AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Testing the stability of welfare estimates in travel cost random utility models of recreation: An application to the Rotorua Lakes, New Zealand

Lena Mkwara[1], Dan Marsh[2] and Riccardo Scarpa[3]

[1]Department of Economics, University of Waikato, New Zealand
lam29@students.waikato.ac.nz

[2]Department of Economics, University of Waikato, New Zealand
dmarsh@waikato.ac.nz

[3]Department of Economics, University of Waikato, New Zealand
ric.scarpa@gmail.com

Contributed paper presented at the 59th AARES Annual Conference,
Rotorua, New Zealand, February 10-13, 2015

Copyright 2015 by Authors names. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Testing the stability of welfare estimates in travel cost random utility models of recreation: An application to the Rotorua Lakes, New Zealand

Lena Mkwara[1], Dan Marsh[2] and Riccardo Scarpa[3]

[1]Department of Economics, University of Waikato, New Zealand
lam29@students.waikato.ac.nz

[2]Department of Economics, University of Waikato, New Zealand
dmarsh@waikato.ac.nz

[3]Department of Economics, University of Waikato, New Zealand
ric.scarpa@gmail.com

Abstract

Assessing the stability of values over time is vital because non-market valuation studies only provide a snapshot of values at a particular point in time. However, policy analysts are often required to extrapolate these values to future scenarios. Studies conducted to explore the stability of values over time report mixed results. A number of factors are said to contribute to changes in values over an extended period of time. Environmental non-market valuation studies testing the stability of values over time have used models that assume scale homogeneity across respondents. In this study we explore the extent to which scale heterogeneity across individuals can contribute to differences in welfare estimates across data sets. The availability of two independent fishing choice data sets for the Rotorua Lakes, collected six years apart, allowed this investigation to be carried out.

Keywords: Trout angling; Travel cost random utility models; stability of welfare estimates; scale heterogeneity

1. Introduction

Much of the growing interest in the stability of values emerged following the introduction of the contingent valuation method (CVM). Temporal stability of values is usually considered to be an indicator of the reliability of a valuation instrument because the values can be reproduced in follow-up experiments (Bliem *et al.*, 2012; Carson *et al.*, 2001; Loomis, 1989; Mitchell & Carson, 1989). Stability of values is also important because stated preference studies only provide a snapshot of values at a particular point in time. On the other hand, policy analysts are often required to extrapolate these values to some future time period (Liebe *et al.*, 2012; Loomis, 1989). Benefit transfer applications, which are often undertaken with a considerable time lag, represent one such scenario.

Interest in the stability of values over time has spanned many fields, including environmental, transportation and health economics. The stability of values is predominantly assessed using a reliability test referred to as a test-retest of the valuation instrument. It involves the repeated administration of the survey to the same subjects or to different samples from the same population over two or more distinct time periods. The time interval may range from a few weeks to several years. A test-retest with a very short time interval is generally not considered to be a true test of reliability because of the high likelihood of carry-over or recall effects (Liebe *et al.*, 2012; Teisl *et al.*, 1995). Some approaches suggest reducing the recall effects by conducting the second test after a sufficiently long time lag, using a different sample, or using an alternative form of valuation question. On the other hand, if the time interval for a test-retest is long there is a high likelihood that respondents' values may actually change. Either way "a reliable […] instrument is the one that reflects the constancy of values when preferences and choice sets do not change, and reflects changes in values when preferences or choice sets have changed" (Teisl *et al.*, 1995, p. 614).

1.1 Temporal stability of environmental values

To gain further insight on the subject, a review of some of the studies conducted with an emphasis on environmental applications is provided in the remainder of this section. A study by Loomis (1989) is one of the earliest applications to test the stability of environmental values over time. The reliability of the CVM was assessed by a test-retest of two target populations, concerning WTP for water quality in Mono Lake in California. In the first survey a sample of California households was used. The retest sample consisted of visitors to Mono Lake contacted on the site. The initial survey was conducted in 1986 and was followed by a retest in 1987, allowing a nine-month interval between the surveys. The estimated WTP values for various water quality levels showed evidence of preference stability between the two periods.

Reiling *et al*. (1990) assessed the stability of estimates of WTP for the control of black flies along a section of the Penobscot River in Maine using household data. Two split samples were used to control for carry-over effects, in which respondents may repeat the responses given in the previous survey. The contingent valuation survey was administered to one half of the sample during the peak black fly season in August and September 1987. The other half of the sample answered the same survey after the black fly season in late October and November 1987. The authors reported similar mean WTP between the two periods. They also noted that there were only six published studies testing the reliability of contingent valuation values, in contrast to a large number of validity studies.

Stevens *et al*. (1994) investigated the temporal stability of existence values for bald eagles in New England over a three year duration, from 1989 to 1992 using the same sample of respondents. The study results showed evidence of stability of WTP values over time.

The study by Cameron (1997) assessed respondents' WTP to improve water quality in the Hawkesbury-Nepean river to a safe level for recreation and watering stock. The same CVM questionnaire was presented to the same group of respondents at yearly intervals from 1993 to 1995. The findings indicated no significant differences in mean WTP over time.

The CVM studies reviewed so far had a relatively short test-retest period of less than 3 years. Using a longer time span, Whitehead & Hoban (1999) used two samples drawn from the same population to test the stability of WTP for an improvement in water pollution and air quality in Gaston County over a five year period. The first survey was administered in 1990 followed by a retest in 1995. It was found that respondents in a retest group had less favourable attitudes towards the environment. After accounting for the change in attitudes, they found that the 1990 and 1995 values were not significantly different from each other.

Similarly, Brouwer & Bateman (2005) compared WTP for flood control and wetland conservation in the Norfolk Broads in the UK across a five-year period (1991 and 1996), and found that WTP estimates changed significantly over time[1]. They also noted that the stability of values over time was mostly reported in CVM studies with a relatively shorter test-retest period, ranging from 2 weeks to 2 years.

More recently, Bliem & Getzner (2012) investigated the stability of WTP bids for river restoration in the Danube National Park in Austria from two identical surveys employed one year apart. The contingent valuation web-based surveys were conducted in November 2007 and December 2008 using two samples with similar socioeconomic characteristics. The study results indicated temporal stability of preferences for river restoration between the two periods.

In contrast, choice experiment applications testing the stability of values in environmental non-market valuation are sparse. The study by Bliem *et al*. (2012) is one of the first choice experiment study to test the stability of values studies in environmental valuation. They assessed the stability of people's preferences for

[1] The CVM survey was applied to the same sample population.

river restoration in Austria using two identical web-based choice experiment surveys that were administered to two independent samples with a one-year lag. The first survey was carried out in 2007 and the second one in 2008. The authors did not find any significant difference in WTP estimates between the two surveys.

Another test-retest choice experiment was carried out by Liebe et al (2012) on landscape externalities of onshore wind power in Central Germany. The survey was presented to the same respondents with a one-year lag. Findings from the study indicate that preferences were fairly stable between the two periods.

