Ecological Economics 69 (2010) 1010–1022

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Methods

Testing different types of benefit transfer in valuation of ecosystem services: New Zealand winegrowing case studies

Ramesh Baskaran [a], Ross Cullen [a,*], Sergio Colombo [b]

[a] Department of Accounting, Economics and Finance, Lincoln University, New Zealand
[b] Department of Agricultural Economics, IFAPA, Granada, Spain

ARTICLE INFO

Article history:
Received 19 January 2009
Received in revised form 7 December 2009
Accepted 31 January 2010
Available online 23 February 2010

Keywords:
Benefit transfer
Choice modeling
New Zealand winegrowing
Ecosystem services

JEL classification:
Q1
Q2
Q5

ABSTRACT

Most ecosystem services (ES) are neither priced nor marketed. Resource managers may fail to take into account degradation of unpriced services in their resource management decisions. Being able to estimate values for ES is fundamental to designing policies to induce resource users to provide (or improve) ES at levels that are acceptable to society. Conducting ecosystem valuation via non-market methods is costly and time consuming. Benefit transfer (BT) using choice experiment (CE) is a potentially cost-effective method for valuing ES by transferring information from existing valuation studies (and study sites) to a target area of interest (policy sites). The prime objective of this paper is to examine the validity of BT and hence whether it is feasible to conduct the transfer process and assist policy making. The paper focuses on the environmental impact of winegrowing practices in two New Zealand winegrowing regions. The two sites, Hawke's Bay and Marlborough, have similar environmental issues and attributes but are geographically separated. The study estimates Willingness to Pay (WTP) and Compensating Surplus (CS) for ES applying CE and, subsequently, given the preferences of respondents across sites and populations, tests the transferability of unadjusted value transfer (WTP) and benefits function (CS) assessing four different types of BT.

1. Introduction

Non-market valuation methods have contributed an important set of new tools, in particular choice experiment, to estimate the value of Ecosystem Services (ES) where they are not priced via markets.[1] Designing cost-effective public policy requires estimates of how valuable improvements in the level of specific ES are to the public. Research is also required to determine which kinds of ES could provide the greatest overall welfare benefits to society (Swinton et al., 2007). Thus, ecosystem service valuation can potentially provide new ways to compare the costs and benefits of different agricultural strategies, using the dollar as the common metric of value.

However, non-market valuation studies are time consuming, labour intensive, and costly. Research funders are interested in finding ways to reduce costs of valuing ES and other non-market items. Value transfer uses value estimates from an existing study and transfers it to another site or alternative context that is of interest. Benefit transfer (BT) is attractive if it can provide acceptable estimates of value at lower cost than would new non-market value studies for each new site or context. Nevertheless, there are concerns about the accuracy of the values that are transferred and research is needed to determine in which circumstances benefit transfer will provide acceptable value estimates. Furthermore, present literature on BT indicates conflicting conclusions on its potential usefulness to policymakers.

This paper has two objectives. The first is to estimate values for selected ES associated with winegrowing. The second objective involves checking if transfer of the estimated ES values across sites and populations is valid. An advanced choice experiment (CE) approach incorporating heterogeneity preferences, known as Random Parameter Logit (RPL) model is used to estimate the selected ES values. Surveys focused on the two largest New Zealand winegrowing regions, Marlborough and Hawke's Bay, are used as case studies. This research is conducted with a goal of applying BT. The research treats each region as both a 'study' site (an original survey site from which to transfer values to other sites) and as a 'policy' site (whereby site values are transferred to the policy site from the original survey site). This study assesses the accuracy of such transfers. By comparing values, the study obtains an estimate of the 'transfer error' (i.e., the difference between the value obtained by surveying a given site and the value obtained by transfer from another site) and provides a series of recommendations about whether it is advisable, and in which conditions, to transfer the ES values resulting from various winegrowing practices.

* Corresponding author.
E-mail addresses: Ramesh.Baskaran@lincoln.ac.nz (R. Baskaran), Ross.Cullen@lincoln.ac.nz (R. Cullen), scolombo@ugr.es (S. Colombo).

[1] Choice experiment is one of the choice modelling approaches. Choice modelling encompasses a range of stated preference techniques that include choice experiments, contingent ranking, contingent rating and paired comparisons.

