ISSN 1327-810X

CHOICE MODELLING RESEARCH REPORTS

CHOICE MODELLING AND TESTS OF BENEFIT TRANSFER*

By M.D. Morrison, J.W. Bennett,
R.K. Blamey and J.J. Louviere

Research Report No. 8
October, 1998

About the authors

Mark Morrison is a Lecturer in the School of Economics and Management, University College, The University of New South Wales, Canberra, ACT.

Jeff Bennett is Associate Professor in the School of Economics and Management, University College, The University of New South Wales, Canberra, ACT.

Russell Blamey is a Research Fellow in the School of Economics and Management, University College, The University of New South Wales, Canberra, ACT.

Jordan Louviere is Professor and Chair in the Department of Marketing, University of Sydney.

* An earlier version of this report was presented at the World Congress of Environmental and Resource Economists, Venice, Italy, June 25-27 1998.

Choice Modelling Research Reports are published by the School of Economics and Management, University College, The University of New South Wales, Canberra 2600 Australia.

These reports represent the provisional findings of the research project 'Using Choice Modelling to Estimate Non-Market Values'.

The project is being funded by the Land & Water Resources Research and Development Corporation and Environment Australia under their joint program on the conservation and management of remnant vegetation. Support for the project is also being provided by the Queensland Department of Primary Industries, the Queensland Department of Natural Resources, the New South Wales National Parks and Wildlife Service and the New South Wales Environment Protection Authority.

Any comments will be gratefully received and should be directed to Associate Professor Jeff Bennett, School of Economics and Management, University College, The University of New South Wales, Canberra ACT Australia 2600. Telephone (02) 6268 8833.
Email: j.bennett@adfa.oz.au Fax: (02) 6268 8450

Table of Contents

Abstract

Choice modelling (CM) is a stated preference technique that is increasingly being used to generate estimates of non-market values. In CM applications, respondents to a survey are presented with several sets of options, which contain common attributes but at varying levels, and are asked to select their preferred alternative from each set. The information provided by respondents' choices is used to derive estimates of value.

The results from three separate CM applications are presented in this report. Estimates are made of the value of improving the quality of wetlands. The first two applications focused on the Gwydir Wetlands in New South Wales, Australia. One survey was conducted in Sydney, a major urban centre, and the second was conducted in Moree, a rural centre close to the Gwydir Wetlands. The third survey, which was also conducted in Sydney, focused on the Macquarie Marshes, another wetland in New South Wales.

An objective of the study is to test the suitability of using CM derived estimates for benefit transfer, both across different populations and across different wetlands. CM is potentially suited to benefit transfer because it is possible to allow for differences in environmental improvements between sites as well as differences in socio-economic characteristics between respondent populations. Several tests aimed at determining the validity of transfers using CM across sites and populations are presented. These include tests of the equality of models, implicit prices and estimates of compensating surplus.

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):

> 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 used.

Resource economists have been somewhat divided in their view about the validity of using benefit transfer. Boyle and Bergstrom (1992) suggest 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.

There have been attempts to develop more rigorous approaches to benefit transfer in recent years. Various 'protocols' have been developed to assist in selecting appropriate studies for use in benefit transfer (see Desvousges, Naughton and Parsons 1992, Boyle and Bergstrom 1992, Smith 1992, and Kask and Shogren 1994). Protocols have been developed in order 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 systematically altered to remove any bias (Boyle and Bergstrom 1992). Analysts need to know the factors that affect willingness to pay and the likely direction and magnitude of the effect to make this decision. Compilation of such data would appear to be critical for use in defensible and valid benefit transfer.

Boyle and Bergstrom (1992) suggest 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 values 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.

Boyle and Bergstrom (1992) also suggest that studies based on primary data be undertaken to determine which variables have a significant effect on value estimates. The objective of this report is to provide information relevant to the first of these lines of investigation, namely the validity of benefit transfer.

In order to achieve this objective, three separate surveys have been conducted in which the non-use value associated with improved wetland quality was estimated. Rather than contingent valuation, we use choice modelling (CM) as the basis for these tests. An advantage of CM for benefit transfer is that it is possible to allow for differences in improvements in environmental quality as well as differences in socio-demographics when transferring value estimates. In CM studies, estimates are derived of the value of individual attributes (eg number of species, hectares of wetland etc). It is therefore possible to value any policy alternative that is within the space described by the attributes used in the survey.

A Lancastrian approach to demand theory is used in CM. Lancaster (1966) argued that goods could be defined as a set of attributes. By examining the relative importance people place on these attributes it is possible to determine their value. For example, potential car buyers typically trade-off between increased price and features such as air-conditioning, power steering etc. In a similar vein, CM surveys present respondents with several sets of resource use options defined by several attributes (eg price, quality), and respondents are asked to choose their preferred option in each set. The trade-offs that respondents make when choosing an option are used to estimate compensating surplus. Examples of the use of CM for non-market valuation include Adamowicz, Louviere and Williams (1994), Adamowicz, Swait, Boxall, Louviere and Williams (1996), Boxall, Adamowicz, Swait, Williams and Louviere (1996) and Morrison, Bennett and Blamey (1998).

The structure of this report is as follows. Existing tests of benefit transfer are reviewed in Section 2. The case studies are described and the questionnaire design is reviewed in Section 3 and 4. The survey results are reported in Section 5. In Section 6, the tests of benefit transfer are presented and in Section 7, conclusions are offered.

2 Existing Tests of Benefit Transfer

Two main types of tests have been conducted to determine the accuracy of benefit transfer[1]. These are tests of the transferability of results across different sites and across different populations. The first type of test aims to determine the effect of physical characteristics of the non-market good on the equivalence of results, while the third consider the effect of characteristics of the population. Studies that have undertaken these tests are reviewed in this section.