Studies investigating the stability of values in the recreational demand literature using revealed preference methods are also limited. Two of these studies are reported here. Bhattacharjee *et al*. (2009) used Kuhn-Tucker demand models to test the stability of households' recreational demand at Iowa lakes. The test-retest survey was carried out in 2002 and 2003 using the same sample of households. They found that the null hypothesis of stability of recreational demand over time could not be rejected.

Parsons & Stefanova (2009) used trip data sets for Delaware residents to beaches in the Mid-Atlantic region collected in 1997 and 2005 to test the stability of recreational preferences over time. Two different samples were used and their study results showed evidence of qualitative stability in consumer preferences over time.

Overall, as noted by Brouwer & Bateman (2005), the stability of values over time is mostly reported in studies with a relatively shorter test-retest period, ranging from 2 weeks to 2 years. In contrast, the stability of environmental values in studies with a test-retest period of five or more years appears to show mixed results. A number of factors can contribute to changes in preferences over an extended period of time, including changes in preferences, choice sets, economic and other social contextual factors (Habib *et al.*, 2013; Teisl *et al.*, 1995).

Additionally, recent empirical evidence from the field of transportation seems to suggest that scale heterogeneity might contribute to differences in mean estimates of WTP across studies. Hensher *et al*. (2011, 2012) compared the value of travel time saving (VTTS) from seven data sets; five Australian and two New Zealand toll road studies conducted between 1999 and 2008. The choice experiment studies were very similar in content and design. Their main objective was to investigate whether there was "greater synergy in the WTP evidence within model form across comparable data sets compared to cross model forms within data sets" (Hensher *et al.*, 2011, p. 1). They found that scale heterogeneity in scaled multinomial logit (S-MNL) and generalized mixed multinomial logit (G-MNL) models appeared to "inordinately contribute more to differences in mean estimates of VTTS across studies" than preference heterogeneity in mixed multinomial logit (MMNL) models (Hensher *et al.*, 2011, 2012).

Precisely, Hensher *et al*. (2011, p.10) reported:

> Empirical evidence seems to suggest that scale heterogeneity appears to exert a greater influence on producing differences in mean estimates of VTTS across studies than does preference heterogeneity (as accounted for in MMNL while ignoring scale heterogeneity). If as it appears, this is the empirical situation, then previous studies that have ignored scale heterogeneity have in effect increased the chance of transferability of VTTS when in fact this is misleading as a consequence of failing to recognise scale heterogeneity in the sampled population.

To the best of our knowledge studies testing the stability of values over time in environmental economics have used models that assume scale homogeneity across respondents. The main question addressed in this study is whether welfare estimates remain stable over time. The extent to which scale heterogeneity can contribute to differences in welfare estimates across data sets is also explored. The work in this study is the first to explore the stability of values over time by using models that account for scale heterogeneity and those that do not. The availability of two independent fishing choice data sets for the Rotorua Lakes, collected six years apart, permit this investigation to be carried out. The methodology used is provided in the subsequent section.

2. Methods

Swait & Louviere (1993) were the first to recognize that parameter estimates in MNL models from different data sets may differ in magnitude due to scale factor differences. Recently, it has been argued that much of the taste heterogeneity typically assumed in MMNL models choice applications can be better described as scale heterogeneity[2] (Louviere, 2001; Louviere & Eagle, 2006; Louviere & Meyer, 2007; Louviere *et al.*, 1999). Typically, the scale and utility weights are confounded and cannot be separately identified unless specific reparameterisations, and hence assumptions, are implemented. This problem is circumvented in logit model estimation by normalising the scale or standard deviation of the idiosyncratic error to a constant. More recently, models that allow for scale heterogeneity to be accounted for at individual level have been developed. Fiebig *et al*. (2009) proposed the estimation of the Generalized Multinomial Logit Model (G-MNL) accounting for both scale and preference heterogeneity using a specific set of assumptions and attendant reparameterisation. The G-MNL is a mixed logit specification that allows for heterogeneity both in error scale and attribute preferences. Greene & Hensher (2010) specify the G-MNL model building on the G-MNL model by Fiebig *et al*. (2009) and mixed logit models by Train (2003). Assuming individual i chooses alternative j in choice situation t, Greene & Hensher (2010, pp. 414-417) specify the G-MNL model as follows starting with the mixed logit model.

[2] In fact they argue that normal mixing distributions used in MMNL models may be seriously mis-specified.

$$Prob(choice_{it} = \ j|x_{itj}, z_i, v_i) = \frac{exp(v_{itj})}{\sum_{j=1}^{J_{it}} exp(v_{itj})} \tag{1}$$

where $v_{itj} = \ \beta_i' x_{itj}$

$\beta_i = \ \beta + \Delta z_i + \Gamma v_i$

x_{itj} = the K attributes of alternative j in choice situation t faced by individual i

z_i= a set of M characteristics of individual i that influence the mean of the taste parameters; and

v_i = a vector of K random variables with zero means and known (usually unit) variances and zero covariances.

The mixed logit formulation above captures both observed heterogeneity, Δz_i and unobserved heterogeneity in preferences, Γv_i. The basic MNL model is derived by assuming $\Delta = 0$ and $\Gamma = 0$.

The G-MNL is obtained by accommodating scale heterogeneity across individuals in the mixed logit model above through random specific constants. The model in equation (1) is modified as follows:

$$\beta_i = \ \sigma_i[\beta + \Delta z_i] + \ [\gamma + \sigma_i(1-\gamma)]\Gamma v_i \tag{2}$$

where σ_i is the individual specific standard deviation of the idiosyncratic error term

$\sigma_i = exp(\overline{\sigma} + \delta' h_i + \ \tau w_i)$

h_i = is a set of M characteristics of individual i and may overlap with z_i,

δ = parameters in the observed heterogeneity in the scale term

w_i = the unobserved heterogeneity which is assumed to be standard normally distributed

$\overline{\sigma}$ = the mean parameter in the variance

τ = the coefficient of the unobserved scale heterogeneity

γ = a weighting parameter that indicates how variance in residual preference heterogeneity varies with scale, with $0 \leq \gamma \leq 1$.

“The weighting parameter, γ, is central to the generalized model. It controls the relative importance of the overall scaling of the utility function, σ_i, versus the scaling of the individual preference weights contained in the diagonal elements of Γ” (Greene & Hensher, 2010, p. 415). If $\gamma = 0$, the G-MNL model reverts to the scaled mixed logit model.

$$\beta_i = \sigma_i[\beta + \Delta z_i + \Gamma v_i] \tag{3}$$

The Scaled MNL model[3] is derived by assuming $\Delta = 0 \; and \; \Gamma = 0$

$$\beta_i = \sigma_i \beta \tag{4}$$

The G-MNL model or any other model forms in equations (3) and (4) above are estimated by maximum simulated likelihood. Fiebig *et al*. (2009) and Greene & Hensher (2010) give a detailed discussion of the complications that arise in model estimation. They note that $\overline{\sigma}$ is not separately identified from τ. To identify the model σ_i is normalized so that $E[\sigma_i^2] = 1$. This is achieved by letting $\overline{\sigma} = -\tau^2/2$ instead of zero. Furthermore, to ensure non-negative values of τ, “the model is fit in terms of λ, where $\tau = exp(\lambda)$ and λ is unrestricted” (Hensher *et al.*, 2011, p. 6).