0921-8009/$ – see front matter
doi:10.1016/j.ecolecon.2010.01.008

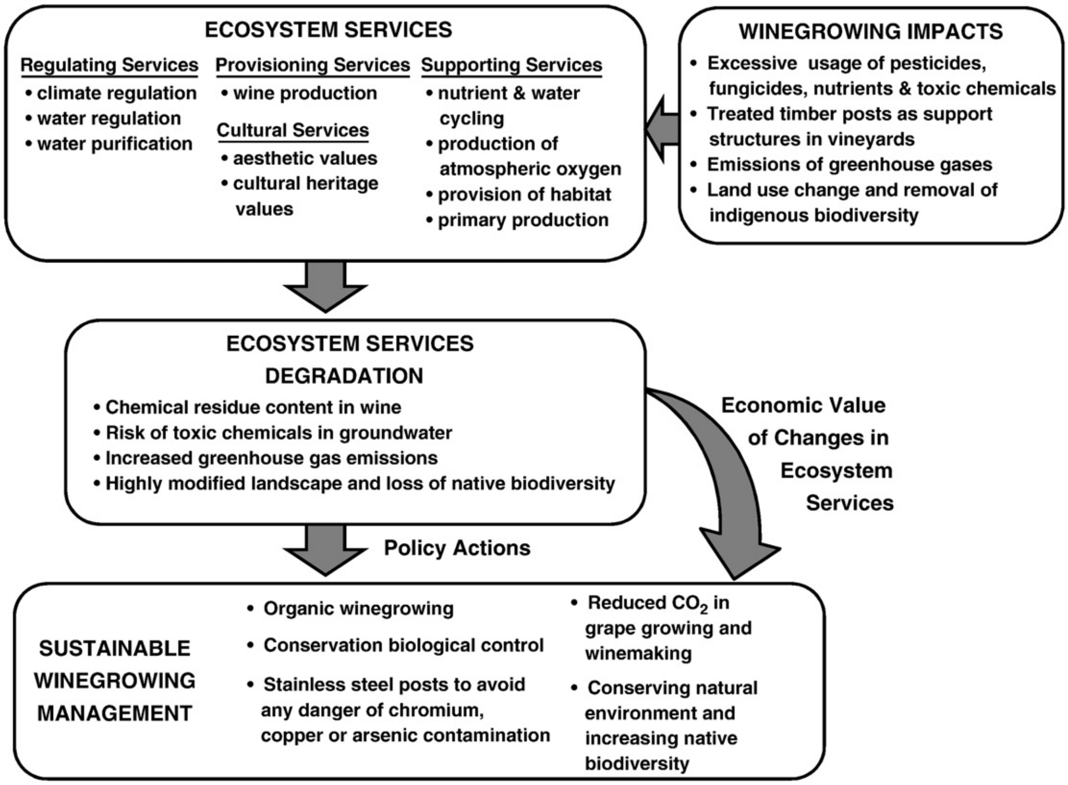

Fig. 1. Conceptual model of winegrowing impacts on ecosystem services.

Several conditions necessary for performing effective and efficient benefit transfers have been considered in the study design, in particular the similarity of site characteristics (Desvousges et al., 1992). The two study regions exhibit some similarities in terms of the importance of winegrowing; demographic profiles of the two populations; the extent and magnitude of the population that may be affected by resource use impacts; the type of value measurement (marginal value); and the period where the studies are carried out (temporality).

The study examines four types of BT tests as suggested by Morrison and Bergland (2006).[2] A key interest in this study is the spatial dimension between the study site and the policy site. Thus, an important hypothesis can be tested: Do the geographically separated Hawke's Bay (North Island) and Marlborough (South Island) regions have the same Willingness to Pay (WTP) estimates for the winegrowing ES considered, and hence, is BT across sites and populations valid?

The validity of four different types of BT is tested by determining whether the WTP estimates are statistically equivalent. Three tests are conducted for each type of BT. In the first test, the equality of the model parameters is examined. Second, the equality of WTP and Compensating Surplus (CS) is tested. Lastly, the study applies a new statistical validity test proposed by Johnston and Duke (2008) incorporating the tolerance level of transfer error for policy purposes. Assessment of this error may allow analysts to judge if the transfer process is reliable and hence, whether in the future it is valid to transfer values from study sites to policy sites without having to conduct new research or surveys.

1.1. Winegrowing environmental impacts on ecosystem services

Fig. 1 provides a representation of winegrowing impacts on ES. ES are defined as both good and services provided by the natural environment that benefit people. The Millennium Ecosystem Assessment (MA) identifies four broad categories of ES: provisioning services, regulating services, cultural services and supporting services. The study assesses winegrowing environmental impact on these ES such as food production (wine residue content), water quality (contamination of groundwater), air quality (climate change due to greenhouse gas emissions), and finally the cultural and aesthetic values (change in agricultural landscapes affecting local biodiversity). While producing grapes, winegrowers can manage land in ways that conflict with the healthy functioning of ecosystems, including excessive use of pesticides and fungicides that leach to groundwater and can become residues in wine, emissions of CO_2 that contribute to climate change, risk of chemicals leaching from wooden posts to groundwater, and removal of indigenous biodiversity. These impacts are not typically reflected in the winegrowers' income and therefore, may be a minor consideration in their decision making even though they directly affect the provision of the ES from winegrowing. Such actions raise questions about the sustainability of current practices. In this context, policy actions with regards to environmentally and economically sustainable winegrowing practices are needed to avoid further degradation of the environment and the consequent reduction of the ES supplied.

