ELSEVIER

Journal of Environmental Management 78 (2006) 183–193

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive

Nick Hanley[a,*], Robert E. Wright[b], Begona Alvarez-Farizo[c]

[a]*Economics Department, University of Stirling, Stirling FK9 4LA, Scotland*
[b]*University of Stirling, Stirling, FK9 4LA, Scotland*
[c]*SIA, Unidad de Economia, 50080 Zaragoza, Spain*

Received 5 January 2004; revised 24 May 2005; accepted 27 May 2005
Available online 22 August 2005

Abstract

The Water Framework Directive is a major regulatory reform of water resources management within the European Union. Integrated catchment management plans must be prepared for all river basins, in order to achieve 'good ecological status' in all EU waters. Ecological status is a broader measure of water quality than the chemical and biological measures that were previously dominant. The Directive calls for a consideration of the economic costs and benefits of improvements to ecological status. In this paper, we use the choice experiment method to estimate the value of improvements in three components of ecological status. Given the high resource cost of valuation studies, benefits transfer methods will be needed in implementing the Directive. We thus also test the ability of choice experiments for benefits transfer across two very similar rivers in the UK.

Keywords: Water quality; Water framework directive; Choice experiments; Benefits transfer

1. Introduction

The Water Framework Directive (2000/60) will bring about major changes in the regulation and management of Europe's water resources. Major changes include:

- a requirement for the preparation of integrated catchment management plans, with remits extending over point and non-point pollution, water abstraction and land use;
- the introduction of an EU-wide target of 'good ecological status' for all surface water and groundwater, except where exemptions for 'heavily-modified' water bodies are granted;
- the introduction of full social cost pricing for water use; and
- the incorporation of estimates of economic costs and benefits in catchment management plans.

How exactly regulators will interpret 'good ecological status' is at present not finalised. However, it is clear that it represents a wider set of parameters than the chemical and biological measures of water quality that have previously dominated EU water quality regulation, such as Biological Oxygen Demand or Ammonia (NH_3) levels. In this paper, we use three indicators of ecological status which ordinary people see as important, but which are also consistent with regulator's expectations about the scientific interpretation of this concept. We take ecological status to be determined by three broad factors: healthy wildlife and plant populations; absence of litter/debris in the river; and river banks in good condition with only natural levels of erosion. Recent assessments for UK waterbodies indicate that a significant fraction of rivers, lochs (lakes), estuaries and coastal waters will require improvements if they are to meet 'good ecological status' (DETR, 1999; Scottish Executive, 2002).

One main focus in this paper is therefore on the values people place on improvements in these three indicators, and thus on the non-market economic benefits of moves towards good ecological status. Whilst benefit estimates do exist for

* Corresponding author. Tel.: +44 1786 466410; fax: +44 1786 467469.
E-mail addresses: n.d.hanley@stir.ac.uk (N. Hanley), r.e.wright@stir.ac.uk (R.E. Wright), balvarez@posta.unizar.es (B. Alvarez-Farizo).

0301-4797/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2005.05.001

implementation of the Water Framework Directive (WFD), these are at present highly incomplete (WRc, 1999; Scottish Executive, 2002). However, we are also interested in the practicalities of environmental management using environmental valuation. Valuation exercises are expensive and time consuming, and regulators are very unlikely to have the time or money to commission original valuation studies for every catchment. Benefits transfer, the process of taking estimates from one context and adjusting and then applying them to another, is therefore likely to be important. Accordingly, we conduct a benefits transfer test across two similar rivers, to see what errors are likely to be experienced if benefits transfer procedures are used as part of implementing the WFD.

In what follows, Section 2 briefly describes the Choice Experiment method of environmental valuation and outlines some current issues in benefits transfer. Section 3 describes the case study rivers and survey design. Section 4 presents results, whilst Section 5 concludes.

2. Methodological approach

2.1. Choice experiments

The methodology we use to estimate the value of improvements in river ecology is Choice Experiments. Choice experiments (CE) are becoming a popular means of environmental valuation (Hanley, Mourato and Wright, 2001; Bennett and Blamey, 2001). Choice experiments are one example of the stated preference approach to environmental valuation, since they involve eliciting responses from individuals in constructed, hypothetical markets, rather than the study of actual behaviour. The Choice Experiment technique is based on random utility theory and the characteristics theory of value: environmental goods are valued in terms of their attributes, by applying probabilistic models to choices between different bundles of attributes. By making one of these attributes a price or cost term, marginal utility estimates can be converted into willingness-to-pay estimates for changes in attribute levels, and welfare estimates obtained for combinations of attribute changes. The decision to use a CE approach here was driven by the desire to estimate values for different component parts, or aspects, of water quality, as interpreted by the WFD. These component parts constitute the attributes in the CE design detailed below.

2.2. Previous studies of river ecology changes using choice experiments

Several authors have previously used CE to estimate the value of improvements in river quality. Adamowicz, Louviere and Williams (1994) studied people involved in water-based recreation in Alberta. They recruited a sample of 1,232 members of the general public, from which a 45% response rate was achieved. The attributes used were landscape terrain, fish size, catch rate, water quality, facilities (eg campsite), distance from home and fish species present. The authors found significant effects on utility from changes in fish size, catch rate, water quality and distance from home.

Burton et al. (2000) studied public preferences for catchment management plans in the Moore Catchment, Australia. This area is subject to problems of salinity, eutrophication and flooding, which are all linked to farming activities. Two populations were surveyed, one in the city of Perth and one in rural towns. The attributes used were area of farmland affected by salinity, area of farmland planted with trees, ecological impacts on off-farm wetlands, risk of major flood, changes in farm income, and annual contribution to management plan. The main findings were that the importance people placed on the cost attribute depended on their attitudes to environmental responsibility. Adverse impacts on wetlands, and losses (but not gains) in farm incomes also had significant impacts on utility.