Greene & Hensher (2010, p. 417) specify the simulated log likelihood function as follows:

$$logL = \sum_{i=1}^{N} log\left\{\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i}\prod_{j=1}^{J_{it}} P\left(j, x_{it,}\beta_{ir}\right)^{y_{itj}}\right\} \tag{5}$$

where r=1,…,R are the draws required for simulation

[3] In the basic MNL model, the standard deviation of the idiosyncratic error term is assumed to be homogenous across the sampled individuals, $\sigma_i = \sigma$; therefore, $\beta_i = \sigma\beta$. It is standard practice to normalize σ to 1, since it is not possible to identify both β and σ.

$$\beta_{ir} = \sigma_{ir}\left[[\beta + \Delta z_i] + [\gamma + \sigma_{ir}(1-\gamma)]\Gamma v_{ir}\right] \tag{6}$$

$$\sigma_{ir} = exp(-\tau^2/2 + \delta' h_i + \tau w_{ir})$$

where v_{ir} and w_{ir} are the simulated draws on v_i and w_i, respectively
y_{itj} equals 1 if individual i chooses alternative j in choice situation t and zero otherwise

The Scaled MNL model is derived by assuming $\Delta = 0$ and $\Gamma = 0$ and accordingly equation (6) reduces to:

$$\beta_{ir} = \sigma_{ir}\beta \tag{7}$$

The probability of individual i choosing alternative j in choice situation t is given by:

$$P(j, x_{it}, \beta_{ir}) = \frac{exp\left(x'_{itj}\beta_{ir}\right)}{\sum_{j=1}^{J_{it}} exp\left(x'_{itj}\beta_{ir}\right)} \tag{8}$$

The G-MNL model also offers a convenient way of reparameterising the model to estimate the taste parameters in WTP space. WTP space models are said to be behaviourally appealing alternative ways of directly obtaining an estimate of WTP over preference space models, where WTP is obtained indirectly as the ratio of the non-monetary attributes to the cost parameter[4]. Recent application of WTP space models include studies by Train & Weeks (2005), Sonnier *et al*. (2007), Scarpa *et*

[4] Estimating models in preference space poses some challenges in panel mixed logit models if taste heterogeneity is assumed for both the cost and non-monetary attributes. This includes obtaining counter-intuitive distributions of WTP values. This can, for example, include the use of the normal and log-normal distribution for the non-monetary and cost attributes, respectively. It is further demonstrated that for most distributions, values of the cost coefficient close to zero may cause the ratio to be very large, causing the WTP distributions to have an excessively long upper tail. The resultant mean and variance may be much higher than otherwise expected (Scarpa *et al.*, 2008; Train & Weeks, 2005).

al. (2008) and Hole & Kolstad (2012). A specific discussion of the advantages that this reparameterisation offers in testing hypotheses on WTP distributions in the estimation stage is provided by Thiene & Scarpa (2009).

Empirical evidence has shown that the G-MNL model is superior to the S-MNL model since it accommodates both scale and preference heterogeneity (Fiebig *et al.*, 2009; Greene & Hensher, 2010). However, the S-MNL model always provides a model fit at least as good as the MNL model, as the latter is a special case of the former.

In this application the S-MNL model is used. The G-MNL is best suited for panel data sets with repeated choice observations. The fishing choice data used in this application is an unbalanced panel data set with a large proportion of anglers reporting visiting the lakes only once over the fishing season. However, this does not mean that such anglers visited the Rotorua lakes only once during the year, but it may simply imply that they were not included in the other sub-samples, since re-sampling was done at two-monthly intervals. The WTP obtained from the S-MNL is compared to that of the MNL models. A detailed description of the data is presented in the following section.

3. Study area and description of data

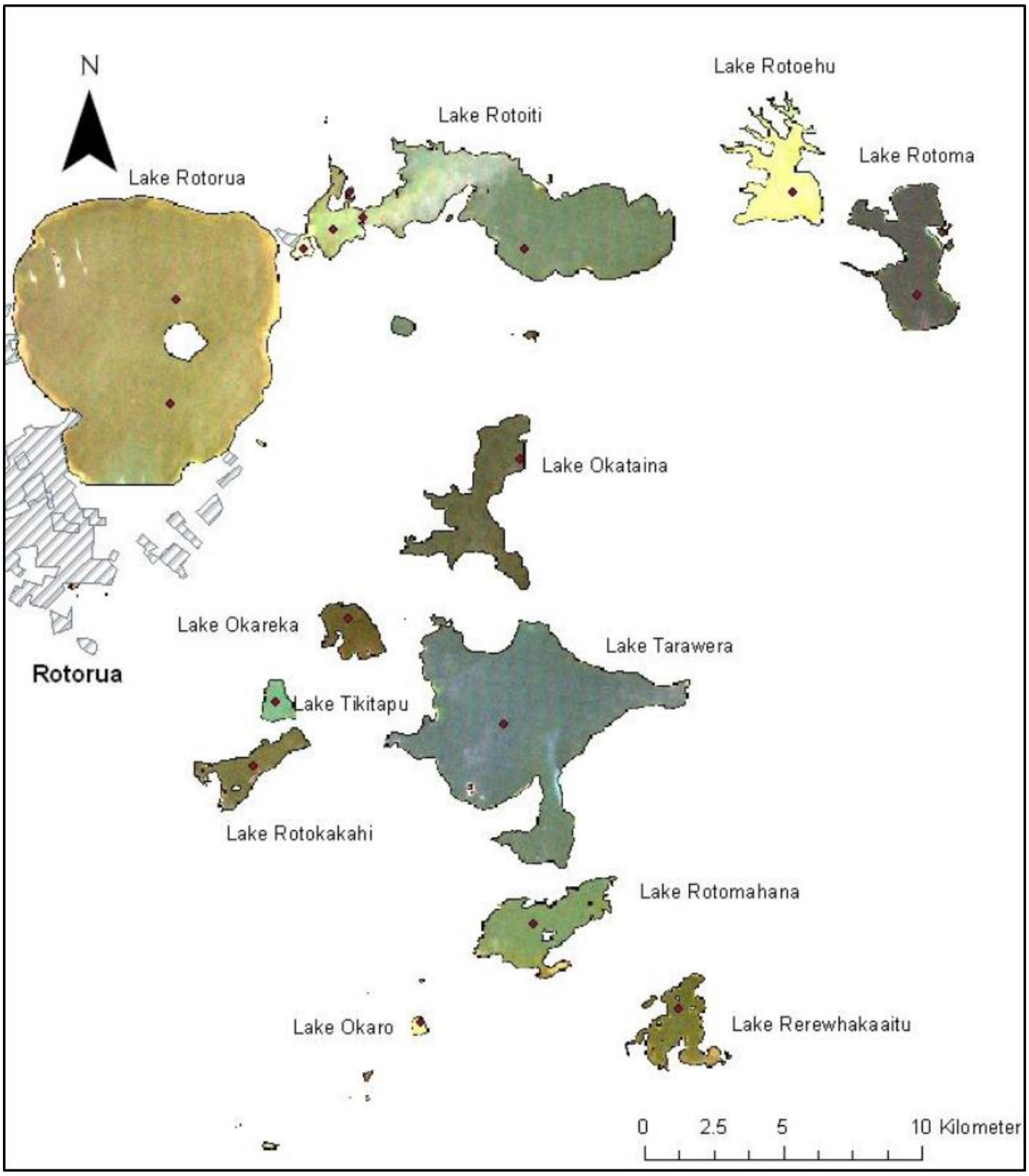
The name Rotorua Lakes refers to twelve main lakes all located in the Rotorua District in Figure 1 below[5]. The Lakes are regarded as having "unique cultural, historical, social and economic value locally, regionally, nationally and internationally[6]". A key element of the recreational value of these lakes is associated with the trout fishery which provides benefits to local residents, visitors, tourists and the local, regional and national economy. Eleven lakes offer a wide