To examine the consequences of ecosystem degradation, or to assess the benefits of sustainable programs, the study will compare ecosystem benefits with and without the sustainable programs: that is, it must compare what would happen if sustainable measures were implemented to what would have happened if they were not. In other words, explicitly estimating the value of ES under the current and sustainable management regimes, and then comparing them. It is important to note that the application of this valuation method is not related to the total value of ecosystems but, rather, to valuing *changes* in ES. For example, a change of value from a highly modified landscape and loss of native biodiversity to a situation of conserved natural environment and protection of native biodiversity can be estimated from society's perspective. Society gains satisfaction from preserving the natural habitat and conserving the biodiversity, so their WTP for better ES provision can be considered as a "proxy" for ES valuation. To

[2] Morrison and Bergland (2006) provide extensive literature on four types of BT tests.

determine the basis for policy intervention to foster the provision of ES from winegrowing, the study measures the public's WTP for sustainable winegrowing practices. Thus, conducting ecosystem valuation is fundamental to designing policies to induce growers to provide (or improve) ES at levels that are demanded by society.

1.2. Types of benefit transfer tests

This research has identified at least four types of BT tests across sites and populations as illustrated in Fig. 2. The four types of BT tests are as follows:

- Type 1 BT: Differences across populations only. This test compares the value estimates held by different populations towards one study site. It tests for population effects, as the site is held constant and the population varied. For example, Hawke's Bay and Marlborough populations valuing Marlborough site (B vs D) and vice versa for the Hawke's Bay site (A vs C).
- Type 2 BT: Differences across sites only. This test involves comparisons of the values held by a single population for multiple sites. The test asks whether the population has similar values for two sites (valuing their own as well as a different site). For example, Hawke's Bay population valuing their own and Marlborough site (A vs B) and vice versa (C vs D).
- Type 3 BT: Differences across sites and equivalent populations which are geographically separated. This involves comparing the values that respondents have for ES within their region to the values that another group of respondents who reside in a different region have for similar ES in their region. For example, Hawke's Bay and Marlborough populations valuing ES in their respective region (A vs D).
- Type 4 BT: Differences across sites and different populations. In general, this test compares values held by different populations towards sites other than their own. For example, Hawke's Bay population valuing Marlborough site (B) and Marlborough population valuing Hawke's Bay site (C).

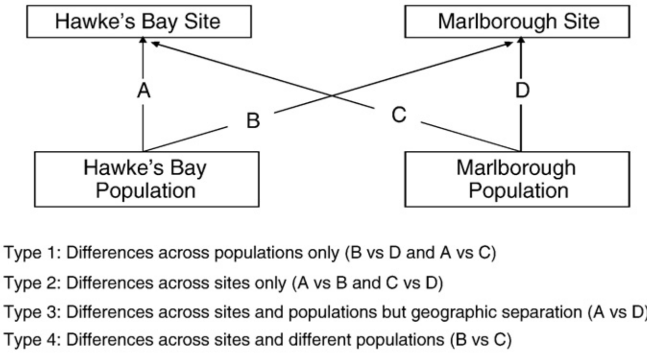

Fig. 2. Types of benefit transfer tests.

2. Method

2.1. Choice experiment (CE)

In this study, a CE method, Random Parameter Logit (RPL) was implemented to elicit respondents' preferences for different hypothetical changes in the quality and quantity of environmental attributes linked to winegrowing practices.[3] The RPL specifications provide the analyst with valuable information incorporating heterogeneity in the data while estimating unbiased parameters estimates (Hanley et al., 2006). In addition, in the context of BT analysis, the inclusion of taste heterogeneity may reduce the magnitude of the transfer error (Colombo et al., 2007). Therefore, a RPL modelling framework is applied to estimate the marginal WTP for improvements in the winegrowing ES and to determine the convergent validity of BT in valuing marginal changes in environmental quality which may differ across sites.

2.2. Welfare measures

The results of the CE estimation can be used to estimate two types of welfare measures: marginal WTP; and Compensating Surplus (CS). Marginal WTP are the marginal rates of substitution (MRS) between the attributes in question and the cost attribute. For a sufficiently small change in attribute, the mean WTP for improvement will be approximately equal to the marginal WTP. Hence, in a model without any socioeconomic interactions with the attributes, the marginal WTP for a single attribute can be estimated as the ratio of the coefficient of that attribute and the coefficient of the cost attribute:

$$\text{Marginal WTP} = -\left(\frac{\beta_{\text{attribute}}}{\beta_{\text{cost}}}\right).$$

CS is the appropriate estimate of the WTP for multiple changes in attributes. In particular, a change in management strategies from the current situation to improved situation. The mean WTP for this suite of changes in attributes can be estimated using Hanemann (1984) utility difference expression:

$$CS = -\frac{1}{\beta_{\text{cost}}}\left\{\ln\left[\sum_{j=1}^{J} e^{V_j^1}\right] - \ln\left[\sum_{j=1}^{J} e^{V_j^0}\right]\right\}$$

where β_{cost} is the parameter estimate on cost (payment vehicle), V^0 represents the utility of the current management plan (before the change) and V^1 represents the level of utility of the improved management plan (after the change) in a given J number of alternatives in each state.