Heberling et al. (2000) studied the benefits of reducing pollution from acid mine drainage in western and central Pennsylvania. Focus groups helped identify the attributes used in the questionnaire: water quality, miles of river restored, travel time from home to site, easy access points, and household costs. Water quality was measured according to what uses could be made of the stream, and took the levels 'drinkable' and 'fishable' and 'swimmable'. The water quality variables were statistically significant determinants of choice in the majority of models, with costs always being significant.

A closely-related technique to choice experiments is contingent ranking. Georgiou et al. (2000) used contingent ranking to estimate the benefits of water quality improvements in the River Tame in Birmingham. People were asked to rank three combinations of four attributes. These were:

- type of fishing (trout/salmon and good game; some game fish species return; a few game fish species return; fish stocks extinct)
- plants and wildlife (otters survive; increase in number and types of insects and greater numbers of bids; more plants and waterfowl; very limited wildlife)
- boating and swimming (both, boating only, swimming only, neither); and
- cost (extra council taxes): £2.50/month, £1.25/month, £0.42/month, zero)

Responses were used to estimate Willingness to Pay for marginal reductions in Biological Oxygen Demand and total ammonia.

Studies using other stated preference techniques to estimate values of ecological changes in river quality may also be found in the literature, notably contingent valuation studies of improvements in low-flow conditions (Hanley et al., 2003; Garrod and Willis, 1996). However, we do not review them here.

A conclusion from this brief review of existing literature is that no study exists which uses choice experiments to estimate the value of improvements in the concept of ecological status as embodied within the Water Framework Directive. Our study is a first step in this policy-relevant direction. Before detailing our study design, however, it is important to review another aspect of environmental valuation relevant to the Directive: Benefits Transfer.

2.3. Benefits transfer

Valuation studies are extensive and time-consuming. For this reason, the policy community has become increasingly interested in benefits transfer techniques (Bateman et al., 2002). Benefits transfer (BT) is a method for taking value estimates from original studies, and adjusting them for use in some new context. The two main approaches to BT are:

- the transfer of adjusted mean values. Mean Willingness to Pay (WTP) estimates taken from the original study or studies are adjusted to account for differences in the environmental characteristics of the new site/context, and/or for differences in the socio-economic characteristics of the affected population at the new site.
- the transfer of benefit functions. Benefit functions are regression equations which explain variations in WTP and/or preferences across individuals according to variations in socio-economic factors and, in some cases, environmental characteristics. A benefits function can be used to produce estimates of WTP.

In both cases, meta analysis (that is, the quantitative analysis of a collection of past studies) can be used to inform the BT process.

Much academic work has taken place in the past 10 years, testing alternative BT methods, and assessing their accuracy. The academic jury is still 'out' on the validity of BT. Studies by Bergland et al. (1995); Barton (2002) and Rozan (2004) largely reject the validity of benefits transfer, both in terms of the transfer of adjusted mean values and the transfer of benefit functions. Brouwer (2000) surveys seven recent benefits transfer studies and finds that the average transfer error is around 20–40% for means and as high as 225% for benefit function transfers, whilst Ready et al. (2001) find a transfer error of around 40% in a multi-country study on the health benefits of reduced air pollution. Shrestha and Loomis (2001) find an average transfer error of 28% in a meta-analysis model of 131 US recreation studies. As Barton points out, though, even fairly small transfer errors (11–26% in his case) can be rejected using the statistical tests favoured by economists. However, this has not stopped the development of large BT software packages, such as the EVRI package, developed by Environment Canada, for use in policy analysis.

One debate on-going at present is whether more complex BT approaches necessarily do better than simple ones. Barton (op cit) finds a simple adjusted means transfer gets closer to original site values than the transfer of benefit functions. The opposite finding, however, is reported in Desvouges et al. (1998). Finding acceptable benefits transfer methods is essential to the wider use of environmental valuation in policy. However, the standards of accuracy required in academic work may exceed those viewed as tolerable by policy-makers, especially in prioritising or filtering alternative investments in water quality.

3. Study design

3.1. Physical context

We located our choice experiment in the context of improvements to the ecology of the River Wear, in County Durham, England; and the River Clyde, in Central Scotland. These were chosen as broadly representative of the kind of waterbodies in the UK where moderate improvements in water quality are likely to be needed in order to meet Good Ecological Status.

The River Wear catchment extends from Burnhope Moor in the Pennines to the North Sea. Population is concentrated in the eastern half of the area, which includes Durham and Sunderland. Throughout much of the 20th century the lower sections of the river were heavily polluted by industry and mining, but have now recovered and support a migratory fishery. The focus for this study is that part of the River Wear which flows through the city of Durham, and which is graded as 'C' on the Environment Agency's General Quality Assessment scheme (interpreted as 'fair' quality). Existing problems include litter, algal growth and acidity problems due to mine drainage. Problems also exist with loss of bankside vegetation, increased erosion, and a decline in habitat and associated fish and wildlife populations. Within the Wear are many man-made structures built in and across river channels. These have important impacts on the way the river functions, altering flows and gravel movements and hindering migration of fish upstream. In terms of recreational uses the River Wear is important as a coarse and game fishery and also as a centre for other water-based recreation (formal and informal). The river also plays an important role in recreation and tourism.

