose B ☐
I would choose neither A or B ☐

This format was used because it was thought that choice sets might be simpler if the alternatives were expressed in increments and participants...
only had to read two and not three alternatives. However, this hypothesis needed to be tested, so several other formats were trialed at the end of each focus group. The first alternative involved fewer attributes (see Table A4.2). This alternative was well received, as shown by the following comments:

You’ve got less to worry about but you’re also learning less. (D1)

1 Just easier on the eye.
2 This covers basically what you need to know—it’s nice. (D2)

Table A4.2: A choice set with fewer attributes

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in wetland area</td>
<td>40,000 ha</td>
<td>80,000 ha</td>
</tr>
<tr>
<td>Increase in area of healthy river redgums</td>
<td>5000 ha</td>
<td>10,000 ha</td>
</tr>
<tr>
<td>Increase in waterbirds breeding</td>
<td>20,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
</tr>
</tbody>
</table>

I would choose A  [ ]  I would choose B  [ ]  I would choose neither A or B  [ ]

Participants were then shown an alternative in which the current situation was specified as ‘Option C’ (see Table A4.3). There was strong support for this alternative:

Option C on that one is good because it gives you a basis. You know what’s what now. (S1)

Its giving us a starting point. We know where we are now, here [the original] you don’t know where you’re coming from. (S2)

Yes I think you need to know what’s actually happening there now...I mean if you’ve got no idea how big it is now... (D1)
<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
<th>Option C: Continue Current Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wetland area</td>
<td>170,000 ha</td>
<td>210,000 ha</td>
<td>130,000 ha</td>
</tr>
<tr>
<td>Area of healthy river redgums</td>
<td>35,000 ha</td>
<td>40,000 ha</td>
<td>30,000 ha</td>
</tr>
<tr>
<td>Number of waterbirds</td>
<td>280,000</td>
<td>360,000</td>
<td>200,000</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>40,000</td>
<td>60,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Waterbird species</td>
<td>40</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
<td>$0</td>
</tr>
</tbody>
</table>

Participants were then asked whether they would prefer a combination of these two alternatives, where there were fewer alternatives and a constant base. There was agreement that this was the preferred alternative in all focus groups.

Participants in the second focus group in Dubbo were also shown a dissonance-minimising (DM) format, similar to that trialed in Bennett et al (1997) (see Table A4.4). The DM format was well received. It was not possible to try this format in the other focus groups because of shortage of time.
### Table A4.4: Dissonance-minimising choice set format

<table>
<thead>
<tr>
<th>Feature</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in wetland area</td>
<td>40,000 ha</td>
<td>80,000 ha</td>
</tr>
<tr>
<td>Increase in area of healthy river redgums</td>
<td>5000 ha</td>
<td>10,000 ha</td>
</tr>
<tr>
<td>Increase in number of waterbirds</td>
<td>80,000</td>
<td>160,000</td>
</tr>
<tr>
<td>Increase in waterbirds breeding</td>
<td>20,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Increase in waterbird species</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Water quality</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td>Increase in water rates</td>
<td>$30</td>
<td>$60</td>
</tr>
</tbody>
</table>

I would choose A

I would choose B

I support the options, but neither is worth this much money to me

I support the options, but I can’t afford this much extra money

I support the options, but not if it requires a levy of any amount

I oppose the scheme, regardless of cost

### A4.2 Alternative Choice Set Formats

*Trialed in the Gwydir Wetlands Focus Groups*

- Balanced and non-balanced formats

In the Gwydir Wetlands focus groups both balanced and non-balanced questionnaire formats were trialed (see Table A4.5 and A4.6). In contrast to the Macquarie Marshes questionnaire, a constant base was included, all of the attributes (except household cost) were expressed as absolute values rather than as increments, and labels were used in the non-balanced format. Respondents have the option of choosing reduced water to the wetlands as well as increased water to the wetlands in the balanced format.
Each of these alternatives appear to have advantages. The balanced format is potentially less biased, as it provides respondents with the option of choosing increased as well as reduced water to the wetlands. However, it was thought that an advantage of the non-balanced format is that respondents may be more discriminating about what environmental improvements they were actually receiving. In other words, a possible
weakness of the balanced format is that respondents may simply choose to increase water to the wetlands without looking carefully at the attributes.

To trial these two formats, participants in each of the focus groups were asked to complete both surveys. Participants in the first focus groups in Sydney and Moree initially received the balanced version, while participants in the second groups initially received the non-balanced version. As expected, some participants found the balanced version broader and less biased:

I preferred that option...leave it the same, increase it, decrease it. Even if you're not going to go anywhere near that option [option 3]. (M1)

To keep the survey accurate you need a point where somebody can say I want more money to go into irrigation. If you're going to get the real cross section of people...you need a box where they can tick that. (M1)

Less biased than the first one (M2)

There was, however, scepticism about the plausibility of the rebate, although some participants thought that the rebate was plausible:

Q: Did you find the rebate believable?
[Several nos]
I thought that it was amusing. (S3)

Q: What about the rebate, did you believe that?
1 Its a possibility
2 Probably just as likely as option 2
3 Its all possible, its just a matter of whether or not they're going to do it or not (M1)

This one offering the $30 rebate. Are we going from the sublime to the ridiculous? Are we saying that people not care about the environment that much...? (M2)

One participant voiced concern about using the balanced version, as he thought that many people might choose the rebate. However, this may be
an exaggerated fear. Out of the 36 people who completed the questionnaire in the focus groups, only one person chose option 3, and in only one choice set.

...including a rebate might be a bit of a risk to the people who are looking to get the wetlands supported financially because many ratepayers would be happy to take a rebate anytime. (S4)

Especially in the first Sydney focus group, participants appeared to be more discriminating under the non-balanced format, as shown by the following comments. Participants appeared to be less likely to focus solely on labels such as ‘Increase water to Gwydir Wetlands’ or ‘Reduce water to Gwydir Wetlands’.

1 It's to nut out what we feel is important... what we feel is a major priority
2 It's a good idea. This one has a few more options. I think it really puts you on the spot to think do you really want to pay your $100, whereas before they we're so easy. $30 was nothing to anyone...
3 I agree... it did really focus your attention on what you were getting. (S3)

Q: Who preferred this one to the other one?
[Lots of yeses]
1 More confronting
2 This one really got to the issue better. (S3)

In summary, it appears that participants thought that the balanced version was less biased although some respondents doubted the plausibility of the rebate. The propensity of some participants to focus on labels rather than attributes under the balanced format (and even to an extent the non-balanced format) is also a concern, although the extent to which participants do this is unclear.

- Combined balanced and non-balanced format

After the first two Sydney focus groups, when the advantages and limitations of the balanced format had become apparent, another format
was developed and trialed in the Moree focus groups. This format is a hybrid of the balanced and non-balanced formats (see Table A4.7). The format is similar to the non-balanced format except that a respondent can still choose to increase water for irrigation and reduce wetland quality. It was expected that this format would be seen as less biased than the original non-balanced format.

Table A4.7: Combined balanced and non-balanced format

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water rates</td>
<td>no change</td>
<td>$100 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Wetland area</td>
<td>400 km(^2)</td>
<td>800 km(^2)</td>
<td>500 km(^2)</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 4 years</td>
</tr>
<tr>
<td>Endangered and protected species</td>
<td>12</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

I would choose option 1, to continue the current situation

I would choose option 2

I would choose option 3

I wouldn't choose any of these options but would prefer more water to be allocated for irrigation and less to the wetlands

Participants in the Moree focus groups voiced moderate support for the use of this format:

1 ...if you're going to put that fourth box in there people might like to suggest ways. I mean they might be totally unachievable but...as it is there it doesn't mean anything. (M1)

1 I think that it helps a bit

2 It helps...(M2)
Inclusion of an employment attribute

A conclusion from the Macquarie Marshes focus groups was that it would be appropriate to include an employment attribute to show the effects of the reallocation on employment on farms and in country towns. However, it is not clear that the purchase of water from farmers would necessarily result in a decline in employment, especially if the reallocation resulted in the installation of more labour intensive irrigation equipment and an increase in employment by graziers. Because of this uncertainty, an employment attribute was only included in the questionnaires given to participants in the first Sydney and first Moree focus groups.