[5] Lake Rotokakahi is not open to the public, therefore the focus is on the remaining eleven lakes.

[6] http://www.hrc.co.nz/human-rights-and-the-treaty-of-waitangi/crown-tangata-whenua-engagement/te-arawa-rotorua-lakes-restoration-programme

range of fishing opportunities. Many of the lakes have a world-class reputation and are within an hour's drive from Rotorua. Rainbow trout are most common in Rotorua's lakes, but there are also brown trout, tiger trout (Lake Rotoma only), and brook trout[7].

Figure 1: The Rotorua Lakes



Source: Allan (2008)

[7] http://eastern.fishandgame.org.nz/

Two fishing choice data sets for these lakes obtained from the New Zealand national angling survey conducted during the 2001/02 and 2007/08 fishing seasons are used in this study. These surveys were carried out jointly by NIWA and FGNZ. The main objectives of the surveys were to obtain consistent estimates of angler usage for all New Zealand lake and river fisheries managed by FGNZ.

Both were telephone sample surveys, based on random samples of anglers drawn from records of fishing licence sales for the angling season, which spans from 1 October to 30 September of each year. Licence holders were asked to identify lakes and rivers they had fished over the previous two months, and the number of days spent on each water. The 2001/02 angling survey was limited to New Zealand residents only, while the 2007/08 survey also included overseas anglers[8]. The surveys were stratified by FGNZ region, time (with the 12 month survey period divided into six two-monthly intervals), and licence type. Licence strata included adult whole-season and family licences, young adult and junior whole season licences and part-season licences (Unwin, 2009; Unwin & Image, 2003).

For the 2001/02 the survey population was limited to the subset of licence holders who were able to be communicated with by telephone. A total of 19,098 licence holders were contacted, of whom 10,847 (56.8%) had fished in at least one of the recognised lake and river fisheries during the two-month survey period of interest (Unwin & Image, 2003).

The 2007/08 consisted of a random sample of 17,739 anglers drawn from a population of 97,215 fishing licence holders. Out of this total, 84,875 were New Zealand resident anglers and 12,340 were overseas anglers (Unwin, 2009).

These surveys did not collect all the information that may be necessary for modelling recreational site choice because such information was not in line with their study objectives. No information was collected on whether the fishing trips

[8] Overseas anglers were contacted by email

undertaken were day trips or involved an overnight stay, or on whether fishing trips were single or multi-purpose. Furthermore, no information is available on whether or not anglers fished in more than one water body during a reported day of fishing. Also missing from the angling survey is information on the amount of time spent fishing on a particular lake[9]. As noted by Phaneuf & Smith (2003), all this information might have implications on how to measure the resources given up in order to access the recreational site.

The angling surveys have been adapted to suit this study in the following ways: The main focus in this application is on single day fishing trips and individual level choice data. To meet these criteria only adult individual licence holders who lived within 240 km of the lakes are included in the sample. This distance measure is considered to be a reasonable benchmark for day trips (McConnell & Strand, 1994; Parsons & Kealy, 1992).

A sample of 524 and 414 anglers fulfilled these criteria for the 2001/02 and 2007/08 fishing seasons, respectively. The total number of fishing days for these samples compared to the total angling days reported in the national angling surveys are presented in Table 1 below. In total 2,200 and 2,292 fishing days were reported for the 2001/02 and 2007/08 samples, respectively.

[9] Information on social economic demographic factors was not collected in the 2001/02 national angling survey. In the 2007/08 national angling survey, only data on age and gender is available for a limited number of anglers.

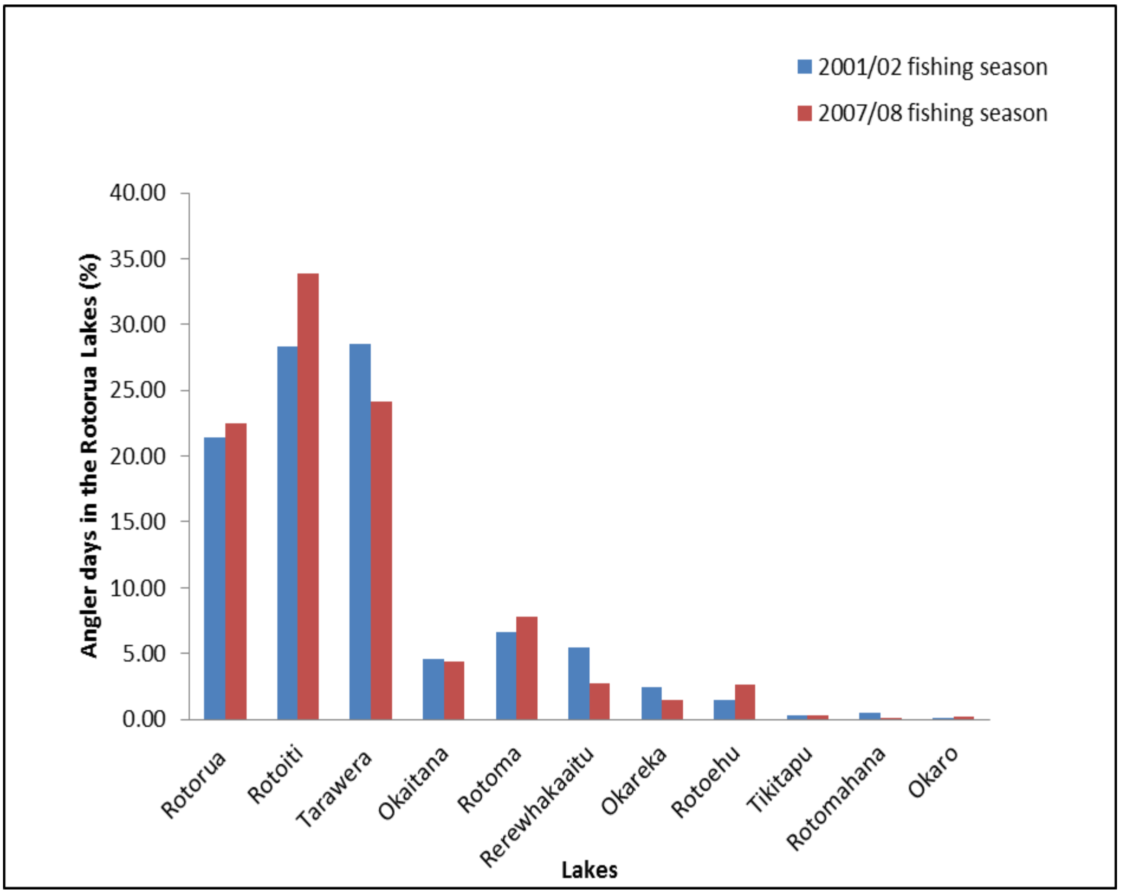
Table 1: Estimated angler days for the 2001/02 and 2007/08 national angling surveys versus samples utilised in this study