The River Clyde is approximately 121 km long. During its journey from its source in the Beattock Hills to its tidal estuarine limits in Glasgow, its quality varies greatly. Discussion with regulators (the Scottish Environmental Protection Agency) led to the selection of the Clyde from Lanark to Cambuslang Bridge as the area for study. This mainly urbanised stretch has recreational and tourist attractions, and encompasses areas of great beauty like the

Environmental attributes and levels used in the choice experiment

	"GOOD" LEVEL	"FAIR" LEVEL
Ecology	• Salmon, trout and course fish (e.g. pike) • A wide range of water plants, insects and birds	• Only course fish (e.g. pike) • A poor range of water plants, insects and birds
Aesthetics / Appearance	• No sewage or litter	• Some sewage or litter
River Banks	• Banks with plenty of trees and plants • Only natural erosion	• Banks with few trees and plants • Evidence of accelerated erosion

Fig. 1. Environmental attributes and levels used in the choice experiment.

Falls of Clyde, but also has some of the most problematic stretches in terms of water quality. Most of this section was graded 'B' using the Scottish river classification system, which is equivalent to the 'C' grade for the Wear under the General Quality Assessment classification system (i.e. fair quality, but in need of improvement to reach 'good ecological status').

3.2. Steps in choice experiment design

Focus groups were recruited from local residents living around the two rivers in both case study areas in order to (i) gauge local attitudes to the rivers and to their problems (ii) investigate current uses of the two rivers and (ii) identify the attributes by which the rivers could best be characterised. We also gauged reaction to the idea of the need to pay for improvements in river ecology. As a result of group discussions, backed up by discussion with officers from both the Environment Agency (the regulator in England) and the Scottish Environmental Protection Agency, three river quality attributes were chosen for the CE. These were in-stream ecology, aesthetics/appearance, and bankside conditions, and are shown in detail in Fig. 1. Each attribute was set at one of two levels. The 'fair' level was described in such a way as to be consistent with current conditions on the Rivers Wear and Clyde. The 'good' level was consistent with regulators' expectations as to what will likely constitute good ecological quality status under the Water Framework Directive. Note that none of these attributes are necessarily consistent with what an ecologist would choose in terms of either indicators of the ecological health of a waterbody, or underlying factors driving changes in ecological status: they merely represent the characteristics of 'water quality' as perceived by the general public.

A cost or price attribute was established as higher water rates payments by households to the local sewerage operator, Northumbria Water, for the R Wear sample; and to the local authority (Lanarkshire Council) for the R. Clyde sample[1]. Focus groups generally accepted the idea that improvements had to be paid for, and water rates were viewed as a realistic payment mechanism. The price vector used in the design was {£2, £5, £11, £15, £24}, and was chosen based on previous contingent valuation studies in the UK of river improvements.

Attributes and levels were then assigned into choice sets using a fractional factorial design. Due to the simple nature of the design, blocking of the choice sets (that is, introducing an additional attribute to dis-aggregate choice sets into manageable groups) was not necessary. Each respondent answered 8 choice questions. Each question consisted of a three-way choice: option A and option B,

[1] Sewage treatment is privatised in England but remains a public service in Scotland.

which gave an improvement in at least one attribute for a positive cost; and the zero-cost, zero-improvement status quo. Each choice card showed the attribute levels pictorially; a preceding section of the questionnaire explained the importance of each attribute to overall ecological quality[2]. Options A and B can be thought of as representing the outcomes of alternative catchment management plans for each river, with their associated costs.

Sampling was undertaken with a randomised quota-sampling approach, using in-house surveys by trained market research personnel in the autumn of 2001. We collected 210 responses for each river. Whilst this sample is rather small, it is comparable to others reported in the CE literature (eg Hanley et al., 2002; Bergmann et al., 2005). A larger sample size would, however, lead to lower standard errors and greater confidence in interpretation of our results.

4. Methodology

4.1. Statistical model

The method of Choice Experiments is an application of Lancaster's 'characteristics theory of value' combined with 'random utility theory', and is therefore firmly based in economic theory. Individuals are asked to choose between alternative goods, which are described in terms of their attributes, one of which is price (or some proxy for price). Consider the two alternatives case ($C=2$). The underlying utility function of individual 'i' is of the form:

$$U_{ij} = U(\boldsymbol{X}_j, P_j) \tag{1a}$$

$$U_{ik} = U(\boldsymbol{X}_k, P_k) \tag{1b}$$

where: '$\boldsymbol{X}_j$' and '$\boldsymbol{X}_k$' are vectors of attributes describing alternatives 'j' and 'k' and 'P_j' and 'P_k' are the prices or costs associated with each of the alternatives. Individual 'i' will choose alternative 'j' over alternative 'k', if and only if:

$$U_{ij} > U_{ik}. \tag{2}$$

That is, the total satisfaction received from 'consuming' alternative 'j' exceeds that received from alternative 'k'. If $U_{ij}=U_{ik}$ then the individual is indifferent between the two alternatives. If $U_{ij}=0$ and $U_{ik}=0$, then the individual receives no satisfaction from either alternative.

The utility functions associated with the comparison in Eq. (2) may be partitioned into two components:

$$U_{ij} = V(\boldsymbol{X}_j, P_j) + \varepsilon(\boldsymbol{X}_j, P_j), \tag{3a}$$

$$U_{ik} = V(\boldsymbol{X}_k, P_k) + \varepsilon(\boldsymbol{X}_k, P_k), \tag{3b}$$

where the first term on the right-hand side of each of these expressions is deterministic and observable (sometimes referred to as an indirect utility function) while the second term is random and unobservable. Ben-Akiva and Lerman (1985) attribute this randomness to a variety of factors, including unobserved attributes, unobserved taste variations and measurement errors. Therefore, the probability that individual 'i' will choose alternative 'i' over alternative 'j' is:

$$\mathrm{Prob}_i(j|\boldsymbol{C}) = \mathrm{Prob}(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik}), \tag{4}$$

where '$\boldsymbol{C}$' is the complete set of alternatives (in this case two alternatives, 'j' and 'k') and 'ε_{ij}' and 'ε_{ik}' are error terms.