2.1 Transfers across sites

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). Two tests of this hypothesis have been conducted in the non-market valuation literature (Loomis 1992, Bergland, Magnussen and Navrud 1995)[2].

Loomis (1992) tested this transferability hypothesis 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 at the 95% significance level. 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 6 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 contingent valuation method. 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 was equivalent using a likelihood ratio test, and found that the models differed statistically. They also

[1] A third type of benefit transfer test is of transfers over time. See for example Kealy, Montgomery and Dovidio (1990), Loomis (1990), Reiling, Boyle, Phillips and Anderson (1990), Stevens, Moore and Glass (1994) and Teisl, Boyle, McCollum and Reiling (1995). Most of these studies have concluded that value estimates are relatively stable over a few years.

[2] Tests have also been conducted in the transport literature (Watson and Westin 1975, Atherton and Ben-Akiva 1976). The results from these studies were generally more positive than those involving valuation of non-market goods.

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 suggests that the differences are not large. Bergland et al (1995) suggest 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 takes changing environmental quality and differing socio-economics explicitly into account'. This comment is supportive of the use of choice modelling for benefit transfer because of its ability to allow for both differences in changes in environmental quality and differences in socioeconomic characteristics.

2.2 Transfers across populations

The next type of test involves determining whether different population groups have the same 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 studies 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 multinomial logit models using travel cost data in order 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 the 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[3]. 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 hypothesis. The context for their study was the avoidance of impacts on natural and social resources due to landfill siting. They estimated separate multinomial logit models for urban and non-urban residents using stated preference 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, although it is possible that this could reflect differences in variance as discussed shortly. A likelihood ratio test was not conducted to determine whether the models were different.

In conclusion, the existing evidence is mixed regarding the validity of transfers across either sites or populations, although the amount of evidence available is limited. 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. Two tests of the validity of benefit transfer are conducted in the following sections. The first is a test of the transferability solely across sites, and the second is of transferability solely across populations. Following the suggestion of Bergland et al (1995), a technique which allows for differences in environmental improvements—choice modelling—is used for this testing. First, the two case study sites are described.

3 Case Studies Used in Benefit Transfer Tests

In this report, estimates are reported of the value of improving the quality of two large ephemeral wetlands in northern New South Wales, Australia. The first wetland is the Macquarie Marshes, which were originally the largest wetlands in New South Wales with an area of about 5000 km^2. A Nature Reserve contained in the Marshes is listed as a wetland of international importance under the Ramsar Convention. The Marshes have a

[3] See Section 6.1.

number of significant environmental values: an important habitat for waterbirds, filter that improves downstream water quality, and production of high quality stock feed. The second wetland is the Gwydir Wetlands, which were originally the third largest wetland in New South Wales, with an area of over 2000 km^2. Similar to the Macquarie Marshes, the Gwydir Wetlands is an important waterbird habitat, has unique vegetation, and produces high quality stock feed.

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 the increasing use of water for irrigation has been a reduction in the amount of water reaching both the Macquarie Marshes and Gwydir Wetlands. These wetlands and their biota have been substantially affected by changes to the flow regime caused by regulation and use of water for irrigation. For example, since 1967 the area of the Marshes has fallen from 5000 km^2 to 1200 km^2 and weeds have affected much of the remaining wetland. The frequency of waterbird breeding events has fallen from every year to every four years, and the number of endangered and protected waterbird species using the wetlands has fallen from 34 to 12 species.

In September 1995 the New South Wales Government announced changes to the allocation of water in the Macquarie and Gwydir Valleys. In the Macquarie Valley extra environmental flows were allocated to the Marshes, and restrictions were placed on the use of flows from tributaries. Restrictions were only placed on the use of water from tributaries in the Gwydir Valley. These changes are expected the help maintain the quality and size of the Macquarie Marshes and Gwydir Wetlands, but some further degradation is expected. The New South Wales Government also announced that river flow objectives would be set for each catchment in the state, and these would be reviewed every two years.

Therefore, there are 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 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. Therefore, these two sites provide a suitable context for using CM to estimate non-use values, and testing the transferability of benefit estimates. The questionnaires used in these CM surveys are described in the next section.

4 Questionnaire Description and Design

The questionnaire[4] used for this case study was developed using the results from eight focus groups and a pretest (see Morrison, Bennett and Blamey 1997). The focus groups were used to help determine which attributes should be included in the choice sets, and refine a draft questionnaire. A pretest of 50 respondents was undertaken in June 1997 in Sydney. On the basis of the pretest only minor modifications to the questionnaire were required.

The questionnaires for the Macquarie Marshes and Gwydir Wetlands were identical, except for site specific information. The questionnaires told respondents that there were three broad options available for the management of the wetlands: to continue the current situation, to increase water for the wetlands, or to increase water for irrigation. The scenario presented to respondents was that it would be possible to purchase water for the wetlands from farmers on the existing water trading market. Respondents were told that the Government did not have sufficient money to purchase the water from existing revenue and that it would be necessary to charge households in New South Wales a one-off levy on water rates in 1998. No mention was made of the other site.

Respondents were then presented with six choice sets showing various options for the wetlands, the first of which was an example (see Table 1). The options in the choice sets were defined using five different attributes: water rates, irrigation related employment, wetlands area, frequency of waterbird breeding and

[4] A copy of the questionnaire is available upon request from the principal author.

endangered and protected species present[5]. Respondents were asked for their preferred choice from each set of options.

Respondents were also asked a series of debrief questions to determine their socio-demographic characteristics, their attitudes towards the environment, and their evaluation of the questionnaire. These data were used in the models reported in the next section.