None of the participants in either the second Sydney or Moree focus groups suggested that an employment attribute should be included in the choice sets. Participants were also asked whether they thought that the purchase of water would result in increased or decreased employment. Most participants were uncertain about the effect:

1 Difficult question to answer
2 You’re going to get more people employed on equipment, but can’t say more or less. (S4)

...if the farmers get access to money they will try and save labour through better machinery
2 Not sure
3 Could be more jobs
4 Might be short term increase and long term decrease
5 If the farmer had more money he’d look for productivity which is bound up with efficiency...how he interprets that is individual. (M2)

These responses suggest that it is not necessary to include an employment attribute. However, there is still a rationale for including an employment attribute. It may serve to reduce bias as well as moderating lexicographic environmental preferences.
• Other choice set formats

Towards the end of the first three focus groups, participants were given a sheet that contained four other methods of presenting choice sets. The first of these formats is identical to the original non-balanced format except that participants were asked to explain their choice. The reaction to this format was mixed:

1 I think everybody's answered it
2 Its too hard
3 You've only got one line there [not enough room]
4 I would tend to keep it the way it is...let the back page cover it. (S3)

1 Rather than writing 'Why did you choose this option' under alternative 1, people sometimes like a comments section so they can say what they couldn't get across in answering the question. (S4)
2 Or you could have a number of questions for people to answer about why they chose a particular option. (S4)

1 I think its good
2 Makes people think about it
Q: Would you like a question or just ask for any comments?
1 No ask a question, if people don't want to answer it they wont, whereas if its just comments they'll skip over it right away
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water rates</td>
<td>no change</td>
<td>$100 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Wetland area</td>
<td>400 km²</td>
<td>800 km²</td>
<td>500 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 4 years</td>
</tr>
<tr>
<td>Endangered and protected species</td>
<td>12</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

Why did you choose this option?

In the second alternative the boxes which participants would tick to indicate their choice were written differently and had a statement next to each (see Table A4.9). This format did not receive much support:

1 Much of a muchness
2 Same dog different legs (S3)

Top one [Alternative 1] is easier (S4)
### Table A4.9: Alternative 2

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: Continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water rates</td>
<td>no change</td>
<td>$100 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Wetland area</td>
<td>400 km²</td>
<td>800 km²</td>
<td>500 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 4 years</td>
</tr>
<tr>
<td>Endangered and protected species</td>
<td>12</td>
<td>20</td>
<td>14</td>
</tr>
</tbody>
</table>

The options were listed in the first column and the attributes were listed in the first row in the third alternative (see Table A4.10). This allows respondents to read an option from left to right, rather than down a column.

### Table A4.10: Alternative 3

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Water rates</th>
<th>Wetland area</th>
<th>Waterbirds breeding</th>
<th>Endangered and protected species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option 1: Continue current situation</td>
<td>no change</td>
<td>400 km²</td>
<td>every 5 years</td>
<td>12</td>
</tr>
<tr>
<td>Option 2: Increase water to Gwydir Wetlands</td>
<td>$100 increase</td>
<td>800 km²</td>
<td>every 2 years</td>
<td>20</td>
</tr>
<tr>
<td>Option 3: Increase water to Gwydir Wetlands</td>
<td>$30 increase</td>
<td>500 km²</td>
<td>every 4 years</td>
<td>14</td>
</tr>
</tbody>
</table>

The options were listed in the first column and the attributes were listed in the first row in the third alternative (see Table A4.10). This allows respondents to read an option from left to right, rather than down a column.
Alternative 4 was identical to Alternative 3, except that participants ticked a box after each option. Both Alternative 3 and 4 received some support. Further testing would, however, be required to determine whether any of these alternatives are improvements over the choice set format used in the focus groups. The format for Alternatives 3 and 4 could be difficult to use with more than four attributes.

1 I think its better because most people read left to right rather than up and down.
2 three or four are the best (S3)

[Alternative 4] Its the best (S3)

1 [Alternative 4] Its like a ballot box
2 Much of a muchness (S4)

Table A4.11: Alternative 4

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Water rates</th>
<th>Wetland area</th>
<th>Waterbirds breeding</th>
<th>Endangered and protected species</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Option 1</strong>: Continue current situation</td>
<td>no change</td>
<td>400 km²</td>
<td>every 5 years</td>
<td>12</td>
</tr>
<tr>
<td><strong>Option 2</strong>: Increase water to Gwydir Wetlands</td>
<td>$100 increase</td>
<td>800 km²</td>
<td>every 2 years</td>
<td>20</td>
</tr>
<tr>
<td><strong>Option 3</strong>: Increase water to Gwydir Wetlands</td>
<td>$30 increase</td>
<td>500 km²</td>
<td>every 4 years</td>
<td>14</td>
</tr>
</tbody>
</table>
Appendix 5: Further Modelling Results

This Appendix contains further analysis of the survey results that were referred to in Chapter 7. Mother logit tests of the IIA property are reported in Section A5.1. MNL models that show the effect of deleting dominated alternatives from the experimental design are reported in Section A5.2. Section A5.3 contains details of several other models which were estimated when attempting to deal with IIA violations.

A5.1 Tests for IIA Violations

A5.1.1 MNL models with choice set attributes

Three mother logit models were estimated to test whether the basic MNL models that only contained choice set attributes suffered IIA violations (see Table A5.1). Each of the models include generic cross effects between alternatives 2 and 3, and 3 and 2, similar to Elrod, Louviere and Davey (1987)\(^9\). As detailed below, these cross-effects show how the attributes for alternative 2 affect the utility of alternative 3, and vice versa:

\(^9\) Including cross effects between alternatives 2 and 3, and 3 and 2 is equivalent to including cross effects between alternatives 1 and 3, and 1 and 2, except that the sign of the cross effects will be opposite. This is because the MNL model differences against the base option. The latter approach is used in the following models.
\[ V_1 = \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES} \]

\[ V_2 = C_2 + \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES} \\
+ \beta_6 \text{RATES}^2 + \beta_7 \text{JOBS}^2 + \beta_8 \text{AREA}^2 + \beta_9 \text{BREED}^2 + \beta_{10} \text{SPECIES}^2 \]

\[ V_3 = C_3 + \beta_1 \text{RATES} + \beta_2 \text{JOBS} + \beta_3 \text{AREA} + \beta_4 \text{BREED} + \beta_5 \text{SPECIES} \\
+ \beta_6 \text{RATES}^2 + \beta_7 \text{JOBS}^2 + \beta_8 \text{AREA}^2 + \beta_9 \text{BREED}^2 + \beta_{10} \text{SPECIES}^2 \]

where \( \text{RATES}^2, \text{JOBS}^2, \text{AREA}^2, \text{BREED}^2 \) and \( \text{SPECIES}^2 \) represent cross-effects between alternatives 2 and 3.

Violations of the IIA property are partly indicated by the existence of significant coefficients for the cross-effects. However, it is necessary to conduct a likelihood ratio test to correctly determine whether the IIA property has been violated. The null hypothesis for this test is that the MNL is the more accurate model. The alternative hypothesis is that the mother logit is the more accurate model. The test statistic is equal to \(-2 (L_R - L_U)\), where \( L_R \) and \( L_U \) are the log likelihood of the restricted and unrestricted models. This statistic is chi-squared distributed, with degrees of freedom equal to the number of restricted parameters.

There is only one significant cross-effect for rates in the first mother logit model, although the cross-effect is significant at the 1% level. Hence it is possible that there will be an IIA violation. The log likelihood is equal to \(-1152.616\), compared to a log-likelihood of \(-1159.135\) in the MNL model. The chi-squared value is therefore: \(-2(-1159.135 - -1152.616) = 13.038\). The critical value at the 5% significance level given five degrees of freedom is 11.07, and at the 1% level is 15.09. Hence we would reject the null hypothesis that the MNL is the more accurate model at the 5% level of significance, but not at the 1% level. This indicates that this model suffers from violations of the IIA property, but that it is not a severe violation.
There are three significant cross-effects in model 2, for RATES2, AREA2 and SPECIES2, therefore it is likely that there will be an IIA violation in this model. The test statistic is equal to: $-2 \times (\text{-1476.999} - \text{-1461.437}) = 31.124$, which is greater than the critical value at the 1% level of 15.086. Hence it appears that there are highly significant violations of the IIA property in the Gwydir Sydney model.