In order to make Eq. (4) empirically tractable, assumptions must be made regarding the structure of the error terms. The usual assumption is that the errors are Gumbel-distributed and independently and identically distributed. This implies that:

$$\mathrm{Prob}_i(j|\boldsymbol{C}) = \exp(\mu V_{ij})/\Sigma_C \exp(\mu V_{iC}), \tag{5}$$

where 'μ' is a scale parameter which is inversely proportional to the standard deviation of the error distribution. This parameter cannot be separately identified and is therefore typically assumed to be one. This assumption implies a constant error variance and also implies that as $\mu \to \infty$ the model becomes deterministic.

In order to derive an explicit expression for this probability, it is necessary to make an assumption regarding the distribution of the error terms (discussed below). As mentioned above, $V(\cdot)$ is composed of attributes describing each alternative. If $V(\cdot)$ is linear in its arguments and additive with a constant term (θ) then the indirect utility functions are:

$$V_{ij} = \theta_0 + \alpha P_j + \boldsymbol{\beta}' \boldsymbol{X}_{ij}, \tag{6a}$$

$$V_{ik} = \theta_0 + a P_k \boldsymbol{\beta}' \boldsymbol{X}_{ik}, \tag{6b}$$

and Eq. (5) becomes:

$$\mathrm{Prob}_i(j|\boldsymbol{C}) = \exp[\mu(\theta_0 + \alpha P_j + \boldsymbol{\beta}' \boldsymbol{X}_{ij})]/\Sigma_C \exp[\mu(\theta_0 + \alpha P_j + \boldsymbol{\beta}' \boldsymbol{X}_{ij})], \tag{7}$$

In order to derive an explicit expression for this probability, it is necessary to know the distribution of the error terms. A typical assumption is that they are independently and identically distributed with an extreme-value (Weibull) distribution. This distribution for the error term implies that the probability of any particular alternative being chosen as the most preferred can be expressed in terms of the logistic distribution, which results in a specification known as the 'conditional logit model' or (less correctly) the 'multinomial logit model' (McFadden, 1974):

[2] For a copy of the questionnaire, please contact the corresponding author.

$$\mathrm{Prob}(j|\boldsymbol{C}) = \exp(\theta_0 + \alpha P_j + \boldsymbol{\beta}'\boldsymbol{X}_j)/\Sigma_C \exp(\theta_0 + \alpha P_C + \boldsymbol{\beta}'\boldsymbol{X}_C). \quad (8)$$

Eq. (8) can be estimated by conventional maximum likelihood procedures (see Green, 1997). Therefore, standard Likelihood ratio-based tests can be used to test restrictions on the parameters (or group of parameters), to test for differences in parameters across sub-groups (e.g. men versus women), and to evaluate goodness-of-fit (e.g. calculate pseudo R^2-values)

Individual-specific characteristics (shifters) that affect utility, such as income, education, marital status etc., can also be included in this specification. However, since these characteristics do not vary across the alternatives, such variables cannot be entered into Eq. (8) as linear arguments (e.g. an individual's education is the same regardless of whether he/she chooses alternative 'j' or 'k'). Such variables can only be included by interacting them multiplicatively with the attributes or the constant.

Adamowicz, Louviere and Williams (1994) show that estimates of consumers surplus associated with changes in the level of attributes can be easily derived from the estimates of this multi-nomial logit model. This calculation is based on an interpretation of the parameter of the price attribute being equal to the marginal utility of income. For the case of two alternatives, this involves summing the marginal values for each attribute when moving from a lower level of the attribute to some higher level of the attribute (for the case of linear demand). More formally, if '$\boldsymbol{X}$' is composed of '$X_1, X_2, \ldots, X_a$' attributes the implicit price (or willingness-to-pay) associated with any individual attribute, 'a' is:

$$p_\mathrm{a} = -\beta_a/\alpha, \quad (9)$$

where from 'α' is the parameter estimate of the price variable 'P' and 'β_a' is the parameter estimate of the specific attribute 'X_a'. Standard errors and confidence intervals can also be calculated for these implicit prices, although there is still considerable discussion relating to what is the most appropriate method to use (Poe et al., 1994, 1997).

4.2. IIA

An important implication of this specification is that selections from the choice set must obey the 'independence from irrelevant alternatives' (IIA) property (or Luce's Choice Axiom; see Luce, 1959). This property states that the relative probabilities of two options being selected are unaffected by the introduction or removal of other alternatives. This property follows from the independence of the error terms across the different options contained in the choice set. If a violation of the IIA hypothesis is observed, then more complex statistical models are necessary that relax some of the assumptions used. These include the multinomial probit (Hausman and Wise, 1978), the nested logit (Wiseman and Koppelman, 1993), the random parameters logit model (Meijer and Rouwendal, 2000; Revelt and Train, 1998; Train, 1998; Train, 2003; Wedel and Kamakura, 2000) and the heterogeneous extreme value logit (Allenby and Ginter, 1995). There are numerous formal statistical tests than can be used to test for violations of the IIA assumption, with the test developed by Hausman and McFadden (1984) being the most widely used.

4.3. Zero-bids and status quo responses

In most CVM studies a significant proportion of respondents usually report 'zero bids'. Likewise in CE studies it is often the case that a significant proportion of respondents select the 'status quo' option. In this sense, status quo responses are analogous to zero bids. In both cases, this implies that they are not willing-to-pay for the changes specified in the design. Zero bids and status quo responses may be categorised into three types. The first are 'genuine zero bids', where the respondent indicates that they not willing to pay anything because they do not value it in a utility sense. The second are 'protest bids', where the respondent reports a zero bid for reasons other than the respondent placing a zero value on the good in question. For example, the respondent disapproves of the principle of paying for environmental protection since they believe it should be required by law. The third are 'don't know' responses, where the respondent is simply uncertain about the amount they are willing-to-pay, noting that this amount could of course be zero.