2.3. Data collection

The CE surveys contained multiple choice questions (choice situations) about alternative policies for improving four ES attributes associated with winegrowing. The questionnaire consisted of three parts.[4] The first part contained questions regarding respondents' opinions and their awareness of current environmental impacts of winegrowing and its impacts on ES. The second part of the survey contained the choice situation questions. Respondents were first briefed about the selected ES attributes and associated cost to the household to achieve change in level of ES. The cost to the household (the payment vehicle) was defined as an additional annual payment to the regional council responsible for the management of the environment over the next five years. This payment vehicle was tested during the pilot study and was judged to be suitable for the ES of interest.

In the choice questions, respondents were asked to select the option they favoured the most out of the three alternatives provided. Each option contained the four attributes and the cost to the household with various levels of attribute combinations. Attributes discussed were toxic chemical residues in wine, risk of toxic chemicals from vineyards reaching groundwater, greenhouse gas emissions from vineyards, and the condition of native wildlife populations in vineyards. Each attribute was presented to respondents as several discrete levels. For example, the attribute of greenhouse gas emissions

[3] The RPL model is a generalisation of the standard conditional logit model that explicitly considers taste variation among individuals. Readers who are interested in the theoretical underpinnings of RPL can refer to the papers of Train (1998), Chapter 6 of Train (2003) or Chapters 15 and 16 of Hensher et al. (2005).

[4] The survey booklet provided 5 pages of information on the impacts of winegrowing practices on ES and is available upon request from the corresponding author.

was presented as having three discrete levels: zero net emissions (highest improved level); 30% reduction; and 'no change' from current emission level. The attributes were effects coded instead of dummy coding because of the advantage of estimating the coefficient estimates of attribute variables in which the effects are uncorrelated with the intercept of the regression (Bech and Gyrd-Hansen, 2005). All of the attributes selected are factors that a policymaker can affect, directly or indirectly and they were selected based on expert advice, current debates in focus groups and wine industry literature. Table 1 provides a more complete description of all explanatory variables and their specified effects coding based on the levels. The final part of the survey contained questions regarding the respondents' socioeconomic status.

Table 1
Definition and coding of variables.

Variable	Description
Attribute variable	
RESORG	Organic wine with fewer residue levels Effect Coding: 1 if organic wine; 0 if zero residue; −1 if current level
RESZERO	Wine with no detectable residue levels Effect Coding: 1 if zero residue; 0 if organic wine; −1 if current level
WATLOW	Low risk of toxic chemical reaching groundwater Effect Coding: 1 if low risk; 0 if no risk; −1 if high risk
WATNO	No risk of toxic chemical reaching groundwater Effect Coding: 1 if no risk; 0 if low risk; −1 if high risk
GHG30	30% reduction on greenhouse gas emissions per hectare per year Effect Coding: 1 if 30% reduction; 0 if zero reduction; −1 if current level
GHGZERO	Zero greenhouse gas emissions per hectare per year Effect Coding: 1 if zero reduction; 0 if 30% reduction; −1 if current level
NAT10	10% increase of natural environment and native wildlife populations Effect Coding: 1 if 10% increase; 0 if 30% increase; −1 if current level
NAT30	30% increase of natural environment and native wildlife populations Effect Coding: 1 if 30% increase; 0 if 10% increase; −1 if current level
COST	Cost to household per year for the next 5 years — NZ$0, 15, 30, 45, 60, 75, 90
Non-attribute variable	
ASC	Alternative specific constant on value of 1 for Alternative 1 and 2, and 0 for the current level
SATIS	How satisfied is respondent with environmental quality (1 = not; 3 = highly)
VISIT	How many times visited the region in the last five years (1 = none; 4 = >5 times)
CLOSE	How close is respondent from the nearest vineyard (1 = >20 km; 5 = <200 m)
VINLAND	Respondents enjoy vineyards with native plant species (1 = strongly disagree; 4 = strongly agree)
WQ	Respondents think that winegrowing damages groundwater (1 = strongly disagree; 4 = strongly agree)
GHGE	Respondents think that winegrowing increase greenhouse gases (1 = strongly disagree; 4 = strongly agree)
HEALTH	Respondents think that winegrowing leaves dangerous residues in wine (1 = strongly disagree; 4 = strongly agree)
WINELABEL	Respondents would like wine bottles to be labelled to show environmental friendly practices in winegrowing (1 = strongly disagree; 4 = strongly agree)
MALE	Respondent sex (1 = male; 0 = female)
AGE	Respondent age
EDU	Respondent education (1 = primary school; 4 = degree/professional)
JOB	Respondent occupation (1= based on agriculture sector; 0 = otherwise)
INCOME	Respondent income (1 = ≤$20,000; 6 = >$100,000)
UNDER	Respondents think the survey was easy to follow (1 = strongly disagree; 4 = strongly agree)