Lake Name	2001/02 national angling survey estimated angler-days ± 1 standard error	Number of fishing days for the 2001/02 sample	2007/08 national angling survey estimated angler-days ± 1 standard error	Number of fishing days for the 2007/08 sample
Rotoiti	43080 ± 3120	668	48070 ± 3710	673
Tarawera	43480 ± 2940	863	34220 ± 3440	548
Rotorua	32640 ± 2580	748	32000± 3200	583
Rotoma	10130 ± 1260	76	11110 ± 2040	233
Okaitana	7050 ± 890	192	6290 ± 1070	95
Rerewhakaaitu	8380 ± 1320	169	3830 ± 800	99
Rotoehu	2190 ± 770	52	3720 ± 1210	33
Okareka	3750 ± 1240	82	2040 ± 530	19
Tikitapu	470 ± 190	7	370 ± 140	3
Okaro	200 ± 120	4	260 ± 170	5
Rotomahana	820 ± 380	7	70 ± 50	1
Total		**2200**		**2292**

Source: Unwin & Image (2003) and Unwin (2009)

In Figure 2 below the estimated angler days on the Rotorua Lakes for the 2001/02 and 2007/08 national angling surveys are further compared.

Figure 2: Angler days at each lake as a percentage of the total angling days at the Rotorua Lakes



The distributions of angling days for the 2001/02 and 2007/08 national angling surveys are broadly similar. In both surveys the highest angling days were reported for lakes Rotoiti, Tarawera and Rotorua. These constitute the three major fishing lakes in the Rotorua Lakes.

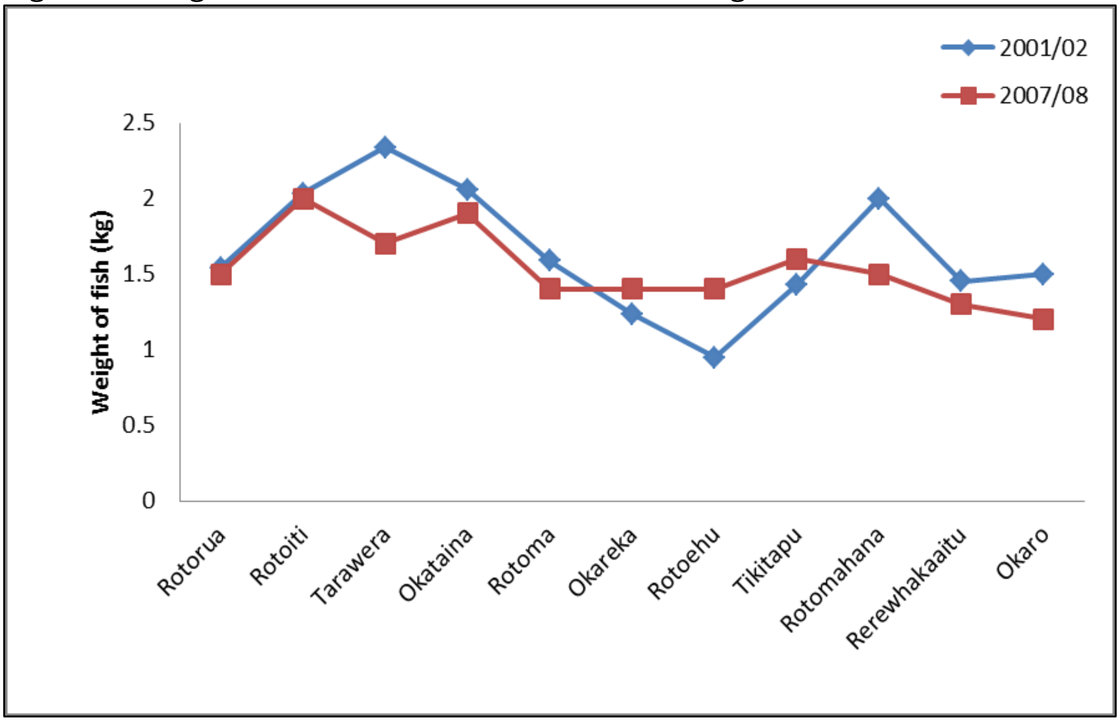
The estimated angler days reported in the national angling survey are used as a benchmark for the true population distribution since the surveys were designed following random sampling procedures. From Table 1 above, there is a clear indication that the samples employed in this application either over-state or under-state the true distribution of fishing days across the lakes. To account for under-sampling and over-sampling, choice-based sampling techniques were used, following procedures outlined by Hensher *et al.* (2005).

Lake attribute data

The lake attribute data used include travel cost, weight of fish, water clarity, size of lake, urban development, facility development, amount of forested land and lake depth. The lake size and depth attributes are invariant across time. Similarly, urban development, facility development and amount of forested land attribute levels were generally constant between the two study periods. On the other hand, there were some slight changes in weight of fish and water clarity for some lakes between the two periods as further elaborated below.

Travel cost includes the cost of fuel expenses only. The opportunity cost of travel time is not accounted for because information on income was not collected in both surveys. Hence the welfare estimates derived are to be considered conservative and a lower bound on the real values. The cost of fuel was estimated at NZ$0.12 and NZ$0.19 per kilometre for the 2001/02 and 207/08 fishing choice data, respectively. The 2001/02 fishing trip costs were recalculated in 2008 New Zealand dollars using the consumer price index. The weight of fish and water clarity for the two study periods are compared in Figure 3.