There are also three significant cross-effects for RATES2, JOBS2 and SPECIES2 in model 3. Hence, IIA violations probably occur in the Macquarie Marshes model. The test statistic is equal to: $-2 \times (\text{-1531.307} - \text{-1519.802}) = 23.010$, which is also greater than the critical value at the 1% level. Therefore, there are highly significant violations of the IIA property in the Macquarie Marshes model. This violation occurred despite the Macquarie Marshes model appearing to be the most robust of the three models.
Table A5.1: Mother logit models with choice set attributes

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1: Gwydir Moree</th>
<th>Model 2: Gwydir Sydney</th>
<th>Model 3: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_2$</td>
<td>-4.298</td>
<td>4.605</td>
<td>7.041*</td>
</tr>
<tr>
<td>$C_3$</td>
<td>-3.836</td>
<td>4.974</td>
<td>7.363**</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.836E-2***</td>
<td>-0.118E-1***</td>
<td>-0.124E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.283E-2*</td>
<td>0.345E-2***</td>
<td>0.111E-2</td>
</tr>
<tr>
<td>AREA</td>
<td>-0.299E-5</td>
<td>-0.202E-3</td>
<td>0.426E-3***</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.035</td>
<td>-0.190**</td>
<td>-0.298***</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.038**</td>
<td>0.511E-2</td>
<td>0.032***</td>
</tr>
</tbody>
</table>

**Cross-Effects**
- RATES2 0.334E-2*** 0.350E-2*** 0.247E-2***
- JOBS2 0.146E-2 -0.945E-3 -0.157E-2*
- AREA2 -0.415E-3 -0.103E-2*** -0.245E-3
- BREED2 0.137 -0.037 0.075
- SPECIES2 -0.011 -0.054*** -0.032***

**Summary statistics**
- Log-likelihood -1152.616 -1461.437 -1519.802
- $\chi^2$ (constants only) 113.638 214.075 405.852
- $\rho^2$ adjusted 0.049 0.065 0.118
- Iterations completed 5 5 5
- Observations 1109 1429 1575

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level standard errors not reported

A5.1.2 Mother logit models with socioeconomic and attitudinal variables

The mother logit models estimated to test whether the IIA property holds for this model specification are reported in Table A5.2. For the Gwydir Moree and Macquarie Marshes model only one cross-effect is significant, suggesting that IIA violations have been reduced. However, for the Gwydir Sydney model three cross effects (RATES2 AREA2 and SPECIES2) are still significant, suggesting that the inclusion of socioeconomic and attitudinal variables has not been successful in dealing with IIA violations. Formal tests for each of these models are as follows.

The chi-squared statistic for model 4 is: $-2(-883.438 - -879.067) = 8.742$. The critical value at the 5% significance level is 11.07. Hence the null hypothesis that MNL is the correct model cannot be rejected. This
indicates that the inclusion of socioeconomic and attitudinal interactions has been successful in removing violations of the IIA property in the Gwydir Moree model.

The test statistic in model 5 is equal to: \(-2 (-1233.092 - -1225.332) = 15.52\). The critical value at the 1% level is 15.09. Therefore we reject the null hypothesis at the 1% level and conclude that the mother logit is the more accurate model and that there are severe violations of the IIA property in the Gwydir Sydney model.

In model 6, which is the Macquarie Marshes model, the test statistic is equal to: \(-2 (-1244.856 - -1239.589) = 10.534\). This is less than the critical value at the 5% level (11.071). Therefore we accept the null hypothesis that the MNL is the more accurate model and conclude that the IIA property has not been violated.
Table A5.2: Mother logit models with socioeconomic and attitudinal variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 4: Gwydir Moree</th>
<th>Model 5: Gwydir Sydney</th>
<th>Model 6: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>-5.779</td>
<td>5.228</td>
<td>13.202</td>
</tr>
<tr>
<td>C3</td>
<td>-5.351</td>
<td>0.0448E-1***</td>
<td>13.549</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.822***</td>
<td>-0.740***</td>
<td>0.380***</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.455***</td>
<td>0.614***</td>
<td>-0.124</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.239E-5</td>
<td>0.594E-5***</td>
<td>0.943E-5***</td>
</tr>
<tr>
<td>INCOME DUMMY</td>
<td>-0.543***</td>
<td>0.158</td>
<td>-0.238</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.462**</td>
<td>0.655***</td>
<td>0.866***</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.454***</td>
<td>-1.036***</td>
<td>1.132***</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-0.903***</td>
<td>0.694***</td>
<td>-0.881***</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.112E-1***</td>
<td>-0.147E-1***</td>
<td>-0.149E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.617E-3</td>
<td>0.0354E-2**</td>
<td>0.265E-2***</td>
</tr>
<tr>
<td>AREA</td>
<td>0.111E-5</td>
<td>0.798E-3**</td>
<td>0.634E-3***</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.171**</td>
<td>-0.144**</td>
<td>-0.364***</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.045***</td>
<td>0.054***</td>
<td>0.061***</td>
</tr>
<tr>
<td>Cross-Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RATES2</td>
<td>-0.289E-2**</td>
<td>-0.255E-2**</td>
<td>-0.137E-2</td>
</tr>
<tr>
<td>JOBS2</td>
<td>-0.101E-2</td>
<td>0.527E-3</td>
<td>0.158E-2</td>
</tr>
<tr>
<td>AREA2</td>
<td>0.525E-3</td>
<td>0.800E-3*</td>
<td>0.171E-3</td>
</tr>
<tr>
<td>BREED2</td>
<td>-0.979E-1</td>
<td>0.050</td>
<td>0.047</td>
</tr>
<tr>
<td>SPECIES2</td>
<td>0.021</td>
<td>0.045***</td>
<td>0.027**</td>
</tr>
<tr>
<td>Summary statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-879.067</td>
<td>-1225.322</td>
<td>-1239.589</td>
</tr>
<tr>
<td>(\chi^2) (constants only)</td>
<td>147.253</td>
<td>334.206</td>
<td>574.789</td>
</tr>
<tr>
<td>(p^2) adjusted</td>
<td>0.072</td>
<td>0.114</td>
<td>0.187</td>
</tr>
<tr>
<td>Iterations completed</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Observations</td>
<td>872</td>
<td>1269</td>
<td>1397</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level, standard errors not reported.
Similar to the previous model specifications, mother logit models are estimated to test whether the IIA property has been violated (see Table A5.3). There are fewer or less significant cross-effects in each of the Gwydir models, although, there is an extra significant cross-effect for JOBS in the Macquarie Marshes model. Hence, overall it appears that this model specification is likely to be less affected by IIA violations. This is demonstrated in the following tests.

The chi-squared statistic for model 7 is: \(-2(-649.347 - -645.507) = 7.68\). This is less than the critical value at the 5% significance level (11.071), hence the null hypothesis that the MNL is the correct model is accepted. Thus the Gwydir Moree model does not suffer from violations of the IIA property.

The test statistic in model 8 is equal to: \(-2 (-999.673 - -996.127) = 7.092\), which is also less than the critical value at the 5% level. Therefore the null hypothesis that the MNL model is the more accurate model is accepted. Hence, similar to the previous model specification, the Macquarie Marshes model does not suffer from IIA violations.

In the final model, model 9, the test statistic is equal to: \(-2 (-1037.286 - -1035.385) = 3.802\), which is less than the critical value at the 5% level. Hence, similar to the Gwydir Sydney model, we accept the null hypothesis and conclude that the MNL is the more accurate model.

---

92 However, the extent of IIA violations does not appear to be substantially different than for the previous model specification. The p-value for JOBS in the previous specification was 0.109 and is 0.086 for this specification.

93 It should be noted that the IIA violations were not removed simply because of the addition of extra variables. When a similar number of irrelevant socioeconomic variables were added the violation was still found to occur.
Table A5.3: Mother logit models with socioeconomic, attitudinal and questionnaire variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 7: Gwydir Moree</th>
<th>Model 8: Gwydir Sydney</th>
<th>Model 9: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_2$</td>
<td>-13.898</td>
<td>10.123</td>
<td>15.422</td>
</tr>
<tr>
<td>$C_3$</td>
<td>-13.964</td>
<td>10.342</td>
<td>15.767</td>
</tr>
<tr>
<td>CHILD</td>
<td>1.07***</td>
<td>-0.578***</td>
<td>0.406**</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.269</td>
<td>0.500**</td>
<td>-0.921E-2</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.338E-5</td>
<td>0.239E-5</td>
<td>0.733E-5***</td>
</tr>
<tr>
<td>INCOME DUMMY</td>
<td>-0.809***</td>
<td>0.402***</td>
<td>-0.257*</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.422*</td>
<td>0.630***</td>
<td>0.812***</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.931***</td>
<td>0.725***</td>
<td>1.211***</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-1.277***</td>
<td>-1.100***</td>
<td>-0.706***</td>
</tr>
<tr>
<td>UNDER</td>
<td>-0.405***</td>
<td>-0.509***</td>
<td>-0.249***</td>
</tr>
<tr>
<td>BIASWET</td>
<td>0.380***</td>
<td>0.020</td>
<td>0.406***</td>
</tr>
<tr>
<td>BIASIRR</td>
<td>0.182*</td>
<td>-0.161</td>
<td>0.098</td>
</tr>
<tr>
<td>MOREINFO</td>
<td>0.212***</td>
<td>0.012</td>
<td>-0.227***</td>
</tr>
<tr>
<td>CONFUSED</td>
<td>-0.328***</td>
<td>-0.074</td>
<td>0.286***</td>
</tr>
<tr>
<td>WILLWORK</td>
<td>-0.372***</td>
<td>-0.139***</td>
<td>-0.087</td>
</tr>
<tr>
<td>ONEOFF</td>
<td>0.523***</td>
<td>0.318***</td>
<td>0.334***</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.129E-2***</td>
<td>-0.158E-2***</td>
<td>-0.160E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>-0.703E-3</td>
<td>0.436E-3**</td>
<td>0.356E-2**</td>
</tr>
<tr>
<td>AREA</td>
<td>0.221E-3</td>
<td>0.103E-3***</td>
<td>0.752E-3***</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.238***</td>
<td>-0.179***</td>
<td>-0.386***</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.055***</td>
<td>0.057***</td>
<td>0.067***</td>
</tr>
</tbody>
</table>