There are four attributes with three levels and the cost attribute with 6 levels ($3^4 \times 6^1$) which were combined in fractional factorial main effects experimental design (Louviere et al., 2000), providing 18 profiles in order to form the choice sets. The choice sets were constructed following the procedure proposed by Street et al. (2005) obtaining choice sets with a D-efficiency rate of 94.85% which were then blocked into 3 versions of 6 choice sets.[5] Each choice question has three alternatives and the third alternative was always the status quo (current plan). In other words, each respondent in each choice set had to choose either an improved environmental management plan (Alternative 1 or 2) or the current plan (Alternative 3).

In February 2008, pilot surveys were conducted on randomly selected residents in Canterbury, New Zealand. During April 2008 a pre-survey card, survey booklet, cover letter, and a reminder post-survey card were mailed to 4392 respondents selected from the New Zealand electoral roll using a random sampling design.[6] The sample was divided into four strata: 1098 respondents were randomly selected from the Marlborough population to value ES in their own region (MARL); 1098 respondents from the Hawke's Bay population to value ES in their own region (HB); 1098 respondents from Hawke's Bay population to value ES in Marlborough region (HBPOP); and lastly, 1098 respondents from Marlborough population to value ES in Hawke's Bay region (MARLPOP). The effective response rates for these samples are 30%, 20%, 18% and 24% respectively. These response rates are relatively low compared to similar non-market valuation studies. This may be due to the lack of awareness of non-market valuation of ES that may have discouraged some of the people sampled. In this circumstance, caution is needed when computing aggregate benefits, to avoid aggregation bias.

3. Results and discussion

3.1. Socioeconomic and attitudinal characterization of the samples

Descriptive statistics and socio-demographic data for the four samples show that the four samples do not differ much from each other but show greater differences in comparison to regional population.[7] These differences between the samples and the total population may influence the model estimates. However, the CE method allows such biases to be corrected by using benefit function estimates rather than point estimates (mean WTP). The inclusion of socioeconomic variables in the benefit function hence allows the adjustment of the WTP (van Bueren and Bennett, 2006).[8]

3.2. HBPOP and HB samples

There were 192 and 197 respondents who provided completed surveys from HBPOP and HB samples, respectively. Of the total number of respondents, 28 (15%) in HBPOP and 26 (13%) in HB samples expressed a protest answer regarding the proposed project. These protest bids were removed from the sample.[9] The majority of

[5] The D-efficiency criterion is defined as "the *p*th root of the ratio of the determinant of the information matrix of the proposed design to that for the optimal design" (Street and Burgess, 2007, page 86).

[6] The New Zealand electoral roll has a complete available adult population record and it is considered to be a representative of general population as it is illegal not to be on the electoral roll.

[7] This data is available upon request from the corresponding author.

[8] This is possible either by interacting the socioeconomic variables with the attributes or by including them in the regression function that is used to estimate the compensating surplus.

[9] It is established in the literature that some respondents do not state their true value for the good in question. Respondents may state a zero WTP although their true WTP is higher than zero or they may state a very high amount which is much greater than their true WTP (Meyerhoff and Liebe, 2008). If protest occurs, this will result in an incorrect economic value estimation of the good in question. In this study, those who expressed a protest answer were identified using a follow up question and deleted from analysis.

respondents who provided a protest response stated 'vineyards as business should pay for any damages they cause' as their main reason for not paying for the improvement plans. All respondents that displayed a genuine zero WTP by always choosing the current policy option (7% in HBPOP and 6% in HB), and those that chose either alternative 1 or 2 at least once were considered in the analysis, giving a total number of 974 and 962 observations for HBPOP and HB models estimation respectively.

Environmental attitudes and beliefs about winegrowing practices were obtained from each of the four samples.[10] The results show that more than three quarters of the sample in HBPOP declared 'don't know' whether they are satisfied with the environmental quality in the Marlborough region. About 56% of the respondents had not visited the area before and may not be aware of the environmental conditions of the area. More than three quarters of the sample in HB are satisfied with the environmental quality in the Hawke's Bay region and more than 75% of the sample live less than 5 kilometers away from a vineyard. In response to a statement 'you enjoy views of vineyard landscapes that include native plant species', about 85% of HBPOP and 53% of HB respondents answered in the affirmative. However, interviewees' preferences in HB are divided into two groups with 53% of the sample agreeing with the statement and 43% in disagreement. Respondents were also asked their opinions on whether winegrowing practices are harmful to groundwater quality, emit greenhouse gases, and affect health through chemical residues in wine. In general, respondents agree that winegrowing practices have the potential to damage the environment unless properly managed, but perceptions vary about these issues. It is surprising to find for both samples that a substantial percentage of respondents did not know the effect of winegrowing on these attributes. As might be expected, more than 75% of respondents in HBPOP would like to see wine bottles labelled so that consumers can be guaranteed that environmentally sustainable practices have been used in grapegrowing and winemaking. However, 79% of the HB respondents stated that they do not want to see wine bottles labelled so that consumers can be guaranteed that environmentally sustainable practices have been used in grapegrowing and winemaking. This suggests that the majority of Hawke's Bay respondents, who are satisfied with their regional environmental quality, are more confident of their wine being produced in an environmentally sustainable manner and hence, labelling is unimportant. If wine is from other regions, they prefer to see the labels provide information on sustainability of production practices.