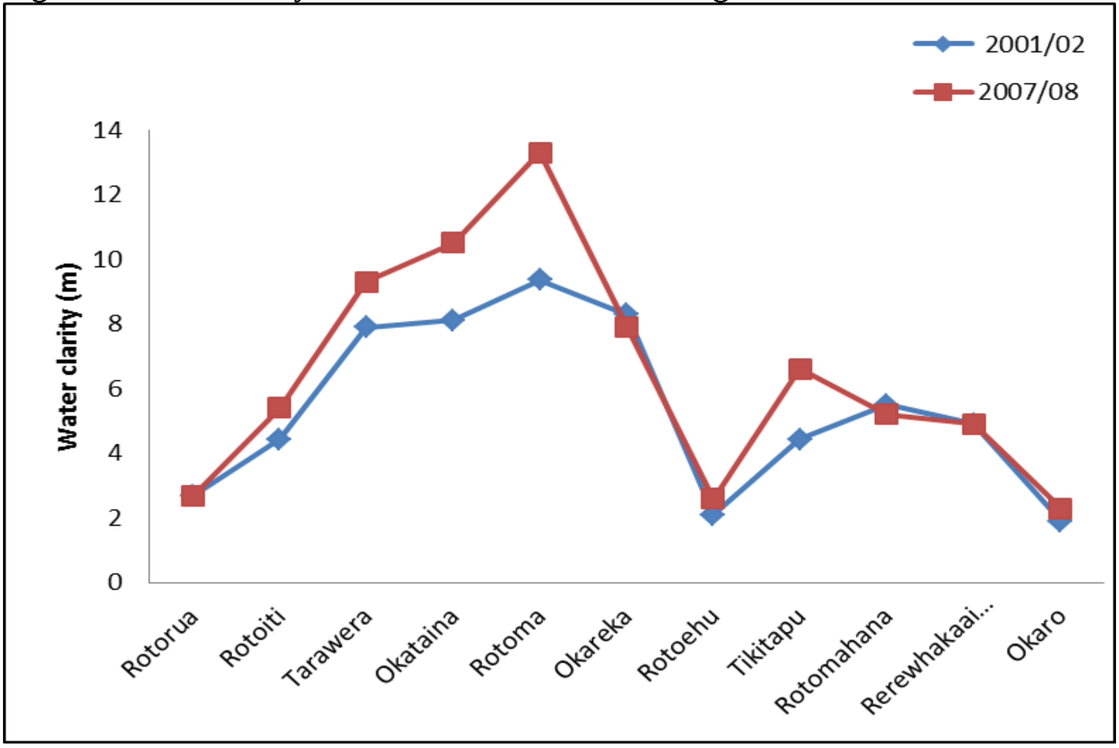
Figure 3: Weight of fish for 2001/02 and 207/08 fishing seasons



Source of data: Eastern Region Fish and Game Region

The annual average weights of fish were generally similar for most lakes, except for Lakes Tarawera, Rotoehu and Rotomahana. For Lake Tarawera the annual average weight of fish was 2.4 kg in the 2001/02 fishing season compared to 1.6 kg during the 2007/08 fishing season. Lake Rotoehu registered an improvement in the average weight of fish from about 1 kg during the 2001/02 fishing season to 1.4 kg in the 2007/08 fishing season. There was a decline in the average weight of fish for Lake Rotomahana from 2 kg in the 2001/02 fishing season to 1.5 kg during the 2007/08 fishing season. The average weight of fish for the 2001/02 fishing season was also compared to that of the previous fishing season (2000/01 fishing season) and was found to be consistent across lakes. Similarly, the average weight of fish for the 2006/07 and 2007/08 fishing seasons was also consistent. In Figure 4 the annual average water clarity during the two survey periods are compared.

Figure 4 Water clarity for 2001/02 and 207/08 fishing seasons



Source of data:(Scholes, 2009)

There was a slight improvement in water clarity for Lakes Rotoiti, Tarawera, Okataina, Rotoma and Tikitapu. For the other lakes water clarity remained stable during the two periods. Improvements in water clarity occurred in lakes which already had good water quality. In general, it is anticipated that an improvement in water clarity in lakes with poorer water quality would be more valued. To account for variability in these attributes between the two study periods in the estimation, year-specific averages of weight of fish and water clarity are used.

The summary statistics for the lake attributes are presented in Table 2 below.

Table 2: Summary statistics of the Rotorua Lakes attributes

Variable	Mean	Std.Dev.	Minimum	Maximum
Weight of fish (kg) (2001/2002 fishing season)	1.65	0.41	0.95	2.43
Weight of fish (kg) (2007/2008 fishing season)	1.54	0.23	1.2	2.0
Water clarity (metres) (2001/2002 fishing season)	5.42	2.65	1.90	9.36
Water clarity (metres) (2007/2008 fishing season)	6.39	3.36	2.3	13.3
Lake size (square km)	18.71	23.31	0.31	80.6
Number of boat ramps[10]	2.27	2.00	1	7
Number of access points	2.36	2.06	0	7
Depth (metres)	29.33	19.68	7	60
Urban development (% of lake catchment area)	1.41	2.27	0	8.1
Amount of forested land (% of lake catchment area)	56.82	26.53	6	94

The estimated results are outlined in the remainder of the sections.

4. Estimated results

The estimated results for the 2001/02 and 2007/08 fishing choice data are presented in Table 3. Parameters are estimated by maximum likelihood using standard routines implemented in Nlogit 4.0.

[10] Boat ramps and number of access points are highly collinear and therefore, boat ramps are used as a proxy for recreational facility development around the lakes.

Table 3: Estimated results for the 2001/02 and 2007/08 fishing seasons

	2001/02 fishing season				2007/08 fishing season			
	Model 1		**Model 2**		**Model 3**		**Model 4**	
	MNL		**S-MNL**		**MNL**		**S-MNL**	
Variable	*Coefficient*	*|t-value|*	*Coefficient*	*|t-value|*	*Coefficient*	*|t-value|*	*Coefficient*	*|t-value|*
Travel cost	-0.166***	22.08	-0.167***	22.12	-0.090***	13.56	-0.121***	9.75
Weight of fish	0.301***	3.02	0.296***	2.97	0.310***	3.88	0.303***	3.0
Water clarity	0.169***	10.82	0.174***	11.16	0.168***	13.69	0.282***	6.11
Lake size	1.978***	16.22	2.033***	16.94	2.946***	14.13	4.533***	5.76
Urban development	-0.282***	13.05	-0.288***	13.31	-0.343***	10.39	-0.509***	4.93
Facility development[11]	0.348***	19.57	0.349***	19.27	0.289***	12.96	0.443***	5.87
Amount of forested land	0.001	0.21	0.001	0.57	0.014***	7.28	0.025***	4.32
Lake depth	-0.035***	9.65	-0.036***	9.87	-0.042***	10.19	-0.070***	4.36
Scale parameter (τ)			0.020	0.51			0.633***	6.19
Summary statistics								
Log-Likelihood	-4641.482		-4638.241		-3830.147		-3824.373	
Mc Fadden R-Squared	0.265		0.271		0.273		0.282	
Number of respondents	524		524		414		414	

***, **, * implies significant at 1%, 5% and 10% level, respectively.

[11] Facility development is the average of boat ramps and number of access points to the lakes because the two attributes were highly correlated and could not enter the utility specification separately.

Models 1 and 2 consist of the estimated results for the 2001/02 fishing season from the MNL and S-MNL models, respectively. The estimated results for the 2007/08 fishing season are presented in Models 3 and 4. In terms of model performance, the S-MNL models perform slightly better than the MNL models in both data sets as indicated by both the log-likelihood and McFadden R-squared.