Cross-Effects

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 7: Gwydir Moree</th>
<th>Model 8: Gwydir Sydney</th>
<th>Model 9: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RATES2</td>
<td>-0.277E-2*</td>
<td>-0.117E-2</td>
<td>-0.731E-3</td>
</tr>
<tr>
<td>JOBS2</td>
<td>-0.232E-2</td>
<td>0.110E-2</td>
<td>0.193E-2*</td>
</tr>
<tr>
<td>AREA2</td>
<td>0.507E-3</td>
<td>0.953E-3*</td>
<td>0.262E-3</td>
</tr>
<tr>
<td>BREED2</td>
<td>-0.148</td>
<td>-0.140</td>
<td>0.346E-2</td>
</tr>
<tr>
<td>SPECIES2</td>
<td>0.025</td>
<td>0.038**</td>
<td>0.035**</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 7: Gwydir Moree</th>
<th>Model 8: Gwydir Sydney</th>
<th>Model 9: Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-645.507</td>
<td>-996.127</td>
<td>-1035.385</td>
</tr>
<tr>
<td>$\chi^2$ (constants only)</td>
<td>257.445</td>
<td>361.416</td>
<td>605.352</td>
</tr>
<tr>
<td>$\rho^2$ adjusted</td>
<td>0.153</td>
<td>0.151</td>
<td>0.224</td>
</tr>
<tr>
<td>Iterations completed</td>
<td>5</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Observations</td>
<td>707</td>
<td>1081</td>
<td>1227</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level standard errors not reported
Five dominated alternatives and two implausible alternatives were deleted from the experimental design in the Gwydir Wetlands surveys. This left twenty alternatives that were blocked into four different questionnaires. In the Macquarie Marshes survey an extra block that consisted of the five dominated alternatives was included.

Five MNL models which include only choice set attributes were estimated to demonstrate the effect of excluding deleted alternatives on model parameters (see Table A5.4). One block has been excluded in each model. In model 1, block 1 only has been excluded; in model 2, block 2 only has been excluded, and so on. The dominated alternatives were in block 5, so they were missing from Model 5.
Table A5.4: Macquarie Marshes MNL models with choice set attributes only and one block of the experimental design deleted from each model

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_2$</td>
<td>-0.309</td>
<td>-0.666***</td>
<td>-0.462**</td>
<td>-0.877***</td>
<td>0.540***</td>
</tr>
<tr>
<td>$C_3$</td>
<td>0.087</td>
<td>-0.286</td>
<td>-0.216</td>
<td>-0.546**</td>
<td>-0.360*</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.130E-1***</td>
<td>-0.126E-1***</td>
<td>-0.121E-1***</td>
<td>-0.120E-1***</td>
<td>-0.129E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.204E-2***</td>
<td>0.216E-2***</td>
<td>0.200E-2***</td>
<td>0.160E-2**</td>
<td>0.454E-3</td>
</tr>
<tr>
<td>AREA</td>
<td>0.489E-3***</td>
<td>0.587E-3***</td>
<td>0.613E-3***</td>
<td>0.684E-3***</td>
<td>0.255E-3*</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.246***</td>
<td>-0.321***</td>
<td>-0.301***</td>
<td>-0.362***</td>
<td>-0.315***</td>
</tr>
<tr>
<td>ENDSPECIES</td>
<td>0.043***</td>
<td>0.052***</td>
<td>0.048***</td>
<td>0.047E-1</td>
<td>0.052***</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-1228.937</td>
<td>-1224.763</td>
<td>-1192.444</td>
<td>-1230.878</td>
<td>-1240.332</td>
</tr>
<tr>
<td>$\chi^2$ (constants only)</td>
<td>301.402</td>
<td>334.692</td>
<td>323.428</td>
<td>332.428</td>
<td>226.941</td>
</tr>
<tr>
<td>$r^2$ adjusted</td>
<td>0.120</td>
<td>0.123</td>
<td>0.119</td>
<td>0.124</td>
<td>0.082</td>
</tr>
<tr>
<td>Iterations completed</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Observations</td>
<td>1275</td>
<td>1275</td>
<td>1235</td>
<td>1282</td>
<td>1233</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level; standard errors not reported

The coefficients for the choice set attributes are significant at least at the 5% level (and generally at the 1% level) in all of the above models except model 5. The coefficient for JOBS is insignificant in model 5, and the coefficient for AREA is only significant at the 10% level. This is the model from which the block of dominated alternatives has been deleted. Hence it appears that the dominated alternatives improve the robustness of a model more than non-dominated alternatives. This could partly explain the relative lack of significance of the AREA and JOBS variables in the Gwydir models.

This finding conflicts with the prevailing view in the literature that dominated alternatives provide less information about marginal rates of substitution than non-dominated alternatives (e.g., Carson et al. 1994). One reason that the dominated alternatives may have had such an effect is that they are less confusing for respondents to answer, thus resulting in
reduced variance. While finding the exact cause of this phenomenon is not the objective of this thesis, it will be the topic of later research.

A5.3 Other Models

Several other models were estimated when attempting to deal with IIA violations. These were the multinomial probit, the heteroscedastic extreme value, and the nested logit models. The multinomial probit model proved to be very difficult to estimate. Several simple model specifications were trialed, however the models didn’t converge even when left for a few days\(^\text{94}\). There was mixed success with the heteroscedastic extreme value model. For the Macquarie Marshes data set the heteroscedastic extreme value model was promising, although for the two Gwydir data sets the sample size appeared to be insufficient to estimate a model that included all socioeconomic and attitudinal interactions (see Table A5.5). The ratios of the scale parameters indicate that the errors for each of the alternatives are identically distributed for both Gwydir data sets. However, it is unlikely that such ideal results are true given the findings from the mother logit tests reported in Section A5.1. It is possible that software problems were being experienced. Mother logit tests are recognised as being one of the more powerful forms of IIA testing (Horowitz 1981b), so the earlier finding should be considered more informative.

\(^{94}\) It is possible that this was due to inadequate computer power. The multinomial probit was estimated using a Pentium 166 with 32 megabytes of RAM.
Table A5.5: Heteroscedastic extreme value models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Gwydir Moree</th>
<th>Gwydir Sydney</th>
<th>Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>-0.665</td>
<td>-0.437</td>
<td>-1.368***</td>
</tr>
<tr>
<td>C3</td>
<td>-0.308</td>
<td>-0.138</td>
<td>-1.137***</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.843</td>
<td>-0.741***</td>
<td>0.378***</td>
</tr>
<tr>
<td>HOUSE</td>
<td>-0.461***</td>
<td>0.616***</td>
<td>-0.124</td>
</tr>
<tr>
<td>INCOME ('000)$^95$</td>
<td>-0.262E-2</td>
<td>0.643E-2***</td>
<td>0.879E-2***</td>
</tr>
<tr>
<td>INCOME DUMMY</td>
<td>-0.564**</td>
<td>0.196</td>
<td>-0.212</td>
</tr>
<tr>
<td>VISIT</td>
<td>0.485**</td>
<td>0.671***</td>
<td>0.832***</td>
</tr>
<tr>
<td>PROGRE</td>
<td>0.461***</td>
<td>0.720***</td>
<td>1.092***</td>
</tr>
<tr>
<td>PRODEV</td>
<td>-0.887</td>
<td>-1.053***</td>
<td>-0.825***</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.859E-2***</td>
<td>-0.119E-1***</td>
<td>-0.117E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>-0.637E-3</td>
<td>0.177E-2</td>
<td>0.166E-2***</td>
</tr>
<tr>
<td>AREA</td>
<td>-0.536E-3</td>
<td>0.250E-3</td>
<td>0.482E-3***</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.128</td>
<td>-0.109*</td>
<td>-0.273***</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.032*</td>
<td>0.036***</td>
<td>0.045***</td>
</tr>
</tbody>
</table>

Ratio of Scale Parameters

<table>
<thead>
<tr>
<th></th>
<th>Alternative A/Alternative C</th>
<th>Alternative B/Alternative C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.000**</td>
<td>1.000***</td>
</tr>
<tr>
<td></td>
<td>1.000**</td>
<td>1.000***</td>
</tr>
</tbody>
</table>

Summary statistics

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-883.439</td>
<td>-1233.117</td>
</tr>
<tr>
<td>$\chi^2$ (constants only)</td>
<td>149.102</td>
<td>322.044</td>
</tr>
<tr>
<td>$\rho^2$ adjusted</td>
<td>0.069</td>
<td>0.110</td>
</tr>
<tr>
<td>Observations</td>
<td>872 (237 skipped)</td>
<td>1269 (160 skipped)</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level
standard errors not reported

Nested logit models were also estimated for the Gwydir Sydney and Macquarie Marshes data sets (see Table A5.6). It was not possible to estimate a model for the Gwydir Moree data set, although several different tree structures were trialed. In the nested logit models estimated for the two Sydney data sets, the two environmental improvement options were grouped in one branch and the continue current situation option was in the other branch. A branch-choice equation was specified in which respondents first choose between ‘more’ water for the wetlands or continuing the ‘current’ situation. The utilities of these two branches depends on an alternative specific constant (which is on the ‘current’ branch) and its interaction with several socioeconomic and attitudinal

$^95$ The heteroscedastic extreme value model has trouble converging when income is entered in thousands.
variables. At the second level of the nest, respondents are assumed to choose on the basis of the attributes of the alternatives.