3.3. MARLPOP and MARL Samples

There are 262 and 301 completed surveys from respondents of MARLPOP and MARL samples respectively. Of the total number of respondents, 27% (MARLPOP) and 12% (MARL) expressed a protest answer regarding the proposed project and were removed from the sample. It is also observed that 54% (MARLPOP) and 57% (MARL) of the protest respondents want vineyards to accept the costs of changed production systems. All respondents that displayed a genuine zero WTP by always choosing the current policy option (6% in MARLPOP and 4% in MARL), and those that chose either alternative 1 or 2 at least once were considered in the analysis, giving a total number of 1134 (MARLPOP) and 1509 (MARL) observations for model estimations.

About 64% of the respondents stated they had never visited the HB region. Almost 83% of the sample in MARLPOP declared they lacked knowledge about the environmental quality in the Hawke's Bay region. In the MARL sample, nearly 88% were satisfied with the environmental quality in the region. The degree of satisfaction is the highest in this region relative to other samples. Respondents who live close to vineyards are most likely to be exposed to spray drift and other external effects of viticulture. They may also be employed in vineyards. Living close to a vineyard may affect responses (i.e., ASCCLOSE). About 37% and 31% of the MARL and HB respondents respectively live less than 1 km from a vineyard. The MARLPOP respondents also have similar tastes for scenery to HBPOP respondents with approximately 81% of the sample concurring with the statement they enjoy views of vineyards landscape that include native plant species. In contrast, MARL respondents differ in the enjoyment they experience viewing vineyards landscape and approximately 90% of the sample disagreed with the statement. As described, respondents were also asked to affirm whether winegrowing practices are harmful for groundwater quality, emit greenhouse gases and damage health via wine residue content. In the MARLPOP sample 59% of the respondents are concerned about groundwater quality issues. Likewise, 53% are aware that toxic chemicals in wine residues are dangerous for health. About 29% of the respondents agree that winegrowing produces greenhouse gas emissions, but 34% did not know whether winemaking contributes to greenhouse gas emissions. In contrast, MARL respondents generally disagree with the statements. In particular, 59% (24% did not know) of the sample disagreed that winegrowing can effect groundwater quality, 42% disagreed (34% did not know) that it contributes to greenhouse gases emissions and 57% disagreed (17% did not know) that pesticides in wine are dangerous for health. Finally, similar to Hawke's Bay residents, people in Marlborough would like to see wine bottles to be labelled especially if they come from other regions. In addition, they also judged that their local wine was produced via environmentally sound practices, and it is not necessary for Marlborough labels to provide that information.

3.4. RPL models

The choice data were analysed using NLOGIT 4.0 statistical software. Table 2 presents RPL models for the four samples in which the socioeconomic and attitudinal characteristics of respondents have been added. The models were estimated using 100 Halton draws and considered the random parameters to be independent.[11] In this study, all the attributes except COST which has a triangular distribution, are assumed to be random variables with normal distribution.[12] Special care has to be taken when treating all attributes as random. Swait (2007: 259) strongly caution that "analyst should err on the side of parsimony when specifying random parameter models." In this study, we started with a simple model constraining several attributes as fixed and gradually allowing the attributes to be random. There is a significant improvement in terms of model prediction by allowing all attributes to be randomly distributed relative to models with fixed attributes.

The normal distribution for the non-monetary attributes was used because respondents may be indifferent to increasing or diminishing quality or quantity of the attributes. For instance, people who completely trust the effectiveness of food safety regulations may not care if the toxic chemical residue content in wine increases a little. The cost attribute is assumed to follow a triangular distribution to ensure non-negative WTP for winegrowing improvements over the entire range of the distribution which guarantees deriving behaviorally meaningful WTP measures while allowing taste heterogeneity for this attribute.[13] The use of a constrained distribution may be restrictive

[10] This data is available upon request from the corresponding author.

[11] All the random parameters models described in this study have been estimated using these settings.