In all the models the travel cost coefficient is negative and highly significant, implying that anglers preferred lakes that were closer to their home regions. Urban development and lake depth are negative and highly significant in both models. These findings suggest that in general lakes with more urban development and deeper ones were less preferred by anglers. Furthermore, the results show that lakes with bigger fish, better water clarity, larger size and more recreational facilities were generally preferred as indicated by positive and highly significant coefficients for these attributes. On the other hand, the coefficient for the amount of forested land is positive but only significant in Models 3 and 4 (2007/08 fishing season).

The scale parameter (τ) for S-MNL model is only significant in the 2007/08 fishing choice data implying greater individual heterogeneity in the 2007/08 sample than the 2001/02 sample. Comparison of parameter estimates obtained from different samples is impossible without accounting for scale factor differences (Hensher, 2012; Swait & Louviere, 1993). Parameter estimates for the S-MNL models (Model 2 and 4) can be compared. Since the concern in this study is to test the null hypothesis of equality of welfare estimates, the equality of utility weights is of less concern. In the remainder of the paper the equality of welfare estimates are tested.

5. Comparison of welfare estimates

The marginal WTP values measured by the ratio of the non-monetary attributes to the travel cost coefficient are presented in Table 4 below.

Table 4: Comparison of marginal WTP values

	MNL 2001/02 sample			**MNL 2007/08 sample**			**S-MNL 2001/02 sample**			**S-MNL: 2007/08 sample**		
Variable	*MWTP*	*95% Confidence Interval*		*MWTP*	*95% Confidence Interval*		*MWTP*	*95% Confidence Interval*		*MWTP*	*95% Confidence Interval*	
Weight of fish	1.81	[0.64	2.99]	3.43	[1.68	5.18]	1.78	[0.60	2.96]	2.53	[0.94	4.13]
Water clarity	**1.02**	**[0.81**	**1.22]**	**1.86**	**[1.56**	**2.16]**	**1.05**	**[0.84**	**1.25]**	**2.33**	**[1.82**	**2.84]**
Lake size	**11.89**	**[10.03**	**13.74]**	**32.63**	**[26.06**	**39.20]**	**12.24**	**[10.39**	**14.09]**	**37.68**	**[27.82**	**47.54]**
Urban development	1.69	[1.42	1.97]	3.80	[2.95	4.66]	**1.73**	**[1.46**	**2.00]**	**4.23**	**[2.88**	**5.58]**
Facility development	2.09	[1.84	2.35]	3.21	[2.58	3.84]	**2.10**	**[1.84**	**2.36]**	**3.68**	**[2.70**	**4.66]**
Amount of forested land	-	-	-	0.15	[0.11	0.20]	-	-	-	0.21	[0.13	0.29]
Lake depth	0.21	[0.16	0.25]	0.46	[0.35	0.57]	**0.21**	**[0.17**	**0.26]**	**0.58**	**[0.37**	**0.79]**

Figures in [ ] are the 95% confidence intervals

Figures in bold imply significant differences in the mean WTP estimates

The marginal WTP and confidence intervals were estimated by simulating approximate distributions of WTP estimates using the Krinsky–Robb procedure with 5000 draws (Krinsky & Robb, 1986).

The mean WTP obtained from MNL and S-MNL models for the 2001/02 sample are not significantly different from each other based on the non-overlapping confidence interval criteria. Similarly, the mean WTP values from the MNL and S-MNL models for the 2007/08 sample are of the same magnitude. These results seem to be supportive of the findings by Greene & Hensher (2010), who reported that accounting for scale heterogeneity without preference heterogeneity in a single study appeared to have little effect on behavioural outputs such as direct elasticities and WTP.

Comparisons of MNL model estimates across the two data sets indicates similar mean WTP for all attributes, except for water clarity and lake size for the 2001/02 and 2007/08 samples. The higher mean WTP estimate for water clarity in the 2007/08 data set could possibly be attributed to the increased need for better water quality over the years since its marked decline in the early 2000s. One possible factor that could explain the higher preference for bigger lakes in the 2007/08 sample is ease of boat launching. It is conjectured that with the increase in the number of anglers using these lakes over time, boat launches in bigger lakes would be relatively more convenient than in smaller lakes. A number of other unknown factors could explain the higher preference for bigger lakes in the 2007/08 sample.

On the contrary, the mean WTP estimates from the S-MNL models for the 2001/02 and 2007/08 samples are significantly different from each other except for the weight of fish attributes. It appears that accounting for scale heterogeneity significantly contributes to differences in mean WTP across the two data sets. Hensher *et al*. (2011, 2012) reported similar findings. They compared the value of travel time saving (VTTS) from seven choice experiment data sets conducted between 1999 and 2008 and found that accounting for scale heterogeneity inordinately contributes to differences in mean estimates of VTTS across studies. Assumptions about scale homogeneity seem therefore to be crucial in testing for equality of mean WTP estimates and hence for preference stability.

6. Conclusion and implications of the study

The main question addressed in this study was whether welfare estimates remain stable over time. The extent to which scale heterogeneity across individuals can contribute to differences in welfare estimates across data sets was also explored. To achieve this objective, welfare estimates obtained from the multinomial logit models (MNL) and scaled-multinomial logit models (S-MNL) for the 2001/02 and 2007/08 fishing choice data sets were compared.

To assess whether the estimated mean WTP for the lake attributes remained stable between the two periods, results from the MNL and S-MNL models were compared both within the same data set and across data sets. The within same data set comparison showed that the mean WTP estimates from the MNL and S-MNL models were not significantly different from each other in both the 2001/02 and 2007/08 data sets. These results seem to support findings by Greene & Hensher (2010), who reported that accounting for scale heterogeneity without preference heterogeneity in a single study appeared to have little effect on behavioural outputs such as direct elasticities and WTP.

On the other hand, comparison of estimated mean WTP from the MNL models across the 2001/02 and 2007/08 data sets showed evidence of relative stability for all attributes except for water clarity and lake size attributes. However, results from the S-MNL model do not support the stability of estimated mean WTP for all attributes except for the weight of fish attribute. It appears that scale heterogeneity across individuals, as accounted for in the S-MNL model, contributed significantly to differences in MWTP across the two samples. Similar findings are reported by Hensher *et al*. (2011).

To the best of our knowledge studies testing the stability of values over time in environmental economics have used models that assume scale homogeneity across respondents. Findings from this study have demonstrated that ignoring scale heterogeneity across the sampled population may result in the erroneous conclusion that mean WTP estimates are stable over time, when in fact they are not. This calls for a re-examination of previous empirical evidence which has not allowed for scale variability, and suggests the need to systematically account for it in future applications.

References

Allan, M. G. (2008). *Remote sensing of water quality in Rotorua and Waikato Lakes.* (Master of Science in Biological Sciences), University of Waikato.