Table 1: Example of a Choice Set from the Macquarie Marshes Questionnaire

Outcome	**Option 1: Continue current situation**	**Option 2: Increase water to Macquarie Marshes**	**Option 3: Increase water to Macquarie Marshes**
Your water rates (one-off increase)	no change	\$20 increase	\$50 increase
Irrigation related employment	4400 jobs	4350 jobs	4350 jobs
Wetlands area	1000 km^2	1250 km^2	1650 km^2
Waterbirds breeding	every 4 years	every 3 years	every year
Endangered and protected species present	12 species	25 species	15 species

☐ 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

5 Survey Results

Three separate CM surveys were conducted. The first survey was conducted in Moree, which is a rural centre close to the Gwydir Wetlands. The second survey also focused on the Gwydir Wetlands, but was conducted in Sydney, which is the main urban centre in New South Wales and over 500 km from the Gwydir Wetlands. The third survey was also conducted in Sydney, but focused on the Macquarie Marshes. Sydney is a similar distance from the Macquarie Marshes. Hence the three surveys were as follows:

1) Gwydir Wetlands case study, survey in Moree (Gwydir Moree)
2) Gwydir Wetlands case study, survey in Sydney (Gwydir Sydney)
3) Macquarie Marshes case study, survey in Sydney

The results from these surveys are described in the following sub-sections.

---

[5] The attribute levels used in the Macquarie Marshes choice sets were: water rates (\$0, \$20, \$50, \$150), employment (4400, 4350, 4250 jobs), wetlands area (1000, 1250, 1650, 2000 km^2), waterbird breeding (4, 3, 2, every year) and species (12, 15, 20, 25 species). In the Gwydir Wetlands choice sets the same attribute levels for rates and species were used. For the remaining attributes the levels were: employment (2800, 2780, 2700 jobs), wetlands area (400, 550, 750, 900 km^2) and waterbird breeding (5, 4, 3, 2 years).

5.1 Survey logistics

A drop-off and pick-up approach was used to distribute the questionnaires based on a cluster sampling technique. The questionnaires for the Gwydir survey in Moree were distributed on 28 and 29 June 1997 and in Sydney between the 12-14 July. The questionnaires for the Macquarie Marshes survey were distributed three months later on the 11 and 12 October.

The response rates for the surveys are listed in Table 2. The response rates from the three surveys were 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 response rate measure, yields a response rate of about 22%.

Table 2: Survey statistics

	Gwydir Wetlands- Moree	Gwydir Wetlands- Sydney	Macquarie Marshes - Sydney
Number of questionnaires distributed (1)	301	349	336
Final (useable) data set (2)	233	294	250
Number of houses with nobody home (3)	606	753	547
Number of rejections (4)	162	259	184
Response rate 1	77.4%	84.2%	74.4%
Response rate 2	50.3%	48.4%	48.1%
Response rate 3	21.8%	21.6%	23.4%

Note: response rate 1 is based on the number of surveys distributed ie [2 ÷1]; 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)].

5.2 Survey statistics

In this section the characteristics of the survey samples are described. The socio-demographics of the two Sydney samples are close to the Sydney average, except for income (see Table 3)[6]. In part the difference can be explained by the inclusion of several regional centres that with lower 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 stated preference surveys. It also may reflect non-response or sampling bias.

Table 3: Socio-demographics of the survey respondents

Variable	Gwydir-Moree	Gwydir-Sydney	Macquarie-Sydney	Sydney average
Age (>17 years)	41.7 years	44.1 years	44.3 years	43.9 years
Sex (% male)	59.4%	55.0%	55.8%	49.2%
Children (%)	74.6%	76.7%	72.1%	67.0%
Own house (%)	54.9%	75.7%	71.3%	67.4%
Education (% > year 12)	70.1%	74.6%	74.6%	77.4%*
Income (household)	$48,127	$51,978	$54,680	$46,184
Employed full or part time (%)	71.8% (> 18 years)	65.6% (> 18 years)	65.7% (> 18 years)	59.3% (>15 years)*
% pro-environment	23.9%	39.5%	35.5%	n.a.
% pro-development	26.7%	10.0%	10.3%	n.a.

*State average

Source: Australian Bureau of Statistics 1996 Census Data

6 Chi-squared tests of independence were conducted to determine whether the Gwydir Sydney and Macquarie Marshes samples had the same socio-demographics as the population. Except for income (both samples) and age (Gwydir Sydney sample only), the null hypothesis of no independence could not be rejected.

5.3 Modelling Results

In this section we present the results of multinomial logit models estimated using the data from the three CM surveys. In each model three different types of independent variables are included: (1) variables representing choice set attributes, (2) variables representing respondents' socio-demographic characteristics and (3) variables representing respondents' evaluation of the questionnaire[7]. Definitions for the coefficients from these models are presented in Table 4.

Table 4: Definitions of variables

Variable	
ASC1, ASC2	Alternative specific constants for options 2, 3 and 4
INCOME	Respondent's household income
INCOMEDUMMY	Dummy variable showing if respondents have not reported their income
CHILD	Dummy variable showing whether respondents have children
VISIT	Dummy variable representing whether a respondent is intending to visit the wetland in the future
PRODEV	Dummy variable showing that a respondent is pro-development
PROGRE	Dummy variable showing that a respondent is pro-environment
UNDER	Likert scale showing whether respondents understood the information in the questionnaire (1=strongly agree, 5=strongly disagree)
WORK	Likert scale showing whether respondents believed the scenario would work (1=strongly agree, 5=strongly disagree)
ONE-OFF	Likert scale showing whether respondents believed that payment would be one-off (1=strongly agree, 5=strongly disagree)
RATE	Water rates
JOBS	Irrigation related employment
AREA	Wetlands area
BREED	Frequency of waterbird breeding
SPECIES	Number of endangered and protected species present

The specification of each of the three models is shown below. There are three utility functions[8] (V_1, V_2 and V_3), with V_1 and V_2 representing the utility of the two alternatives that lead to improved wetland quality, and V_3 representing the utility of the current situation. Socioeconomic, attitudinal and questionnaire evaluation variables were included through interactions with the alternative specific constants[9]. These interactions represent the effect of these variables on the probability that a respondent will choose either alternative 2 or 3.