[12] A referee has argued that researchers in the choice experiment analysis regularly choose not to distribute one or more variables of RPL models to assist in ensuring identification. Following Ruud (1996), allowing all the parameters to be randomly distributed often leads to problem with identification and it is advisable to leave at least one parameter fixed. In this study, we include a constant term as fixed in estimating the model.

[13] Following Hensher et al. (2005), a constrained triangular distribution was used in which the variance (spread) of the distribution is made equal to the mean. Such a constraint forces the Cost coefficient to have the same sign across the entire distribution. This is useful where a change of sign does not make sense.

and therefore, the analyst should be cautious of its limitations. We tested both HB and MARL datasets by estimating models where the cost variable was not constrained. In the HB case the unconstrained model was not significantly better than the constrained one, given that the spread of the cost attribute was less than double the mean value. In the MARL sample the unconstrained model is statistically superior to the constrained one, although significant changes in the signs and magnitudes of the attributes were not identified. Due to the need to maintain the same model specification across all datasets and to derive behaviorally meaningful WTP, we finally opted for the constrained cost distribution.[14]

In the analysis, all the attributes except for cost were effect-coded, using the level of the current winegrowing management practices as reference point or base level for each attribute.[15] The effects coding system does not directly estimate the parameter of the base level; nonetheless it can be inferred from the estimation of the two effect-coded corresponding attribute parameters. In other words, the parameter of the current level of management practices for each attribute is equal to the negative sum of the estimated coefficients for that attribute. For example, the coefficient of the current (base level) water quality in HBPOP model is equal to $-(\beta_3+\beta_4)=-(1.0975+1.2212)=-2.3187$. The coefficients of all the attribute base levels have similar negative sum values. Thus, in Table 2, it is noticeable that respondents have an aversion to the current winegrowing management in all the four samples.

Overall the models are highly significant and show an excellent fit to the data.[16] All the significant attribute coefficients have the *a priori* expected sign for the models. In all the models, it is shown that the attribute RESORG is insignificant. This suggests that reducing the residue content in wine is a matter that significantly affects people's satisfaction only if the reduction is complete (zero level) rather than a marginal reduction. However, with the exception of the HB sample, the high heterogeneity associated with this attribute (note that the standard deviations are far larger than the mean values) show that one proportion of the sample derive utility from reducing the residues content in wine, whilst the other is not interested in the quantity of wine residues. The preferences almost evenly distributed along the two sides of the zero are probably the cause of the insignificance (statistically) of the coefficient of this attribute.[17] In addition, this could also be due to the lack of knowledge and high levels of disagreement among respondents with the statement that chemical residues in wine are of concern. The effects that winegrowing has on underground water quality is deemed extremely important (highly significant and large coefficients) by the four samples of respondents, and a reduction in the risk of toxic chemicals reaching groundwater increases respondents' utility.

Reduction in the emissions of greenhouse gases is also of concern to all the sample respondents. Nevertheless, only a reduction of greenhouse gas emissions by 30% increases HB and MARLPOP respondents' utility, and they are indifferent about a reduction to zero emissions. All four sample's respondents prefer increasing the native wildlife population in vineyards by at least 10% or 30% relative to the current condition. For example, HB respondents favour increasing the native wildlife population in vineyards by at least 30% relative to the current condition. A smaller improvement (10% increase) is not of interest to them. On the other hand, MARL respondents get more utility from a 10% increase. As expected, cost is highly significant and has a negative sign for all the samples, showing that the higher the cost associated with a policy option, the less likely a given respondent is to choose that option.

By interacting individual socioeconomic and attitudinal variables with the alternative specific constant (ASC), it is possible to enrich information about a particular sample and also to explain a part of respondent heterogeneity.[18] It is surprising to note that most of the ASC in the models except MARL are negative with a large coefficient and are highly significant, showing that there are systematic reasons other than the attribute values that drove respondents when choosing the status quo option. MARL sample respondents who are satisfied with environmental quality prefer to hold on to current management practices instead of improving them. It is noticeable that HB sample respondents who live closer to vineyards are more likely to stick with the current winegrowing management. This may reflect concerns about possible loss of employment if higher environmental standards were to be introduced.

Although the HB and HBPOP respondents are from the same region, the results only show some similarities in terms of socioeconomic interactions. For example, both sets of sample respondents are in favour of winegrowing management practices that lead to more wildlife in the landscape, reduced wine residues, and more informative labelling (production methods) of wine bottles. The HBPOP respondents who are young and highly educated support better environmental management. By contrast, HB residents, in particular females and older people are more likely to choose the improvement plans over the current environmental management. Highly educated residents are also not in favour of improving the present conditions. Household income is significant and has opposite signs for HB and HBPOP models indicating that higher income people are more likely to support the proposed winegrowing management practices in their own region but not outside their region.