For both the Gwydir Sydney and Macquarie Marshes data sets the nested logit models appear to be as robust as the MNL models. The increased level of significance of both the JOBS and AREA coefficients is suggestive of an improved model fit for the Gwydir Sydney data.

The inclusive value provides an indication of the relatedness of alternatives within a branch. Inclusive values that are closer to zero than one indicate a high degree of relatedness between alternatives (See Train, McFadden and Ben-Akiva 1987). Another way to test for IIA violations in the basic MNL is to determine whether the inclusive value is less than one, as this implies relatedness between alternatives and hence correlated error terms. T-tests indicate that there are significant differences from unity for both inclusive values. Hence this tests indicates that the MNL with the above variables included would suffer from IIA violations, due to the relatedness of the alternatives. This contrasts in part with the finding in Chapter 7 that the main cause of IIA violations was the existence of random taste variations. However, the specification of the models presented in Chapter 7 and those presented above are not identical. Hence, it appears that there were several factors leading to IIA violations.

The nested logit models were not used in the benefit transfer tests presented in this thesis. The reasons for this are that they are more complex than the MNL model, and it was not possible to estimate a model using the Gwydir Moree data set. Hence it would not have been possible to conduct all of the hypothesis tests presented in Chapter 10 if nested logit models were used. Yet these models appear to be robust, and their use for estimating non-use values and for benefit transfer appears to be deserving of further research.
Table A5.6: Nested logit models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Gwydir Sydney</th>
<th>Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>-0.363***</td>
<td>-0.267***</td>
</tr>
<tr>
<td>$C_3$</td>
<td>-5.469*</td>
<td>-2.411</td>
</tr>
<tr>
<td>CHILD</td>
<td>0.446***</td>
<td>-0.343**</td>
</tr>
<tr>
<td>INCOME (’000)</td>
<td>-0.830E-5***</td>
<td>-0.704E-5***</td>
</tr>
<tr>
<td>INCOME DUMMY</td>
<td>-0.307</td>
<td>0.409*</td>
</tr>
<tr>
<td>VISIT</td>
<td>-0.635***</td>
<td>-0.545***</td>
</tr>
<tr>
<td>PROGRE</td>
<td>-0.681***</td>
<td>-1.266***</td>
</tr>
<tr>
<td>PRODEV</td>
<td>1.027***</td>
<td>1.115***</td>
</tr>
<tr>
<td>RATES</td>
<td>-0.147E-1***</td>
<td>-0.164E-1***</td>
</tr>
<tr>
<td>JOBS</td>
<td>0.393E-2***</td>
<td>0.261E-2***</td>
</tr>
<tr>
<td>AREA</td>
<td>0.789E-3**</td>
<td>0.576E-3***</td>
</tr>
<tr>
<td>BREED</td>
<td>-0.150***</td>
<td>-0.393***</td>
</tr>
<tr>
<td>SPECIES</td>
<td>0.508E-1***</td>
<td>0.682E-1***</td>
</tr>
<tr>
<td><strong>Inclusive value</strong></td>
<td>0.521***</td>
<td>0.697***</td>
</tr>
</tbody>
</table>

Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Gwydir Sydney</th>
<th>Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-1247.536</td>
<td>-1038.754</td>
</tr>
<tr>
<td>$\chi^2$ (constants only)</td>
<td>492.392</td>
<td>524.566</td>
</tr>
<tr>
<td>Iterations completed</td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td>rho2 adjusted</td>
<td>0.160</td>
<td>0.197</td>
</tr>
<tr>
<td>Observations</td>
<td>1429 (146 skipped)</td>
<td>1233 (103 skipped)</td>
</tr>
</tbody>
</table>

Note: *** significant at 1% level, ** significant at 5% level, * significant at 10% level standard errors not reported
Before water from the Gwydir River was diverted for irrigation, the Gwydir Wetlands were the third largest wetlands in NSW. They had an area of about 2000 square kilometres that contained reeds, rushes and other aquatic plants as well as acacia and coolabah woodlands.

The Gwydir Wetlands were an important habitat for birds. Before irrigation, some 225 bird species were found in the area, with 125 species breeding there. The wetlands were one of the largest breeding areas in NSW for waterbirds such as ibis, egrets, spoonbills and herons, species which breed only in large groups. These waterbirds breed in only a few other wetlands in NSW. Before irrigation, the number of waterbirds in a single breeding event stimulated by flooding ranged from 5000 pairs to well over 200,000 pairs.

Before irrigation, the Gwydir Wetlands were used as habitat by 19 bird species listed as endangered by the National Parks and Wildlife Service and 18 migratory waterbirds protected under international agreements.

All of the Gwydir Wetlands are owned by private landholders. Several whole properties have been declared wildlife refuges. The Gwydir Wetlands have been used for grazing sheep and cattle since the 1800s. The wetlands provide a high quality stock feed known as ‘water couch’.
Options for the Gwydir Wetlands: A Community Survey

Dear Respondent

I would like to ask for your help in completing this survey as part of a University of New South Wales research project. The survey is being undertaken to find out the views of the community on a range of possible options for the Gwydir Wetlands.

The survey should take about 20 minutes to complete. Your participation is voluntary, but your support would be much appreciated. We would like to get as many opinions from as many different people as possible. Your answers will be kept strictly confidential.

Funding for this survey has been provided by several NSW and Federal Government agencies in order to provide an understanding of the community’s views about the Gwydir Wetlands.

If you would like to receive a summary of the results of this survey, please fill out the form below and put it inside your completed questionnaire.

If you have any other questions about the survey I can be contacted on (06) 268 8833.

Jeff Bennett
ASSOCIATE-PROFESSOR
10 June 1997

---

Yes, I would like to receive a summary of the results of this survey:

Name

Address
Options for the Gwydir Wetlands in Northern New South Wales

A Community Survey
1997
How To Fill Out This Questionnaire

In most cases you only need to tick the box which is closest to your view. Here is an example:

EXAMPLE

Do you think the government should spend more or less on education?

- Spend more on education
- Spend less on education
- Don’t know

If you think the government should spend more on education, you would tick the box as shown.

If you have any questions regarding this survey, please contact:
Associate-Professor Jeff Bennett of
The University of New South Wales on (06) 268 8833

Completion of this survey is voluntary

All your answers will be kept strictly confidential

The information in this questionnaire has been primarily derived from the following sources:

5. Royal Australian Ornithological Union (1920s-1930s). The Emu. Various articles
Your views on the Gwydir Wetlands

The Gwydir Wetlands are in the Gwydir Valley, about 600 kilometres north of Sydney and about 70 kilometres west of Moree (see the enclosed map).

The building of Copeton Dam on the Gwydir River in 1976 enabled the development of a large area of irrigated agriculture. This has provided significant benefits to the regional economy. However, the use of water for irrigation has meant that less water now flows into the Gwydir Wetlands. This has caused reductions in the size and quality of the Gwydir Wetlands.

There are several options available to the government in allocating water to irrigation and to the Gwydir Wetlands. We would like to ask you about your views on these options.

What issues are important?

Question 1

Before considering what should be done about the Gwydir Wetlands, we would first like you to think about a range of environmental issues. Please RANK the following five environmental goals by placing the numbers 1 (most deserving of continued government funding) to 5 (least deserving) in the following boxes:

2. Improving water quality in NSW rivers
5. Improving the quality of wetlands in NSW
4. Reducing soil erosion in rural areas in NSW
1. Reducing logging in the native forests of NSW
3. Improving air quality in NSW urban centres
THE GWYDIR WETLANDS

Before water from the Gwydir River was diverted for irrigation, the Gwydir Wetlands were the third largest wetlands in NSW. The wetlands are still an important habitat for birds, especially for waterbird breeding. This includes habitat for endangered birds and visiting migratory waterbirds. More detailed information about the wetlands is included under the map.

How has less water affected the Gwydir Wetlands?