Zero bids and status quo responses do not necessarily mean that an individual is unwilling to pay anything. It is likely the case that many of the respondents who report that at not willing-to-pay anything in hypothetical questioning, would actually pay something if they were required to do so 'in reality'. In CVM studies individuals who report zero bids are often excluded from the modelling. That is, the analysis is restricted to only individuals who report positive bids or only to only individuals who report positive bids and genuine zero bids, with individuals who report protest bids or who give 'don't know' responses being excluded (such information can be obtained with ancillary questions).

These empirical strategies are problematic since the samples used may be 'self-selected.' Individuals who report positive and genuine zero bids may be very different to those individuals who report protest bids or 'don't know'. If this is the case, they it is likely the case that any regression-based modelling aimed at evaluating the impact of factors thought to impact on willingness-to-pay may be biased. Alvarez-Farizo et al. (1999) have developed a statistical framework aimed at addressing this form of 'sample selection bias' in the estimation of CVM bid curves. However, the authors are unaware of any research that has tried to apply a similar framework in CE. Such an extension should be possible in principle, although the econometrics required to estimate a model along these lines are not straightforward. No attempt has been made to do this in this paper. Therefore, the reader should keep in mind this potential weakness in the modelling that follows.

4.4. Benefit transfer and pooling

An important concept in environmental economics is the notion of 'benefit transfer'. In a nutshell, benefit transfer is the ability to take the results from one 'study site' and apply them to other 'policy sites'. That is, being able to construct estimates of willingness-to-pay applicable based on set of parameters that applicable to a wide range of sites. In terms the substantive problem considered in this paper, it means being able to take the estimates of willingness-to-pay obtained from the River Wear and apply them to the River Clyde (or vice-versa). The time and cost advantages of being able to do this are obvious.

From a statistical point of view, the assessment of benefit transfer concerns testing for the equality of parameters and willingness-to-pay values 'across equations'. Such tests are straightforward to carryout. In terms of the multi-nominal logit model, the test of the equality of parameters across models is the maximum likelihood extension of the 'Chow test for a structural break' (Chow, 1960). The test of equality of willingness-to-pay estimates across models is an application of the 'Wald test for non-linear restrictions' (Wald, 1939, 1943). These tests are described further below and an excellent discussion of the technical details can be found in Greene (2003). More generally, applicability of the principle of benefit transfer is based on the extent to which data from different samples can be pooled. These tests provide evidence about whether pooling is 'statistically acceptable' which therefore provides the researcher with 'hard evidence' on whether benefit transfer is advisable.

5. Results[3]

Table 1 reports the results of the Hausman test for IIA. This test was carried out on a pooled sample of both survey sites ('Both Rivers') and individually for each survey site ('River Wear' and 'River Clyde'). In all three cases the acceptance of IIA was firmly rejected with the Hausman statistic being very large and statistically significant well below the one per cent level. This suggests that estimating the model as a multi-nomial logit could generate misleading results.

As mentioned above, there are various models that can be used to estimate a CE model in the presence of IIA. The approach that we follow here is the random parameters logit model, which is becoming increasingly popular in applied research. One limitation of the mult-nomial logit model is that it assumes that preferences are homogenous and only one parameter is estimated for each attribute. However, it is likely that preferences differ across individuals. In addition, an individual's preference should not vary across the eight choice set questions they were asked in the survey. The random parameters logit model allows for such variation in preferences across individuals and adjusts for error correlation across the choices made by each individual. Application requires assumptions being made about the distribution of preferences. It is assumed that preferences relating to the three attributes are heterogeneous and follow a normal distribution while preferences towards price are assumed to be homogenous. Therefore, separate parameters are estimated for each individual for each of the three attributes along with a single parameter for all for price.

Table 1
Hausman test for IIA

(1)	(2)	(3)
Sample:	Statistic	Significance level
Both rivers	124.7	$P<0.01$
River wear	104.6	$P<0.01$
River clyde	35.5	$P<0.01$

Before turning to a discussion of the results, it is worth noting that in the River Wear sample, 23.8 per cent of respondents selected the 'status quo' while in the River Clyde sample 27.1 per cent did so. However, only a small number of individuals selected the 'status quo' for all eight of the choice set questions they were asked. These percentages are lower that what is often the case in CVM studies but are not small, so it is unclear whether the results may be biased because of protest bidding, 'don't know' responding, etc.

The estimates presented below are based on models that do not include any covariates. In a set of models not reported here, a series of respondent-specific control variables were included in the specification. These variables were: Whether the respondent has ever visited the site; household income; amount of water bill, the respondent's age; and whether the respondent has children. Inclusion of these variables did not have much impact on the estimates so the discussion below is based on the 'simplest' model that includes the three attributes (a_1, a_2 and a_3) price (p) and alternative specific-constants [$\alpha(A)$ and $\alpha(B)$]. These constants can be thought of as representing all other determinants of utility for each option not captured by the attributes.

Table 2 presents the estimates of the model. Turning first to the multi-nomial logit estimates [Columns (3), (5) and (7)], in all three samples the three attributes have the expected positive signs and all are statistically significant below the one percent level. Likewise, in all three samples, price has the expected negative sign. However, price is not statistically significant at even the generous ten per cent level in the River

[3] The statistical package LIMDEP with NLOGIT (Version 3) was used to estimate the models and perform the statistical tests (Greene, 2003).