[7] It was necessary to include the latter type of variables to allow for random taste variations that were found to cause violations of the assumption of identically and independently distributed (IID) error terms. A consequence of the IID assumption is the property of independence of irrelevant alternatives. The IID assumption can be violated for several different reasons. One reason is the existence or random taste variations (ie heterogeneous preferences). In some applications it is possible to remove IID violations due to random taste variations by including additional socio-demographic or attitudinal variables. If not, it may be necessary to use more complex model specifications.

[8] Respondents who chose option 4 (less water to the wetlands) were recoded as choosing the current situation.

[9] It is not possible to include these variables directly in the model as they are invariant across any choice set, and thus would cause a Hessian singularity.

$V_1 = C_1 + C_{CHILD} + C_{INCOME} + C_{INCOMEDUM} + C_{VISIT} + C_{PRODEV} + C_{PROGRE} + C_{UNDER} + C_{WORK} + C_{ONE\text{-}OFF}$
$+\beta_1.RATE+\beta_2.RATES +\beta_3.JOBS+\beta_4.AREA+\beta_5.BREED+\beta_6.SPECIES$

$V_2 = C_2 + C_{CHILD} + C_{INCOME} + C_{INCOMEDUM} + C_{VISIT} + C_{PRODEV} + C_{PROGRE} + C_{UNDER} + C_{WORK} + C_{ONE\text{-}OFF}$
$+\beta_1.RATE+\beta_2.RATES +\beta_3.JOBS+\beta_4.AREA+\beta_5.BREED+\beta_6.SPECIES$

$V_3 = \beta_1.RATE+\beta_2.RATES +\beta_3.JOBS+\beta_4.AREA+\beta_5.BREED+\beta_6.SPECIES$

where C_1 and C_2 are alternative specific constants, and the betas are taste parameters

It is possible to determine *a priori* the expected signs of several of the above variables. PROGRE should have a positive sign because respondents with a pro-environmental orientation would be expected to choose options 2 or 3 more frequently; similarly, PRODEV should to have a negative sign. VISIT should have a positive sign if respondents who intend to visit the wetlands in the future have positive option value (see Bishop 1982). INCOME should have a positive sign as respondents with higher income should have a greater capacity to pay. The *a priori* sign for CHILD, however, is 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. For the questionnaire evaluation variables, if respondents act conservatively, it would be expected that doubts about the questionnaire would result in a reduced likelihood to move away from the status quo and hence choose alternatives 2 or 3. Hence the sign for UNDER and WILLWORK is expected to be negative, and the sign for ONEOFF should be positive.

The results of the multinomial logit models for each of the three surveys are presented in Table 5. The design variables refer to the attributes in the choice sets. In each model the coefficients for RATES, BREED and SPECIES are significant at the 5% level or higher and have *a priori* expected signs. 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, it is significant in the Macquarie Marshes model if a one-tailed t-test is used. Neither the coefficients for JOBS or AREA are significant in the Gwydir Moree model. The insignificance of the JOBS coefficient in Moree is surprising given that this is where impacts on employment would occur.

The coefficients for VISIT, PROGRE, PRODEV, UNDER, WILLWORK and ONE-OFF were all significant at the 1% level and have *a priori* expected signs. The sign of the coefficient for CHILD was positive in the Macquarie Marshes and Gwydir Moree models but negative in the Gwydir Sydney model. One explanation for this result is that, in the first two surveys, the bequest motives had a relatively greater effect than reduced disposable income. The sign of INCOME also differed across the first three surveys. In the two Sydney surveys INCOME, as expected by theory, had a positive sign. In the Moree survey it had a negative sign. To some extent the negative sign could reflect the fact that many higher income people in Moree are involved in the cotton industry, and therefore have a vested interest in seeing less water going to the Gwydir Wetlands.

The explanatory power of the models is satisfactory, with adjusted rho-squared of 13-17%[10]. The chi-squared statistics indicate that each model is significant overall.

[10] Rho-squared is similar to R^2 in conventional regression analysis. It is equal to one minus the ratio of the unrestricted log-likelihood over the restricted log likelihood. Hensher and Johnson (1981) comment 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.'

Table 5: Multinomial logit models

Variables	Macquarie Marshes Sydney	Gwydir Wetlands Sydney	Gwydir Wetlands Moree
ASC1	-1.001**	1.606***	0.878*
	(-2.256)	(3.535)	(1.814)
ASC2	-0.757*	1.897***	1.180***
	(-1.704)	(4.148)	(2.440)
Interactions with ASC's			
CHILD	0.395**	-0.447**	1.157***
	(2.296)	(-2.421)	(5.738)
INCOME	0.529E-5**	0.545E-5**	-0.492E-5*
	(2.006)	(2.237)	(-1.786)
INCOMEDUMMY	-0.334	0.573**	-0.766***
	(-1.292)	(2.377)	(-2.865)
VISIT	0.552***	0.558***	0.614***
	(0.162)	(0.163)	(2.674)
PROGRE	1.149***	0.576***	0.856***
	(6.102)	(3.359)	(3.619)
PRODEV	-1.120***	-1.103***	-1.356***
	(-4.876)	(-4.533)	(-4.279)
UNDER	-0.284***	-0.521***	-0.404***
	(-2.883)	(-5.146)	(-3.681)
WORK	-0.225***	-0.364***	-0.546***
	(-2.706)	(-4.314)	(-7.075)
ONE-OFF	0.343***	0.287***	0.258***
	(4.624)	(4.009)	(2.905)
Design variables			
RATES	-0.145E-1***	-0.136E-1***	-0.106E-1***
	(-12.084)	(-12.109)	(-7.830)
JOBS	0.155E-2	0.291E-2**	-0.700E-3
	(1.591)	(2.209)	(-0.444)
AREA	0.484E-3***	0.527E-3*	-0.480E-3
	(2.936)	(1.640)	(-1.173)
BREED	-0.350***	-0.128**	-0.163**
	(-5.784)	(-2.306)	(-2.322)
SPECIES	0.614E-1***	0.437E-1***	0.410E-1***
	(5.156)	(4.026)	(2.911)
Summary statistics			
Log-likelihood	-944.204	-1041.689	-703.209
Chi-squared (constants only)	406.128	361.570	219.618
Rho-squared adjusted	0.171	0.148	0.128
Number of observations	1045 (178 skipped)	1121 (308 skipped)	742 (367 skipped)