Bhattacharjee, S., Kling, C. L., & Herriges, J. A. (2009). Kuhn-Tucker estimation of recreation demand – A study of temporal stability. Paper presented at the Agricultural & Applied Economics Association 2009: AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26-29, 2009 http://ageconsearch.umn.edu/bitstream/49408/1/611716AAEA.pdf

Bliem, M., & Getzner, M. (2012). Willingness-to-pay for river restoration: differences across time and scenarios. Environmental Economics and Policy Studies, 14(3), 241-260. doi: 10.1007/s10018-012-0029-3

Bliem, M., Getzner, M., & Rodiga-Laßnig, P. (2012). Temporal stability of individual preferences for river restoration in Austria using a choice experiment. Journal of Environmental Management, 103, 65-73. doi: http://dx.doi.org/10.1016/j.jenvman.2012.02.029

Brouwer, R., & Bateman, I. J. (2005). Temporal stability and transferability of models of willingness to pay for flood control and wetland conservation. Water Resources Research, 41(3), DOI: 10.1029/2004WR003466.

Cameron, J. I. (1997). Applying socio-ecological economics: A case study of contingent valuation and integrated catchment management. Ecological Economics, 23(2), 155-165. doi: http://dx.doi.org/10.1016/S0921-8009(97)00051-7

Carson, R., Flores, N., & Meade, N. (2001). Contingent Valuation: Controversies and Evidence. Environmental and Resource Economics, 19(2), 173-210. doi: 10.1023/A:1011128332243

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2009). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Marketing Science, Articles in Advance, pp. 1–29.

Greene, W., & Hensher, D. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. Transportation, 37(3), 413-428. doi: 10.1007/s11116-010-9259-z

Habib, K. M., Swait, J., & Salem, S. (2013). Using repeated cross-section travel surveys to enhance forecasting robustness: Accounting for changing mode preferences. Retrieved 2 June 2014

Hensher, D. A. (2012). Accounting for scale heterogeneity within and between pooled data sources. Transportation Research Part A: Policy and Practice, 46(3), 480-486. doi: http://dx.doi.org/10.1016/j.tra.2011.11.007

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). Applied choice analysis. Cambridge, UK: Cambridge University Press.

Hensher, D. A., Rose, J. M., & Li, Z. (2011). Does the choice model method and/or the data matter? . (Working Paper ITLS-WP-11-14). Retrieved 23 July 2014, from Institute of Transport and Logistics studies, University of Sydney http://sydney.edu.au/business/__data/assets/pdf_file/0011/104006/ITLS-WP-11-14.pdf

Hensher, D. A., Rose, J. M., & Li, Z. (2012). Does the choice model method and/or the data matter? Transportation Research Board, 39(2), 351-385. doi: http://dx.doi.org/10.1007/s11116-011-9329-x

Hole, A., & Kolstad, J. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. Empirical Economics, 42(2), 445-469. doi: 10.1007/s00181-011-0500-1

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. Review of Economics and Statistics, 68(Nov.), 715-719.

Liebe, U., Meyerhoff, J., & Hartje, V. (2012). Test–Retest Reliability of Choice Experiments in Environmental Valuation. Environmental Resource Economics, 53, 389–407. doi: 10.1007/s10640-012-9567-1

Loomis, J. B. (1989). Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses. American Journal of Agricultural Economics, 71(1), 76-84. doi: 10.2307/1241776

Louviere, J. J. (2001). What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon. Journal of Consumer Research, 28(3), 506-511. doi: 10.1086/323739

Louviere, J. J., & Eagle, T. (2006). Confound it! That pesky little scale constant messes up our convenient assumptions. Proc. 2006 Sawtooth Sofware Conf, 211-228.

Louviere, J. J., & Meyer, R. J. (2007). Formal Choice Models of Informal Choices: What Choice Modeling Research Can (and Can't) Learn from Behavioral Theory. Review of Marketing Research, 4(1), 3-32.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., .Irwin, J. (1999). Combining Sources of preference data for modelling complex decision processes. Mark. Lett., 10(3), 205–217.

McConnell, K. E., & Strand, I. E. (1994). The economic value of Mid and South Atlantic Sportfishing.: Report to the U.S. EPA and NOAA. Department of Agricultural and Resource Economics, University of Maryland.

Mitchell, R. C., & Carson, R. T. (1989). Using surveys to value public goods: The contingent valuation method. Washington, DC: John Hopkins University Press.

Parsons, G. R., & Kealy, M. J. (1992). Randomly drawn opportunity sets in a random utility model of lake recreation. Land Economics, 68(1), 93-106.

Parsons, G. R., & Stefanova, S. (2009). Temporal stability of recreation choices. Paper presented at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26-29.

Phaneuf, D. J., & Smith, K. (2003). Recreational demand models. In M. Karl-Goran & R. V. Jeffrey (Eds.), Handbook of Environmental Economics. North Holland: Elsevier.

Reiling, S. D., Boyle, K. J., Phillips, M. L., & Anderson, M. W. (1990). Temporal Reliability of Contingent Values. Land Economics, 66(2), 128-134. doi: 10.2307/3146362

Scarpa, R., Thiene, M., & Train, K. (2008). Utility in willingness to pay space: A tool to address confounding random scale effects in destintion choice to the Alps. Amer. J. Agric. Econ., 90(4), 994-1010. doi: http://dx.doi.org/10.1111/j.1467-8276.2008.01155.x

Scholes, P. (2009). Rotorua Lakes water quality report: Prepared for Environment Bay of Plenty Regional Council.

Sonnier, G., Ainslie, A., & Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. Quantitative Marketing and Economics, 5(3), 313-331. doi: 10.1007/s11129-007-9024-6

Stevens, T. H., More, T. A., & Glass, R. J. (1994). Interpretation and Temporal Stability of CV Bids for Wildlife Existence: A Panel Study. Land Economics, 70(3), 355-363. doi: 10.2307/3146535

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(3), 305-314. doi: 10.2307/3172883

Teisl, M. F., Boyle, K. J., McCollum, D. W., & Reiling, S. D. (1995). Test-Retest Reliability of Contingent Valuation with Independent Sample Pretest and Posttest Control Groups. American Journal of Agricultural Economics, 77(3), 613-619. doi: 10.2307/1243229

Thiene, M., & Scarpa, R. (2009). Deriving and testing efficient estimates of WTP distributions in destination choice models. Environmental and Resource Economics, 44, 379–395. doi: http://dx.doi.org/10.1007/s10640-009-9291-7

Train, K. (2003). Discrete choice methods with simulation. Cambridge: Cambridge University Press.

Train, K., & Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In R. Scarpa & A. Alberini (Eds.), Applications of simulation methods in environmental and resource economics (pp. 1-16). Dordrecht: The Netherlands: Springer Publisher.

Unwin, M. (2009). Angler usage of lake and river fisheries managed by Fish & Game New Zealand: results from the 2007/08 national angling survey. In M. Unwin (Ed.): A Report Prepared by the National Institute of Water and Atmospheric Research Ltd on behalf of Fish & Game New Zealand.

Unwin, M., & Image, K. (2003). Angler usage of lake and river fisheries managed by Fish & Game New Zealand: results from the 2001/02 National Angling Survey. NIWA Client Report: CHC2003- 114: Fish & Game New Zealand.

Whitehead, J. C., & Hoban, T. J. (1999). Testing for Temporal Reliability in Contingent Valuation with Time for Changes in Factors Affecting Demand. Land Economics, 75(3), 453-465. doi: 10.2307/3147190