Table 2
Model estimates

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sample:	Both Rivers		River Wear		River Clyde	
Estimator		Logit	RP Logit	Logit	RP Logit	Logit	RP Logit
a_1	River Ecology	0.281^{***} [0.034]	0.336^{***} [0.048]	0.298^{***} [0.048]	0.346^{***} [0.066]	0.267^{***} [0.048]	0.333^{***} [0.069]
a_2	Aesthetics	0.236^{***} [0.36]	0.290^{***} [0.048]	0.294^{***} [0.053]	0.343^{***} [0.071]	0.188^{***} [0.050]	0.246^{***} [0.067]
a_3	Banksides	0.300^{***} [0.025]	0.362^{***} [0.040]	0.308^{***} [0.036]	0.340^{***} [0.062]	0.298^{***} [0.036]	0.370^{***} [0.056]
p	Price	-0.028^{***} [0.005]	-0.037^{***} [0.007]	-0.048^{***} [0.007]	-0.057^{***} [0.011]	-0.089 [0.065]	-0.172^{**} [0.009]
$\alpha(A)$	Constant for Option A	0.719^{***} [0.063]	0.806^{***} [0.089]	1.022^{***} [0.093]	1.108^{***} [0.138]	0.430^{***} [0.08]	0.499^{***} [0.118]
$\alpha(B)$	Constant for Option B	0.509^{***} [0.074]	0.581^{***} [0.093]	0.0822^{***} [0.106]	0.884^{***} [0.144]	0.203^{***} [0.104]	0.271^{***} [0.131]
Standard deviations of parameters							
$\sigma(a_1)$	River Ecology	–	0.501^{***} [0.173]	–	0.512^{**} [0.241]	–	0.494^{**} [0.249]
$\sigma(a_2)$	Aesthetics	–	0.682 [0.191]	–	0.533 [0.328]	–	0.793^{***} [0.252]
$\sigma(a_3)$	Banksides	–	0.008 [0.34]	–	0.075 [0.96]	–	0.028 [0.356]
$-2\ln L$		7,021	7,014	3,472	3,470	3,523	3,517
N		420	420	210	210	210	210

(1) Standard error in parentheses; (2) '*' = statistically significant at the 10 per cent level; '**' = 5 per cent level; and '***' = 1 per cent level.

Clyde sample [Column (7)]. Turning next to the random parameters logit estimates [Columns (4), (6) and (8)], in all three samples the three attributes have the expected positive signs and all are statistically significant below the one percent level. Therefore, with respect to the attributes both estimators are generating similar results. It is important to note that in the River Clyde sample price is now statistically significant at the five per cent level. These estimates confirm that people 'value' and are prepared to pay for water quality improvements and such improvements are valued 'even more' the lower the cost associated with obtaining them.

Also shown in Table 2 are the standard deviations and standard errors for the parameters of the random parameters logit estimates. It is interesting to note that the standard deviation for the 'river ecology' attribute is statistically significant at the five per cent level or lower in all three samples. The standard deviation for the 'aesthetics' attribute is only statistically significant (below the one percent level) in the River Clyde sample. The standard deviation of the 'banksides' attributes is not statistically significant in any of the samples. These results suggest two things relating to preferences. The first is that the major component of preference heterogeneity is preferences towards 'river ecology'. The second is that preference heterogeneity in the River Clyde sample compared to the River Wear sample is 'larger'. Put slightly differently, preferences appear to be more homogenous amongst River Wear respondents. Of course the key question that needs to be answered is why do preferences differ so much across these sites?

It is clearly the case that if the parameters were identical in numeric values for the River Wear and River Clyde samples then it would be acceptable to estimate the model on the two samples pooled together and apply the results to both sites. That is benefit transfer would be applicable. As mentioned above the degree of similarity between the parameters can be formally tested by the maximum likelihood analogue of the Chow test. More formally, this test is a test of the difference in the parameters across the two samples:

$$\beta'\text{s(Wear)} - \beta'\text{s(Clyde)} = 0 \tag{10}$$

The results of this test for both the multi-nomial logit and random parameters logit models are shown in Table 4. In both cases, the equality of parameters is firmly rejected. The Chi-square values are large and are statistically significant well below the one per cent level. This suggests that the structures of the choice models—which are representative of underlying indirect utility functions—are significantly different from each other for the two samples. More generally it suggests that benefit transfer is not advisable.

Table 4 reports the implicit prices, along with their standard errors, obtained by applying Eq. (9). These values are the amount of money individuals are willing-to- pay for the specified improvement given in the table. Most of these prices are statistically significant below the one percent. It is important to note that the multinomial logit and random parameter logit models generate a set of implicit prices that are very similar for the River Wear sample [Columns (5) and (6)]. This suggests that preference heterogeneity is likely not a factor of much importance and the prices are robust. For the

Table 3
Likelihood ratio test of parameter equality

(1)	(2)	(3)	(4)
Logit		Random parameters logit	
$\chi 2$ value	Significance level	$\chi 2$ value	Significance level
27.1	$P \leq 0.01$	27.8	$P \leq 0.01$

(1) Test is: $\beta'\text{s(Wear)} - \beta'\text{s(Clyde)} = 0$.

Table 4
Willingness-to-pay (implicit prices) estimates for improvements in water quality

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:		Both Rivers		River Wear		River Clyde	
Attribute	Improvement	Logit	RP Logit	Logit	RP Logit	Logit	RP Logit
River Ecology	From 'fair' to 'good'	£20.17*** [3.03]	£18.19** [2.75]	£12.54*** [1.92]	£12.19** [1.99]	£60.08 [40.13]	£38.70** [16.95]
Aesthetics	From 'fair' to 'good'	£16.91*** [3.00]	£15.68*** [2.62]	£12.35*** [2.15]	£12.07*** [2.09]	£42.38 [29.29]	£28.57** [13.05]
Banksides	From 'fair' to 'good'	£21.53*** [3.48]	£19.57*** [2.83]	£12.92*** [1.92]	£12.67*** [1.85]	£67.08 [47.83]	£42.99** [19.49]

(1) Standard error in parentheses; (2) '*' = statistically significant at the 10 per cent level; '**' = 5 per cent level; and '***' = 1 per cent level.