Note: ***significant at 1%, **significant at 5%, *significant at 10%
t-statistics are in brackets

6 Benefit Transfer Tests

Tests of two different benefit transfer hypotheses are reported in this section. The first hypothesis is that benefit transfer *across sites* is valid. Tests of this hypothesis are reported in Section 6.1. The comparison is, therefore, between the Gwydir Sydney and Macquarie Marshes survey results. The second hypothesis is that benefit transfer *across populations* is valid. Tests of this hypothesis are reported in Section 6.2. These tests involve a comparison of the Gwydir Sydney and Gwydir Moree survey results. Hence this is a comparison between a population from a rural town which is close to the environmental good of interest, and a predominantly urban population that is some distance away.

6.1 Tests of transferability across sites

Four different tests are used to test the hypothesis 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 overall models are equivalent across sites

The first test of this hypothesis is whether the overall models are equivalent. Stated formally, the null and alternative hypotheses are:

H_0: $\beta_{MM} = \beta_{GS}$
H_1: $\beta_{MM} \neq \beta_{GS}$

where β_{MM} and β_{GS} are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney datasets.

A likelihood ratio test is used to determine whether $H1_0$ should be rejected. The test statistic is:

$$-2[L_{1+2}-(L_1-L_2)] \sim \chi^2 \text{ with r degrees of freedom}$$

where L_{1+2} is the log-likelihood calculated using both data sets, L_1 and L_2 are the log-likelihoods calculated using the first and second data sets respectively, 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 5. The log-likelihood of the combined model, which is reported in Table 6, 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 6: MNL models estimated using the combined Gwydir Sydney and Macquarie Marshes data sets

Variables	Combined model without rescaling	Combined model with rescaling
C_1	0.328 (0.306)	0.587** (0.252)
C_2	0.586** (0.308)	0.845*** (0.253)
CHILD	0.245E-2 (0.122)	-0.884E-2 (0.137)
INCOME	0.526E-5*** (0.173E-5)	0.530E-5*** (0.194E-5)
INCOMEDUMMY	0.162 (0.171)	0.772E-1 (0.194)
VISIT	0.544*** (0.112)	0.591*** (0.125)
PROGRE	0.854*** (0.123)	0.971*** (0.140)
PRODEV	-1.095*** (0.167)	-1.266*** (0.187)
UNDER	-0.354*** (0.684E-1)	-1.446*** (0.752E-1)
WORK	-0.302*** (0.572E-1)	-0.374*** (0.613E-1)
ONE-OFF	0.315*** (0.496E-1)	0.336*** (0.548E-1)
RATES	-0.139E-1*** (0.813E-3)	-0.158E-1*** (0.914E-3)
JOBS	0.223E-2*** (0.766E-3)	0.245E-2*** (0.825E-3)
AREA	0.313E-3** (0.136E-3)	0.390E-3*** (0.144E-3)
BREED	-0.223*** (0.404E-1)	-0.249*** (0.449E-1)
SPECIES	0.512E-1*** (0.797E-2)	0.565E-1*** (0.897E-2)
Summary statistics		
Log-likelihood	-2012.048	-2007.504
Chi-squared (constants only)	719.888	728.976
ρ^2 adjusted	0.151	0.153
Number of observations	2166 (486 skipped)	2166 (486 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 involves determining whether the models are equivalent after allowing for differences in variance. As discussed in Swait and Louviere (1993), multinomial logit models have a scale parameter (λ) which is inversely proportional to variance, but is confounded with the β 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 datasets 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 multinomial logit model estimated using the stacked data set is maximised. It is then possible to test the hypothesis that two data sets are equal, except for differences in variance (Swait and Louviere 1993) [11]. This is the objective of the second test:

H_0: $\beta_{MM} = \lambda\beta_{GS}$
H_1: $\beta_{MM} \neq \lambda\beta_{GS}$

where β_{MM} and β_{GS} are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney datasets, and λ 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:

$-2[L_\lambda-(L_{MM}-L_{GS})] \sim \chi^2$ with r+1 degrees of freedom

where L_λ 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 6). 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 was allowed to vary between the models. The degrees of freedom is therefore 17, and the critical value at the 5% level is 27.59. Therefore the null hypothesis is again rejected and it can be concluded that the two models are different, even after rescaling[12].

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. While implicit prices are generally not appropriate to use as 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 are as follows:

H_0: $\beta(MM)_{i1}/\beta(MM)_{M1} = \beta(GS)_{i2}/\beta(GS)_{M2}$
H_1: $\beta(MM)_{i1}/\beta(MM)_{M1} \neq \beta(GS)_{i2}/\beta(GS)_{M2}$

where β_i is a non-monetary choice set attribute (i= 1,…, n-1), β_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 for the previous tests. This is because standard errors for implicit prices are not directly calculated in the multinomial logit model. It is necessary to use alternative procedures to derive confidence intervals. The Krinsky and Robb (1986) procedure is used in this report. This procedure involves randomly drawing a large number of parameter vectors from a multivariate normal distribution with mean and variance equal to the β vector and a variance-covariance matrix from the estimated multinomial logit 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. To estimate confidence

[11] Swait, Louviere and Williams (1994) and Louviere and Swait (1996) recommend 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.