Wetland area

Since 1976, water that once would have flowed into the wetlands has been used for irrigation. As a result, the area of the Gwydir Wetlands has fallen from 2000 to 400 square kilometres (for comparison, the entire area of Sydney is about 4000 square kilometres). In what remains of the wetlands, the area of weeds—such as lippia and roly-poly—has increased significantly.

Waterbirds

Before the construction of Copeton Dam, waterbird breeding used to occur every 2 years. It now occurs every 5 years. The number of endangered bird species using the Gwydir Wetlands as habitat has fallen from 19 to 5 species since 1976. The number of visiting migratory species protected under international agreements which use the wetlands as habitat has fallen from 18 to 7.

IRRIGATION IN THE GWYDIR VALLEY

Production

The most widespread irrigated crop is cotton. The Gwydir Valley is the largest cotton producing area in Australia, with up to 900 square kilometres harvested each year. The annual revenue from cotton production and other irrigated crops is shown in the following chart:

- Irrigated cotton production $110 million
- Other irrigated crops $25 million
- Dryland agricultural production $100 million
Employment

About 1600 people are employed on irrigated farms in the Gwydir Valley. Two towns in the valley—Moree and Mungindi—rely heavily on the revenue from irrigated agricultural production. It is estimated that irrigation creates about 1200 jobs in these towns. The total employment due to irrigation is about 2800 jobs, which is about 30% of employment in the region.

Water use

Most of the water in Copeton Dam, about 500,000 megalitres, is allocated for irrigation (for comparison, it takes about 1 megalitre to fill an olympic swimming pool). After Copeton Dam was built, farmers received licences to irrigate from the Department of Water Resources. These licences specify an amount of water that will be allocated to farmers if water is available. However, water shortages mean that farmers with licences receive the full amount allocated to them about 50% of the time. Some farmers supplement the water they are allocated by collecting water from tributaries to the Gwydir River and from groundwater bores.

Water trading

Water licences are currently bought and sold in a water trading market which covers the entire Gwydir Valley. In this market, farmers who do not want all of the water they have been allocated may sell the rights to receive this water to the highest bidder. The price of water licences in the Gwydir Valley is currently between $600 to $700 per megalitre.

OPTIONS FOR THE GWYDIR WETLANDS

One option for the Gwydir Wetlands is to continue the current situation. The amount of water allocated for both irrigation and the Gwydir Wetlands would remain the same. Under this option, wetlands size and quality are expected to remain at their current levels.

Another option is to increase the amount of water allocated to the Gwydir Wetlands. This would increase wetlands size and improve wetlands quality.

One way of increasing water for the Gwydir Wetlands is for the NSW Government to purchase water licences permanently from farmers using the existing water trading market. The sale of water licences would be voluntary.
Farmers would be expected to sell their water licences only if the price offered was high enough. The money received from selling their water licences could enable farmers to purchase more efficient irrigation equipment, such as drip irrigation. This would reduce their water needs. Other farmers may convert to dryland farming. The reduction in the use of water for irrigation may cause a decline in employment on farms and in Moree and Mungindi.

Purchasing water licences from farmers would be costly and the NSW Government does not have money to purchase the water licences from existing taxation revenue. If the government were to purchase water licences from farmers, it would be necessary to charge all households in NSW a one-off levy on water rates, payable only in 1998. The revenue would go into a special fund that would only be used to purchase water licences permanently from farmers during the following year.

WHICH OPTION DO YOU PREFER?

We are interested to find out your views on some possible options for the Gwydir Wetlands. To do this, we have prepared five sets of possible options. We would like to know which option you prefer in each of the five sets of options shown below.

The outcomes in each of the options have been specifically defined so that you have a broad range of choices. Within this range, some options may seem strange according to your experience, but bear in mind that there are many different ways of managing water. For example, wetland vegetation requires regular floods of various sizes to remain healthy. In contrast, waterbird breeding requires flooding of sufficient size at a suitable time of year.

When thinking about which option you prefer, keep in mind your available income and all the other things that you spend money on. It is possible that, in the future, other environmental projects may cost you additional money.
EXAMPLE

Here is an example of how to answer questions 2 to 6. Suppose options 1, 2 and 3 are the only available options for the Gwydir Wetlands. If you preferred option 1, you would tick the box for option 1 as shown below:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$150 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>2800 jobs</td>
<td>2780 jobs</td>
<td>2700 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>400 km²</td>
<td>900 km²</td>
<td>550 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 4 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

☑️ I would choose option 1

☐  I would choose option 2

☐  I would choose option 3

☐  I would not choose any of these options because I would prefer more water to be allocated for irrigation
Now please answer ALL questions from 2-6

**Question 2**

Suppose options 1, 2 and 3 are the **ONLY** ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Your water rates (one-off increase)</strong></td>
<td>no change</td>
<td>$20 increase</td>
<td>$50 increase</td>
</tr>
<tr>
<td><strong>Irrigation related employment</strong></td>
<td>2800 jobs</td>
<td>2800 jobs</td>
<td>2780 jobs</td>
</tr>
<tr>
<td><strong>Wetlands area</strong></td>
<td>400 km²</td>
<td>750 km²</td>
<td>900 km²</td>
</tr>
<tr>
<td><strong>Waterbirds breeding</strong></td>
<td>every 5 years</td>
<td>every 3 years</td>
<td>every 3 years</td>
</tr>
<tr>
<td><strong>Endangered &amp; protected species present</strong></td>
<td>12 species</td>
<td>25 species</td>
<td>25 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1

☑ I would choose option 2

☐ I would choose option 3

☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation
**Question 3**

Suppose options 1, 2 and 3 are the **ONLY** ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$20 increase</td>
<td>$20 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>2800 jobs</td>
<td>2700 jobs</td>
<td>2800 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>400 km²</td>
<td>750 km²</td>
<td>750 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 3 years</td>
<td>every 2 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

- [ ] I would choose option 1
- [ ] I would choose option 2
- [x] I would choose option 3

- [ ] I would not choose any of these options because I would prefer more water to be allocated for irrigation
Question 4

Suppose options 1, 2 and 3 are the ONLY ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$50 increase</td>
<td>$150 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>2800 jobs</td>
<td>2780 jobs</td>
<td>2800 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>400 km²</td>
<td>750 km²</td>
<td>750 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 2 years</td>
<td>every 3 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>15 species</td>
<td>25 species</td>
</tr>
</tbody>
</table>

- [ ] I would choose option 1
- [ ] I would choose option 2
- [ ] I would choose option 3
- [ ] I would not choose any of these options because I would prefer more water to be allocated for irrigation
Question 5

Suppose options 1, 2 and 3 are the ONLY ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$50 increase</td>
<td>$50 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>2800 jobs</td>
<td>2700 jobs</td>
<td>2700 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>400 km²</td>
<td>900 km²</td>
<td>750 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 4 years</td>
<td>every 3 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1

☐ I would choose option 2

☑ I would choose option 3

☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation
**Question 6**

Suppose options 1, 2 and 3 are the **ONLY** ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to Gwydir Wetlands</th>
<th>Option 3: Increase water to Gwydir Wetlands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$150 increase</td>
<td>$150 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>2800 jobs</td>
<td>2780 jobs</td>
<td>2780 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>400 km$^2$</td>
<td>750 km$^2$</td>
<td>900 km$^2$</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 5 years</td>
<td>every 4 years</td>
<td>every 2 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>20 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

![ ] I would choose option 1

![ ] I would choose option 2

![ ] I would choose option 3

![ ] I would not choose any of these options because I would prefer more water to be allocated for irrigation
In this section of the questionnaire we would like to ask you some further questions about the options for the Gwydir Wetlands

Question 7

Which of the following outcomes were the most important to you in answering questions 2 to 6? (tick up to two boxes)

- [ ] Your water rates
- [ ] Irrigation related employment
- [ ] Wetlands area
- [ ] Waterbirds breeding
- [ ] Endangered and protected species present

Question 8

Thinking about the information provided earlier, please indicate how strongly you agree or disagree with each of the following statements (circle the number which is closest to your view):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neither Agree or Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I understood the information in the questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The information in the questionnaire was biased in favour of the wetlands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The information in the questionnaire was biased in favour of irrigation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I needed more information than was provided</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I found questions 2 to 6 confusing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I thought the proposal to purchase water from the farmers would work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don't trust the government to make the increase in water rates one-off</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We would now like to ask you a few questions to make sure the people we are surveying are from a wide range of backgrounds

**Question 9**

Over the years, when you have heard about proposed projects where there is a conflict between development and the environment, have you tended to:

- [ ] Favour preservation of the environment more frequently
- [ ] Favour development more frequently
- [x] Favour development and environmental preservation equally

**Question 10**

Have you ever visited the Gwydir Wetlands?

- [ ] Yes
- [x] No

Do you think you will ever visit the Gwydir Wetlands?