River Clyde sample, the multi-nomial logit model generates prices that are not statistically significant to zero [Column (7)]. However, the random parameters logit model gives prices for the River Clyde sample that are statistically significant at the five per cent level [Column (8)]. The fact that these prices are not significant in the multi-nomial logit model but are significant in the random parameters logit model demonstrates the potential importance of controlling for preference heterogeneity in choice experiments.

Ignoring statistical significance for the moment, the point estimates in Table 3 suggest that willingness-to-pay is higher in the River Clyde sample compared to the River Wear sample. This is the case for both estimators. It is also interesting to note that for the River Clyde sample the random parameters logit model gives prices that are about two-thirds of those given by the multi-nomial logit model. In the River Clyde sample, controlling for preference heterogeneity 'pushes' the prices closer to those found in the River Wear sample. In other words, controlling for preference heterogeneity appears to increase the probability that benefit transfer is advisable (Table 4).

In order to explore the notion of benefit transfer more formally, Table 5 shows the results of the Wald test aimed at evaluating whether the implicit prices are the same across the two models. More formally, this test is concerned with the difference in the estimated prices:

$$p_j(\text{Wear}) - p_j(\text{Clyde}) = 0 \qquad (11)$$

This test is carried out for the prices generated by both the multi-nomial logit and random parameters logit models. In all cases, equality of prices is firmly rejected. The Chi-square values are all large and all are statistically significant well below the one per cent level.

Table 5
Wald test for willingness-to-pay (implicit price) equality

(1)	(2)	(3)	(5)	(6)
Estimator	Logit		Random parameters logit	
Price	Wald $\chi 2$ statistic	Significance level	Wald $\chi 2$ statistic	Significance level
River Ecology (p_1)	616.6	$P \leq 0.01$	172.2	$P \leq 0.01$
Aesthetics (p_2)	194.4	$P \leq 0.01$	61.4	$P \leq 0.01$
Banksides (p_3)	795.9	$P \leq 0.01$	234.1	$P \leq 0.01$

Test is: $p_j(\text{Wear}) - p_j(\text{Clyde}) = 0$.

6. Conclusions

In this paper, we were interested in seeing (1) what values people place on improvements to watercourses such as are envisaged under the Water Framework Directive and (2) whether choice experiments provide encouraging evidence for benefits transfer in this context. With regard to the former point, three attributes were selected to represent the concept of 'good ecological status' under the Directive: river ecology, which represents aquatic life including fish, plants and invertebrates; aesthetics, which represent the amount of litter in the river; and the quality of banksides both in terms of vegetation and in terms of erosion. For the River Wear, we found that people place insignificantly different values on these three aspects of the quality of rivers. One possible interpretation of this is that all three are seen as equally valid indicators of a 'healthy river', which is all people really care about. Another is that the amount of information provided to respondents was insufficient for them to distinguish between the three attributes. Given this finding, it would have in fact been more straightforward to use Contingent Valuation to value the change from fair to good water quality: however, this was not something the researchers could know prior to undertaking the CE. For the River Clyde, larger differences were found in attribute values, with aesthetic improvements being valued appreciably lower than either river ecology or bankside conditions.

The second purpose of this paper was to carry out tests of benefits transfer. This was thought to be important, since the Water Framework Directive will impose a considerable burden on regulators to compare the costs and benefits of river basin management plans. Finding acceptably-accurate means of benefits transfer will be a vital component of this task. We used an identical survey instrument to value identical improvements on two rivers which are both classified as being of 'fair' quality currently. However, both benefits transfer tests were rejected here: preferences and values differ significantly across the two samples. This is a similar finding to that reached by Morrison et al. (2002), who largely reject transferability of values and preferences in a choice experiment study of wetlands conservation in Australia. We found that people living near the Clyde valued improvements to their local river more highly than people in Durham valued identical improvements to their local river; despite the fact that the former sample was lower income than the latter. This is surprising to the extent that the demand for environmental quality is typically assumed to increase with income—although the elasticity of WTP with respect to income is less than one, implying that poorer groups are willing to give up higher fractions of their income for environmental improvements than richer groups (Kristrom and Reira, 1996; Hokby and Soderqvist, 2003). This is what we find: people living near the Clyde appear willing to exchange a larger fraction of their income for local environmental improvements than better-off people living near the Wear. Other possible reasons why those living near the Clyde might value improvements more highly than those living near the Wear include differences in quality of nearby rivers (substitute sites), differences between the two rivers in terms of their natural characteristics (eg hydrology, scale), differences in cultural attitudes to the two rivers, and different uses to which the two rivers are currently put.

Finally, work clearly needs to progress on finding acceptable methods of benefits transfer for water quality improvements under the Water Framework Directive, since it is hard to see how it can be fully implemented in Europe without such a benefits transfer system being set up. Choice experiments do seem promising in this regard, since they can incorporate variations in both environmental quality and socio-economic characteristics across sites, which would seem a priori to be the biggest drivers of differences in value. The present study shows that simple choice experiments may not be capable of delivering such benefit transfers within conventional limits of statistical significance. However, it may well be that policy-makers will view much lower levels of accuracy as acceptable in practice. The question is: how close is close enough?

Acknowledgements

This paper is based on data collected during a study for the UK Department of the Environment, Food and Rural Affairs. We thank Maggie Dewar for research assistance.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining stated and revealed preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26, 271–292.

Allenby, G., Ginter, J., 1995. The effects of in-store displays and feature advertising on consideration sets. International Journal of Research in Marketing 12, 67–80.

Alvarez-Farizo, B., Hanley, N., Wright, R.E., Macmillan, D., 1999. Estimating the benefits of agri-environmental policy: econometric issues in open-ended contingent valuation studies. Journal of Environmental Planning and Management 42, 23–43.