[12] 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 effect 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.

intervals 500 draws were used. The (100 – α)% confidence interval is then found by ranking the 500 resulting implicit prices and removing the top and bottom α/2% of the implicit prices from the final ranking.

Implicit prices and 95% confidence intervals for the two models are shown in Table 7. For each of the four implicit prices the confidence intervals overlap, indicating a degree of similarity.

Table 7: Implicit prices and confidence intervals for the Gwydir Sydney and Macquarie Marshes models

	Jobs	Area	Breed	Species
Macquarie Marshes	10.7 cents (-2.7, 23.8)	3.4 cents (1.1, 5.7)	\$24.15 (\$15.83, \$33.72)	\$4.27 (\$2.69, \$5.98)
Gwydir Sydney	21.8 cents (5.1, 40.1)	3.9 cents (0.9, 8.4)	\$9.81 (\$2.40, \$17.42)	\$3.21 (\$1.50, \$4.71)
P-values (Poe et al test)	0.142	0.400	0.006	0.198

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[13]. 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 7. 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[14].

Test 4: the equivalence of estimates of compensating surplus across sites

This test is the most important of the four hypotheses, 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. It is possible that even if the overall models and implicit prices are equal, that estimates of compensating surplus will be different, or vice versa. Stated formally this hypothesis 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

[13] To illustrate their argument, Poe et al (1994) use as an example a 90% confidence interval for the mean of an estimate ie X ± 1.6450*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.

[14] The equality of mean estimates were also tested using non-parametric Mann-Whitney tests (see Morrison, Bennett, Blamey and Louviere 1998). 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.

represents the Gwydir Sydney model, and there are i= 1,…,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 report, 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 8). 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 8 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[15] 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.

[15] 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, Parsisson and Crabtree 1998; Hanley, Wright and Adamowicz 1998). Given the statistical equivalence of three out of four implicit prices, a larger number of the estimates of compensating surplus reported above would be expected to be equivalent if the alternative specific constants were excluded. 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.

Table 8: Estimates of compensating surplus

Alternative	Change in attributes	Macquarie Marshes	Gwydir Sydney	P-value (Poe et al)
1	Jobs-no change; Area +150 km^2; Breed +1 year; Species +4	$27.83 ($6.64, $48.55)	$80.96 ($63.27, $101.81)	0.000
2	Jobs-no change; Area +350 km^2; Breed +2 years; Species +13	$97.20 ($79.60, $113.64)	$127.57 ($111.08, $146.06)	0.010
3	Jobs-no change; Area +500 km^2; Breed +3 years; Species +8	$105.13 ($90.27, $121.76)	$127.25 ($106.27, $147.57)	0.056
4	Jobs -20; Area +150 km2; Breed +2 years; Species +8	$66.90 ($51.62, $82.79)	$99.25 ($85.28, $116.62)	0.000
5	Jobs -20; Area +350 km2; Breed +3 years; Species +4	$80.81 ($65.22, $98.40)	$104.11 ($87.21, $120.00)	0.026
6	Jobs -20; Area +500 km2; Breed +1 year; Species +13	$76.03 ($59.07, $92.77)	$119.31 ($100.69, $139.18)	0.000
7	Jobs -100; Area +150 km2; Breed +3 years; Species +13	$103.80 ($85.98, $123.02)	$107.63 ($89.10, $128.82)	0.414
8	Jobs -100; Area +350 km2; Breed +1 year; Species +8	$41.00 ($26.22, $58.01)	$79.86 ($62.67, $97.42)	0.000
9	Jobs -100; Area +500 km2; Breed +2 years; Species +4	$53.20 ($38.63, $67.07)	$82.75 ($64.84, $102.15)	0.004

Note: 95% confidence interval is in brackets

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 indicate that benefit transfer may be valid in certain circumstances. For the two models reported here, it is valid for larger improvements. However, 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.

6.2 Tests of transferability across populations

The results from the Gwydir Wetlands surveys in Moree and Sydney are compared in this section to test the hypothesis that the transfer of non-use value estimates across populations is valid. The same tests that were used in the previous section are relevant for this section. However, because only three of the five design variables are correctly signed and significant in the Moree model, it is self-evident that the hypothesis for the fourth test will be rejected. 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 β_{GM} and β_{GS} are taste parameter vectors corresponding to the Gwydir Moree and Gwydir Sydney datasets.

In Table 5 the log-likelihoods for the Gwydir Wetlands Sydney and Moree models are reported. The log-likelihood of the combined model, which is reported in Table 9, is 1780.09. The test statistic is therefore:

$$-2\ [-1780.09 - (-703.21 + -1041.69)] = 70.38$$

The critical value for this test given 16 degrees of freedom is 26.30 at the 5% level. 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 for the transfer across populations.

Table 9: MNL models estimated using the combined Gwydir Sydney and Gwydir Moree data sets

Variables	Combined model without rescaling	Combined model with rescaling
C_1	1.268*** (0.319)	1.381*** (0.290)
C_2	1.558*** (0.320)	1.665*** (0.291)
CHILD	0.274** (0.126)	0.358*** (0.137)
INCOME	-0.124E-7 (0.72E-5)	-0.106E-5 (0.190E-5)
INCOMEDUMMY	0.517E-1 (0.173)	-0.497E-1 (0.189)
VISIT	0.295*** (0.115)	0.448*** (0.125)
PROGRE	0.776*** (0.133)	0.823*** (0.150)
PRODEV	-1.062*** (0.183)	-1.211*** (0.203)
UNDER	-0.379*** (0.708E-1)	-0.448*** (0.770E-1)
WORK	-0.455*** (0.542E-1)	-0.509*** (0.571E-1)
ONE-OFF	0.270*** (0.534E-1)	0.278*** (0.593E-1)
RATES	-0.122E-1*** (0.852E-3)	-0.133E-1*** (0.943E-3)
JOBS	0.134E-2 (0.995E-3)	0.131E-2 (0.110E-2)
AREA	0.105E-3 (0.250E-3)	0.352E-4 (0.276E-3)
BREED	-0.140*** (0.433E-1)	-0.148*** (0.476E-1)
SPECIES	0.413E-1*** (0.855E-2)	0.448E-1*** (0.944E-2)
Summary statistics		
Log-likelihood	-1780.090	-1777.093
Chi-squared (constants only)	522.386	528.380
ρ^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 β_{MM} and $\beta_{GS}$are taste parameter vectors corresponding to the Macquarie Marshes and Gwydir Sydney datasets, and λ represents the ratio of scale factors.