- [x] Yes
- [ ] No

**Question 11**

What is your age?

- [ ] 18-29
- [ ] 30-39
- [x] 40-49
- [ ] 50-59
- [ ] 60-69
- [ ] 70-79
- [ ] 80+
**Question 12**

What is your sex?  

☐ Male  

☒ Female

**Question 13**

Do you have any children?  

☐ Yes  

☒ No

**Question 14**

Do you own (or are you paying off) the house that you are living in?  

☒ Yes  

☐ No

**Question 15**

What is the highest level of education you have obtained or are obtaining?  

☐ Never went to school  

☐ Secondary/Year 12

☐ Primary only  

☐ Diploma or certificate

☐ Junior/Year 10  

☒ Tertiary degree

Other (please specify) ____________________________________________
**Question 16**

To the best of your knowledge please indicate the total income (before taxes) that you and your spouse (if applicable) earned last year.

- □ Under $6,239
- □ $10,400-15,599
- □ $20,800-25,999
- □ $31,200-36,399
- □ $41,600-51,999
- □ $78,000-103,999
- □ Don't know
- □ $6,240-10,399
- □ $15,600-20,799
- □ $26,000-31,199
- □ $36,400-41,599
- □ $52,000-77,999
- □ More than $104,000

**Question 17**

What is your current work status?

- □ Employed full or part time
- □ Unemployed/looking for work
- □ Retired/pensioner
- □ Full-time student
- □ Home duties (home-maker)

Other (please specify)________________________________________________________________________
If you would like to make any other comments about this issue, or about this questionnaire, please make them in the following space.

Your completed questionnaire will be collected by a surveyor at an agreed time. If you are unable to give the questionnaire to the surveyor at this time, please use the reply paid envelope to mail it to:

Reply Paid 207
Associate-Professor Jeff Bennett
School of Economics and Management
The University of New South Wales
Locked Bag 25
Kingston ACT 2604

Thank you very much for doing this questionnaire. We hope that you enjoyed taking part in this survey.
Options for the Macquarie Marshes: A Community Survey

Dear Survey Respondent

Recently you were given a survey about options for the Macquarie Marshes. When we came to pick up your completed questionnaire nobody was at home.

Your views are very important to us. We would like to get as many different opinions from as many different people as possible.

Could you please put your completed questionnaire in the replied paid envelope and mail it back to the following address as soon as possible;

   Reply Paid 207  
   Associate-Professor Jeff Bennett  
   School of Economics and Management  
   The University of New South Wales  
   Locked Bag 25  
   Kingston ACT 2604

If you have any questions about the survey please contact Associate-Professor Jeff Bennett of The University of New South Wales on (06) 268 8833.

Thank you for your assistance.
Options for the Macquarie Marshes: A Community Survey

Dear Respondent

I would like to ask for your help in completing this survey as part of a University of New South Wales research project. The survey is being undertaken to find out the views of the community on a range of possible options for the Macquarie Marshes.

The survey should take about 20 minutes to complete. Your participation is voluntary, but your support would be much appreciated. We would like to get as many opinions from as many different people as possible. Your answers will be kept strictly confidential.

Funding for this survey has been provided by several NSW and Federal Government agencies in order to provide an understanding of the community's views about the Macquarie Marshes.

If you would like to receive a summary of the results of this survey, please fill out the form below and put it inside your completed questionnaire.

If you have any other questions about the survey I can be contacted on (06) 268 8833.

Jeff Bennett
ASSOCIATE-PROFESSOR
7 October 1997

Yes, I would like to receive a summary of the results of this survey:

Name__________________________
Address__________________________
Before water from the Macquarie River was diverted for irrigation, the Macquarie Marshes were the largest wetlands in NSW. They had an area of about 5000 square kilometres that contained reeds, rushes and other aquatic plants as well as river redgum and coolabah woodlands.

Before irrigation, some 200 bird species were found in the Macquarie Marshes. The marshes were the largest breeding area in NSW for waterbirds such as ibis, egrets, spoonbills and herons, species which bred only in large groups. These waterbirds breed only in few other wetlands in NSW. Before irrigation, the number of waterbirds in a single breeding event stimulated by flooding ranged from 5000 pairs to 1,000,000 pairs.

Before irrigation, the Macquarie Marshes were used a habitat by 12 species listed as endangered by the National Parks and Wildlife Service and 19 waterbirds protected under international agreements.

The Macquarie Marshes acts as a filter that improves downstream water quality in the Macquarie River. It does this by causing sediment to settle out and by trapping other impurities. 

Within the Macquarie Marshes is a Nature Reserve covering approximately 180 square kilometres. The Nature Reserve is listed as a wetland of international importance.

Parts of the Macquarie Marshes have been used for grazing sheep and cattle since the 1830's. The wetlands provide a high quality of stock feed known as 'water couch'.
Options for the Macquarie Marshes in Northwest New South Wales

A Community Survey
1997
How To Fill Out This Questionnaire

In most cases you only need to tick the box which is closest to your view. Here is an example:

EXAMPLE

Do you think the government should spend more or less on education?

☑ Spend more on education

☐ Spend less on education

☐ Don’t know

If you think the government should spend more on education, you would tick the box as shown.

If you have any questions regarding this survey, please contact:
Associate-Professor Jeff Bennett of
The University of New South Wales on (06) 268 8833.

Completion of this survey is voluntary

All your answers will be kept strictly confidential

The information in this questionnaire has been primarily derived from the following sources:

Your views on the Macquarie Marshes

The Macquarie Marshes are in the Macquarie Valley, about 500 kilometres north-west of Sydney and about 70 kilometres north-west of Dubbo (see the enclosed map).

The building of Burrendong Dam on the Macquarie River in 1967 enabled the development of a large area of irrigated agriculture. This has provided significant benefits to the regional economy. However, the use of water for irrigation has meant that less water now flows into the Macquarie Marshes. This has caused reductions in the size and quality of the Macquarie Marshes.

There are several options available to the government in allocating water to irrigation and to the Macquarie Marshes. We would like to ask you about your views on these options.

What issues are important?

Question 1

Before considering what should be done about the Macquarie Marshes, we would first like you to think about a range of environmental issues. Please RANK the following five environmental goals by placing the numbers 1 (most deserving of continued government funding) to 5 (least deserving) in the following boxes:

- [ ] Improving water quality in NSW rivers
- [ ] Improving the quality of wetlands in NSW
- [ ] Reducing soil erosion in rural areas in NSW
- [ ] Reducing logging in the native forests of NSW
- [ ] Improving air quality in NSW urban centres
THE MACQUARIE MARSHES

Before water from the Macquarie River was diverted for irrigation, the Macquarie Marshes were the largest wetlands in NSW. The wetlands are still an important habitat for birds, especially for waterbird breeding. This includes habitat for endangered birds and visiting migratory waterbirds. More detailed information about the wetlands is included under the map.

How has less water affected the Macquarie Marshes?

Wetland area

Since 1967, water that once would have flowed into the Macquarie Marshes has been used for irrigation. As a result, the area of the Marshes has fallen from 5000 to 1200 square kilometres (for comparison, the entire area of Sydney is about 4000 square kilometres). In what remains of the wetlands, the area of weeds—such as lippia and roly-poly—has increased significantly.

Waterbirds

Before the construction of Burrendong Dam, waterbird breeding used to occur almost every year. It now occurs every 4 years. The number of endangered bird species using the Macquarie Marshes as habitat has fallen from 15 to 5 species since 1967. The number of visiting migratory species protected under international agreements which use the wetlands as habitat has fallen from 19 to 7.

IRRIGATION IN THE MACQUARIE VALLEY

Production

The most widespread irrigated crop grown in the valley is cotton. Other irrigated crops include lucerne, pasture cereals and vegetables. The Macquarie Valley is one of the largest cotton producing areas in Australia, with up to 400 square kilometres harvested each year. The annual revenue from cotton production and other irrigated crops is shown in the following chart:
Employment

About 2400 people are employed on irrigated farms in the Macquarie Valley. Three towns in the valley—Trangie, Warren and Narromine—rely heavily on the revenue from irrigated agricultural production. It is estimated that irrigation creates about 2000 jobs in these towns. The total employment due to irrigation is about 4400 jobs, which is about 10% of employment in the region.

Water use

Most of the water in Burrendong Dam, about 525,000 megalitres, is allocated for irrigation (for comparison, it takes about 1 megalitre to fill an Olympic swimming pool). After Burrendong Dam was built, farmers received licences to irrigate from the Department of Water Resources. These licences specify an amount of water that will be allocated to farmers if water is available. However, water shortages mean that farmers with licences receive the full amount allocated to them about 65% of the time. Some farmers supplement the water they are allocated by collecting water from tributaries to the Macquarie River and from groundwater bores.