Barton, D., 2002. The transferability of benefits transfer. Ecological Economics 42, 147–164.

Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R., Swanson, J., 2002. Economic Valuation with Stated Preference Techniques: a Manual. Edward Elgar, Cheltenham.

Bennett, J. and Blamey R., 2001. The Choice Modelling Approach to Non-Market Valuation, Cheltenham UK and Northampton, MA, USA: Edward Elgar.

Bergland, O., Magnussen, K., Navrud, S., 1995. Benefit Transfer: Testing for Accuracy and Reliability. Discussion Paper #D-03/1995. Department of Economics and Social Sciences, The Agricultural University of Norway.

Bergmann, A., Hanley, N., Wright, R., 2005. Valuing the attributes of renewable energy investments. Energy Policy forthcoming.

Brouwer, R., 2000. Environmental value transfer: state of the art and future prospects. Ecological Economics 32 (1), 137–152.

Burton, M., Marsh, S., Patterson, J., 2000. Community Attitudes towards Water Management in the Moore Catchment. Paper to Agricultural Economics Society conference, Manchester.

Chow, G., 1960. Tests of equality between sets of coefficients in two tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.

Desvouges, W.H., Johnson, F.R., Banzhaf, H.S., 1998. Environmental Policy Analysis with Limited Information. Edward Elgar, Cheltenham.

DETR, 1999. Prospective costs and benefits of implementing the proposed EU water resources framework directive. Report to DETR, London, prepared by WRc.

Garrod, G., Willis, K., 1996. Estimating the benefits of environmental enhancement: a case study of the River Darrent. Journal of Environmental Planning and Management 39 (2), 189–203.

Georgiou, S., Bateman, I., Cole, M., Hadley, D., 2000. Contingent ranking and valuation of water quality improvements. CSERGE discussion paper 2000-18, University of East Anglia.

Greene, W., 1997. Econometric Analysis, New York: Macmillan.

Greene, W.H., 2002. NLOGIT (Version 3), Econometric Software Inc. Plainview, New York.

Greene, W.H., 2003. Econometric Analysis, fifth ed. Prentice Hall, New York.

Hanley, N., Mourato, S., Wright, R., 2001. Choice modelling approaches: a superior alternative for environmental valuation? Journal of Economic Surveys 15, 3.

Hanley, N., Wright, R.E., Koop, G., 2002. Modelling recreation demand using choice experiments: rock climbing in Scotland. Environmental and Resource Economics 22, 449–466.

Hanley, N., Schlapfer, F., Spurgeon, J., 2003. Aggregating the benefits of environmental improvements: distance-decay functions for use and non-use values". Journal of Environmental Management 68, 297–304.

Hausman, J., Wise, D., 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 42, 403–426.

Hausman, J., McFadden, D., 1984. Specification tests for the multinomial logit model. Econometrica 52, 1219–1240.

Heberling, M., Fisher, A., Shortle, J., 2000. How the Number of Choice Sets Affects Responses in Stated Choice Surveys. Cincinnati, US EPA, Cincinnati.

Hokby, S., Soderqvist, T., 2003. Elasticities of demand and willingness to pay for environmental services in Sweden. Environmental and Resource Economics 26 (3), 361–383.

Kristrom, B., Reira, P., 1996. Is the income elasticity of environmental improvements less than one? Environmental and Resource Economics 7, 45–55.

Louviere, J., Henscher D., Swait, J., 2000. Stated Choice Methods: analysis and application. Cambridge: Cambridge University Press.

Louviere, J., 2001. Choice experiments: an overview of concepts and issues. In: Bennett, J., Blamey, R. (Eds.), The Choice Modelling Approach to Environmental Valuation. Edward Elgar, Cheltenham.

Luce, R.D., 1959. Individual Choice Behaviour: a Theoretical Analysis. Wiley, New York.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15, 447–470.

Meijer, E., Rouwendal, J., 2000. Measuring welfare effects in models with random coefficients, SOM-theme F: Interactions between Consumers and Firms. AKF, Copenhagen.

Morrison, M., Bennett, J., Blamey, R., 2002. Choice modelling and tests of benefits transfer. American Journal of Agricultural Economics 84 (1), 161–170.

Poe, G.L., Severance-Lossin, E.K., Welsh, 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., Welsh, M.P, Champ, P.A., 1997. Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. Land Economics 73, 255–267.

Ready, R., Navrud, S., Day, B., Doubourg, R., Machado, F., Mourato, S., Spanninks, F., Rodriquez, M., 2001. Benefits Transfer: are Values Consistent across Countries?. Department of Agricultural Economics, Penn State University.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households choices of appliance efficiency level. Review of Economics and Statistics 80, 647–657.

Rozan, A., 2004. Benefit transfer: a comparison of WTP for air quality between France and Germany. Environmental and Resource Economics 29 (3), 295–306.

Scottish Executive, 2002. Costs and Benefits of Implementation of the EU Water Framework Directive (2000/60) in Scotland. Report to the Scottish executive, Edinburgh by WRc.

Shrestha, R., Loomis, J., 2001. Testing a meta analysis model for benefit transfer in international outdoor recreation. Ecological Economics 39 (1), 67–84.

Train, K., 1998. Recreation demand models with taste differences over people. Land Economics 74, 230–239.

Train, K., 2003. Discrete Choice Models with Simulation. Cambridge University Press, Cambridge.

Wald, A., 1939. Contributions to the theory of statistical estimation and testing hypothesis. Annals of Mathematical Statistics 10, 299–326.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426–482.

Wedel, M., Kamakura, W., 2000. Market Segmentation: Conceptual and Methodological Foundations. Kluwer, Boston.

WRc, 1999. Potential Costs and Benefits of Implementing the Proposed Water Resources Framework Directive. Final report to DETR, London.