The log-likelihood for the combined and rescaled model is 1777.09[16] (see Table 9). The test statistic is therefore:

$$-2 [-1777.09 - (-703.21 + -1041.69)] = 64.38$$

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 ar:

H_0: $\beta(GM)_{i1}/\beta(GM)_{M1} = \beta(GS)_{i2}/\beta(GS)_{M2}$
H_1: $\beta(GM)_{i1}/\beta(GM)_{M1} \neq \beta(GS)_{i2}/\beta(GS)_{M2}$

where β_i is a non-monetary choice set attribute (i= 1,…, n-1), β_M is the monetary attribute, GM represents the Gwydir Moree model, GS represents the Gwydir Sydney model

Implicit prices and 95% confidence intervals for the two models are shown in Table 10. The confidence intervals for each of the four implicit prices again overlap. However, the overlapping coefficients for JOBS and AREA are not very meaningful because of the insignificant coefficients for these variables. However, the Poe et al (1994) tests show that only the implicit prices for BREED and SPECIES are equivalent.

Table 10: Implicit prices and confidence intervals for the Gwydir Sydney and Gwydir Moree models

	Jobs	Area	Breed	Species
Gwydir Moree	-7.7 cents* (-42.9, 21.7)	-4.7 cents* (-12.6, 3.5)	\$15.18 (\$2.00, \$29.95)	\$3.86 (\$1.15, \$6.73)
Gwydir Sydney	21.8 cents (5.1, 40.1)	3.9 cents (0.9, 8.4)	\$9.81 (\$2.40, \$17.42)	\$3.21 (\$1.50, \$4.71)
P-value (Poe et al)	0.056	0.030	0.260	0.348

* insignificant coefficients in the Gwydir Moree model

Overall, the results from these three hypothesis tests are generally less positive than for the previous set of hypothesis tests. 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 design variables. The incorrect signs would be expected to cause estimates of implicit prices and compensating surplus to diverge substantially, and hence benefit transfer would be subject to considerable

[16] The ratio of scale parameters was 0.84, indicating that the Gwydir Moree data set has larger variance than the Gwydir Sydney data set.

error. The implication is that transfers between rural and urban populations should be treated with considerable caution.

7 Conclusion

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 objective of this report 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, especially regarding the role of alternative specific constants.

As reported in Section 2, 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 report 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 very exacting. This notion of validity, however, may not win much support from policy makers. 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.

Some evidence to answer this question has been provided in this report. The results from this report suggest that larger errors are likely to be found for transfers across sites when larger 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 in this report 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.

Bibliography

Adamowicz, W., Boxall, P., Williams, M. and Louviere, J. (1998). Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics*, 80: 64-75.

Adamowicz, W., Louviere, J. & Williams, M. (1994). Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management*, 26: 271-292.

Adamowicz, W., Swait, J., Boxall, P., Louviere, J. & Williams, M. (1996). Perceptions versus Objective Measurement of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management*, 32:65-84.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100: 464-477.

Atherton, T.J. and Ben-Akiva, M.E. (1976). Transferability and Updating of Disaggregate Travel Demand Models. *Transportation Research Record*, 610: 12-18.

Bergland, O., Magnussen, K. and Navrud, S. (1995). *Benefit Transfer: Testing for Accuracy and Reliability*. Paper presented at the 6th Annual Conference of the European Association of Environmental and Resource Economists. Umea, Sweden, June 17-20.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M. & Louviere, J. (1996). A Comparison of Stated Preference Methods for Environmental Valuation. *Ecological Economics*, 18: 243-253.

Boyle, K.J. and Bergstrom, J.C. (1992). Benefit Transfer Studies: Myths, Pragmatism, and Idealism. *Water Resources Research*, 28(3): 657-663.

Boyle, K.J., Poe, G.L. and Bergstrom, J.C. (1994). What Do We Know About Groundwater Values: Preliminary Implications from a Meta Analysis of Contingent Valuation Studies. *American Journal of Agricultural Economics*, 76: 1955-1061.

Desvousges, W.H., Naughton, M.C. and Parsons, G.R. (1992). Benefit Transfer: Conceptual Problems in Estimating Water Quality Benefits Using Existing Studies. *Water Resources Research*, 28(3): 675-683.

Foster, V. and Mourato, S. (1998). *Elicitation Format and Part-Whole Bias: Do Dichotomous Choice Contingent Valuation Surveys and Stated Preference Experiments Give the Same Results?.* Paper presented at the First World Congress of Environmental and Resource Economists. Venice, June 25-27.

Hanley, N., MacMillan, D., Wright, R.E., Bullock, C., Simpson, I., Parsisson, D. and Crabtree, B. (1998). Contingent Valuation Versus Choice Experiments: Estimating the Benefits of Environmentally Sensitive Areas in Scotland. *Journal of Agricultural Economics*, 49(1): 1-15.

Hanley, N., Wright, R.E. and Adamowicz, V. (1998). Using Choice Experiments to Value the Environment. *Environmental and Resource Economics*, 11(3,4): 413-418.