In 1995 the NSW Government reduced the amount of water allocated for irrigation by 75,000 ML or 12%, and increased the amount of water for the Macquarie Marshes. This extra water will help maintain the quality and size of the Macquarie Marshes. However further reductions in the size of the Macquarie Marshes are expected.

Water trading

Water licences are currently bought and sold in a water trading market which covers the entire Macquarie Valley. In this market, farmers who do not want all of the water they have been allocated may sell the rights to receive this water to the highest bidder. The price of water licences in the Macquarie Valley is currently between $400 to $500 per megalitre.

OPTIONS FOR THE MACQUARIE MARSHES

One option for the Macquarie Marshes is to continue the current situation. The amount of water allocated for both irrigation and the Macquarie Marshes would remain the same. Under this option, the size of the Marshes is expected to decline from 1200 square kilometres to 1000 square kilometres.

Another option is to increase the amount of water allocated to the Macquarie Marshes. This would increase wetlands size and improve wetlands quality. One way of increasing water for the Macquarie Marshes is for the NSW Government to purchase water licences permanently from farmers using the existing water trading market. The sale of water licences would be voluntary.
Farmers would be expected to sell their water licences only if the price offered was high enough. The money received from selling their water licences could enable farmers to purchase more efficient irrigation equipment, such as drip irrigation. This would reduce their water needs. Other farmers may convert to dryland farming. The reduction in the use of water for irrigation may cause a decline in employment on farms and in Trangie, Warren and Narromine.

Purchasing water licences from farmers would be costly and the NSW Government does not have enough money to purchase the water licences from existing taxation revenue. If the government were to purchase water licences from farmers, it would be necessary to charge all households in NSW a one-off levy on water rates, payable only in 1998. The revenue would go into a special fund that would only be used to purchase water licences permanently from farmers during the following year.

**WHICH OPTION DO YOU PREFER?**

We are interested to find out your views on some possible options for the Macquarie Marshes. To do this, we have prepared five sets of possible options. We would like to know which option you prefer in each of the five sets of options shown below.

The outcomes in each of the options have been specifically defined so that you have a broad range of choices. Within this range, some options may seem strange according to your experience, but bear in mind that there are many different ways of managing water. For example, wetland vegetation requires regular floods of various sizes to remain healthy. In contrast, waterbird breeding requires flooding of sufficient size at a suitable time of year.

When thinking about which option you prefer, keep in mind your available income and all the other things that you spend money on. It is possible that, in the future, other environmental projects may cost you additional money.
EXAMPLE

Here is an example of how to answer questions 2 to 6. Suppose options 1, 2 and 3 are the only available options for the Macquarie Marshes. If you preferred option 1, you would tick the box for option 1 as shown below:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$150 increase</td>
<td>$30 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4350 jobs</td>
<td>4250 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>2000 km²</td>
<td>1250 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every year</td>
<td>every 3 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

I would choose option 1

I would choose option 2

I would choose option 3

I would not choose any of these options because I would prefer more water to be allocated for irrigation
Question 2

Suppose options 1, 2 and 3 are the ONLY ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$20 increase</td>
<td>$20 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4350 jobs</td>
<td>4400 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>2000 km²</td>
<td>2000 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every year</td>
<td>every 2 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>20 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

- [ ] I would choose option 1
- [x] I would choose option 2
- [ ] I would choose option 3
- [ ] I would not choose any of these options because I would prefer more water to be allocated for irrigation
**Question 3**

Suppose options 1, 2 and 3 are the **ONLY** ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$50 increase</td>
<td>$20 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4400 jobs</td>
<td>4350 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>1250 km²</td>
<td>1650 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every 2 years</td>
<td>every 2 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>20 species</td>
<td>20 species</td>
</tr>
</tbody>
</table>

- [ ] I would choose option 1
- [ ] I would choose option 2
- [x] I would choose option 3
- [ ] I would not choose any of these options because I would prefer more water to be allocated for irrigation
**Question 4**

Suppose options 1, 2 and 3 are the **ONLY** ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$50 increase</td>
<td>$150 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4250 jobs</td>
<td>4400 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>1250 km²</td>
<td>1250 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every 2 years</td>
<td>every year</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>20 species</td>
<td>25 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1

☑ I would choose option 2

☐ I would choose option 3

☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation
Question 5

Suppose options 1, 2 and 3 are the ONLY ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$150 increase</td>
<td>$50 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4400 jobs</td>
<td>4350 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>2000 km²</td>
<td>1250 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every 2 years</td>
<td>every 2 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>15 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1

☐ I would choose option 2

☑ I would choose option 3

☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation
Question 6

Suppose options 1, 2 and 3 are the ONLY ones available, which would you choose?

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Option 1: continue current situation</th>
<th>Option 2: Increase water to the Macquarie Marshes</th>
<th>Option 3: Increase water to the Macquarie Marshes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your water rates (one-off increase)</td>
<td>no change</td>
<td>$150 increase</td>
<td>$150 increase</td>
</tr>
<tr>
<td>Irrigation related employment</td>
<td>4400 jobs</td>
<td>4250 jobs</td>
<td>4350 jobs</td>
</tr>
<tr>
<td>Wetlands area</td>
<td>1000 km²</td>
<td>1250 km²</td>
<td>1650 km²</td>
</tr>
<tr>
<td>Waterbirds breeding</td>
<td>every 4 years</td>
<td>every year</td>
<td>every 3 years</td>
</tr>
<tr>
<td>Endangered &amp; protected species present</td>
<td>12 species</td>
<td>25 species</td>
<td>15 species</td>
</tr>
</tbody>
</table>

☐ I would choose option 1

☑ I would choose option 2

☐ I would choose option 3

☐ I would not choose any of these options because I would prefer more water to be allocated for irrigation
In this section of the questionnaire we would like to ask you some further questions about the options for the Macquarie Marshes

**Question 7**

Which of the following outcomes were the most important to you in answering questions 2 to 6? (tick up to two boxes)

- Your water rates
- Irrigation related employment
- Wetlands area
- Waterbirds breeding
- Endangered and protected species present

**Question 8**

Thinking about the information provided earlier, please indicate how strongly you agree or disagree with EACH of the following statements (circle the number which is closest to your view):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither Agree or Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I understood the information in the questionnaire</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The information in the questionnaire was biased in favour of the marshes</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The information in the questionnaire was biased in favour of irrigation</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>I needed more information than was provided</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>I found questions 2 to 6 confusing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>I thought the proposal to purchase water from the farmers would work</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>I don't trust the government to make the increase in water rates one-off</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
We would now like to ask you a few questions to make sure the people we are surveying are from a wide range of backgrounds

**Question 9**

Over the years, when you have heard about proposed projects where there is a conflict between development and the environment, have you tended to:

- [ ] Favour preservation of the environment more frequently
- [ ] Favour development more frequently
- [x] Favour development and environmental preservation equally

**Question 10**

Have you ever visited the Macquarie Marshes?

- [ ] Yes  
- [x] No

Do you think you will ever visit the Macquarie Marshes?

- [x] Yes  
- [ ] No

**Question 11**

What is your age?

- [ ] 18-29  
- [ ] 30-39  
- [x] 40-49

- [ ] 50-59  
- [ ] 60-69  
- [ ] 70-79

- [x] 80+
Question 12

What is your sex?

- [ ] Male
- [x] Female

Question 13

Do you have any children?

- [x] Yes
- [ ] No

Question 14

Do you own (or are you paying off) the house that you are living in?

- [ ] Yes
- [x] No

Question 15

What is the **highest** level of education you have obtained or are obtaining?

- [ ] Never went to school
- [x] Secondary/Year 12
- [ ] Primary only
- [ ] Diploma or certificate
- [ ] Junior/Year 10
- [ ] Tertiary degree

Other (please specify) ______________________________________________________

Question 16
To the best of your knowledge please indicate the total income (before taxes) that you and your spouse (if applicable) earned last year.

- Under $6239
- $6240-10,399
- $10,400-15,599
- $15,600-20,799
- $20,800-25,999
- $26,000-31,199
- $31,200-36,399
- $36,400-41,599
- $41,600-51,999
- $52,000-77,999
- $78,000-103,999
- More than $104,000
- Don't know

Question 17

What is your current work status?

- Employed full or part time
- Full-time student
- Unemployed/looking for work
- Home duties (home-maker)
- Retired/pensioner

Other (please specify) ________________________________
If you would like to make any other comments about this issue, or about this questionnaire, please make them in the following space.

Your completed questionnaire will be collected by a surveyor at an agreed time. If you are unable to give the questionnaire to the surveyor at this time, please use the reply paid envelope to mail it to:

Reply Paid 207
Associate-Professor Jeff Bennett
School of Economics and Management
The University of New South Wales
Locked Bag 25
Kingston ACT 2604

Thank you very much for doing this questionnaire. We hope that you enjoyed taking part in this survey.