Hensher, D.A., and Johnson, L.W (1981). *Applied Discrete Choice Modelling.* John Wiley and Sons, New York.

Kask, S.B. and Shogren, J.F. (1994). Benefit transfer protocol for long-term health risk valuation: A case of surface water contamination. *Water Resources Research,* 30(10): 2813-2823.

Kealy, M.J., Montgomery, M. and Dovidio, J.F. (1990). Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter? *Journal of Environmental Economics and Management*, 19: 244-263.

Kenkel, J.L. (1989). *Introductory Statistics for Management and Economics*. Duxbury Press.

Krinsky, I. and Robb, A.L. (1986). On Approximating the Statistical Properties of Elasticities. *Review of Economics and Statistics*, 68: 715-719.

Lancaster, K. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74: 132-157.

Loomis, J.B. (1990). Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques. *Journal of Environmental Economics and Management*, 18: 78-85.

Loomis, J.B. (1992). The Evolution of a More Rigorous Approach to Benefit Transfer: Benefit Function Transfer. *Water Resources Research*, 28(3): 701-705.

Louviere, J.J. and Swait, J. (1996). Searching for Regularities in Choice Processes, Or The Little Constants That Could. Working Paper, Sydney University**.**

Morrison, M.D., Bennett, J.W. & Blamey, R.K. (1997). *Designing Choice Modelling Surveys Using Focus Groups: Results from the Macquarie Marshes and Gwydir Wetlands Case Studies*. Choice Modelling Research Report No.5. School of Economics and Management, University College, The University of New South Wales.

Morrison, M.D., Bennett, J.W. & Blamey, R.K. (1998). *Valuing Improved Wetland Quality Using Choice Modelling*. Choice Modelling Research Report No.6. School of Economics and Management, University College, The University of New South Wales.

Parsons, G.R. and Kealy, M.J. (1994). Benefits transfer in a random utility model of recreation. *Water Resources Research*, 30(8): 2477-2484.

Poe, G.L., Severance-Lossin, E.K. and Welsch, M.P. (1994). Measuring the Difference (X-Y) of Simulated Distributions: A Convolutions Approach. *American Journal of Agricultural Economics*, 76: 904-915.

Poe, G.L., Welsch, M.P. and Champ, P.A. (1997). Measuring the Difference in Mean Willingness to Pay When Dichotomous Choice Contingent Valuation Responses are Not Independent. *Land Economics*, 73(2): 255-267.

Reiling, S.D., Boyle, K.J., Philips, M.L., and Anderson, M.W. (1990). Temporal Reliability of Contingent Values. Land Economics, 66(2): 128-134.

Smith, V.K. (1992). On Separating Defensible Benefit Transfers From “Smoke and Mirrors”. *Water Resources Research*, 28(3): 685-694.

Smith, V.K. and Osborne, L.L. (1996). Do Contingent Valuation Estimates Pass a “Scope” Test? A Meta-Analysis. *Journal of Environmental Economics and Management*, 31: 287-301.

Swait, J. & Louviere, J. (1993). ‘The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models’. *Journal of Marketing Research*, 30: 305-314.

Swait, J., Louviere, J.J. and Williams, M. (1994). A Sequential Approach to Exploiting the Combined Strengths of SP and RP data: Application to freight shipper choice. *Transportation*, 21: 135-152.

Swallow, S.K., Weaver, T., Opaluch, J.J. and Michelman, T.S. (1994). Heterogeneous Preferences and Aggregation in Environmental Policy Analysis: A Landfill Siting Case. *American Journal of Agricultural Economics*, 76: 431-443.

Tiesl, M.F., Boyle, K.J., McCollum, D.W. and Reiling, S.D. (1995). ‘Test-Retest Reliability of Contingent Valuation with Independent Sample Pretest and Posttest Control Groups’. *American Journal of Agricultural Economics*, 77: 613-619.

Watson, P.L. and Westin, R.B. (1975). Transferability of Disaggregate Mode Choice Models. *Regional Science and Urban Economics*, 5: 227-249.

Previous Research Reports in this Series

Morrison, M.D., Blamey, R.K., Bennett, J.W. & Louviere, J.J. (1996). *A Comparison of Stated Preference Techniques for Estimating Environmental Values*. Choice Modelling Research Report No.1. University College, The University of New South Wales, Canberra.

Morrison, M.D. & Bennett, J.W. (1997). *Water Use Trade-offs in the Macquarie and Gwydir Valleys*. Choice Modelling Research Report No.3. University College, The University of New South Wales, Canberra.

Rolfe, J.C., Blamey, R.K. & Bennett, J.W. (1997). *Remnant Vegetation and Broadscale Tree Clearing in the Desert Uplands Region of Queensland*. Choice Modelling Research Report No.3. University College, The University of New South Wales, Canberra.

Blamey, R.K., Rolfe, J.C., Bennett, J.W. & Morrison, M.D. (1997). *Environmental Choice Modelling: Issues and Qualitative Insights*. Choice Modelling Research Report No.4. University College, The University of New South Wales, Canberra.

Morrison, M.D., Bennett, J.W & Blamey, R.K. (1997), *Designing Choice Modelling Surveys Using Focus Groups: Results from the Macquarie Marshes and Gwydir Wetlands Case Studies,* Choice Modelling Research Report No.5, University College, The University of New South Wales, Canberra.

Morrison, M.D., Bennett, J.W & Blamey, R.K. (1998), *Valuing Improved Wetland Quality Using Choice Modelling*, Choice Modelling Research Report No.6, University College, The University of New South Wales, Canberra

Blamey, R.K., Bennett, J.W., Morrison, M.D., Louviere, J.J. and Rolfe, J.C. (1998), *Attribute Selection in Environmental Choice Modelling Studies: The Effect of Causally Prior Attributes*, Choice Modelling Research Report No. 7, University College, The University of New South Wales, Canberra.