Comparing MARL and MARLPOP samples, it is clear that males and younger people in MARL are more likely to choose the improvement plans, the opposite to what is observed in the MARLPOP model. For the MARL sample, neither household income nor education affects the choice of the improvement alternatives relative to the status quo. Respondent occupation significantly affects choice of the current situation relative to the various alternatives. In particular, people who work in agriculture or resource based sectors are more likely to prefer current winegrowing management over alternative management practices. This may be due to apprehension of incurring extra costs or losing income if there is a change in their management or a cultural reason to continue with the current management practices. It is also interesting to observe that in MARLPOP sample, higher income respondents indicate their support for reduced environmental degradation happening in Hawke's Bay region. Lastly, both MARL and MARLPOP respondents who found difficulty in understanding the environmental issues described in the questionnaire have lower probability of choosing the two improvement alternatives relative to current winegrowing management.

All of the standard deviation terms are significant for all the models (except for RESORG in HB and NAT10 in MARLPOP and HB) indicating preference heterogeneity does indeed exist.[19] This may be expected given the differing opinions of respondents about the effects of winegrowing management on groundwater quality, wildlife,

[14] If the cost distribution contains the zero value, then there is higher likelihood of getting draws very close to zero which causes an “explosion” of the WTP value. If the distribution is constrained to be positive, the likelihood of getting these values is reduced and hence, generally do not enter in the 95% confidence interval of the resulting empirical WTP distribution.

[15] Refer to Louviere et al., 2000 for detailed discussion of the different coding schemes.

[16] Simulations by Domencich and McFadden (1975) suggest values of ρ^2 between 0.2 and 0.4 are comparable to values between 0.7 and 0.9 for R^2 in the case of ordinary linear regression.

[17] Given the mean and standard deviation values of the RESORG coefficient, we observe that around 50% of the sample hold negative preferences and 50% positive preferences towards reducing the residue content in wine. For instance, in the HBPOP sample, 51% of the population is not interested in reducing the residues content in wine, whilst 49% would prefer otherwise.

[18] ASC is basically the constant term and it is used to capture the influence on choice of unobserved attributes relative to specific alternatives (in this study, alternatives 1 and 2).

[19] Note that the parameter estimate for the standard deviation of Cost is exactly the same as that of the absolute value of its mean for all the samples, which is due to the constraints imposed on the cost distribution.

Table 2
RPL model results for HB, HBPOP, MARL and MARLPOP.

Variable	HB	HBPOP	MARL	MARLPOP
Random parameters				
RESORG	−0.1476	−0.0235	−0.0841	−0.0809
RESZERO	0.3162**	0.3372**	0.2647***	0.2202**
WATLOW	0.9633***	1.0975***	0.9059***	0.8639***
WATNO	1.0528***	1.2212***	1.1871***	1.0765***
GHG30	0.5649***	0.3574**	0.2224**	0.4544***
GHGZERO	0.1709	0.5503***	0.4408***	−0.0418
NAT10	−0.0486	0.2582*	0.4712***	0.1843*
NAT30	0.5824***	0.3887***	0.2980***	0.3888***
COST	−0.0385***	−0.0362***	−0.0195***	−0.0263***
Non-random parameters				
ASC	−13.8075***	−16.0329***	0.7256	−14.4547***
ASCSATIS	−0.9589	0.2903	−1.4961***	−0.4947
ASCVISIT		0.0204		0.4055
ASCCLOSE	−1.5122***		0.1247	
ASCVINLAND	1.7121***	0.9195**	0.4064	1.2863***
ASCWQ	0.6836	0.9356*	0.0145	1.2695***
ASCGHGE	−0.1013	−0.7729*	0.2620	1.6587***
ASCHEALTH	0.2937	1.7350***	0.5261	0.9586**
ASCWINELABEL	2.2091***	1.9368***	0.5036	−0.5136
ASCMALE	−1.7411***	0.7043	0.8138**	−1.2178**
ASCAGE	0.0336*	−0.0477***	−0.0283**	0.0143
ASCEDU	−0.5607*	0.5770*	0.4316	−0.0099
ASCJOB	−1.2128	−0.1570	−1.0253***	0.1571
ASCINCOME	0.7009***	−0.4174**	−0.1218	0.8135***
ASCUNDER	1.0474*	0.2544	−1.0895***	−1.3794***
Standard deviation				
NsRESORG	0.1306	0.8921***	0.6465***	0.7116***
NsRESZERO	0.6587***	0.3364**	0.6850***	0.4539**
NsWATLOW	1.1437***	1.5257***	0.8677***	0.7495***
NsWATNO	0.8849***	1.2116***	1.5643***	1.1668***
NsGHG30	0.7047***	0.6057***	0.7147***	0.4183*
NsGHGZERO	0.8218***	1.1937***	0.5930***	0.5699***
NsNAT10	0.0724	0.9174***	0.8336***	0.3138
NsNAT30	0.8222***	0.8469***	0.9011***	0.8054***
TsCOST	0.0385***	0.0362***	0.0195***	0.0263***
Model statistics				
N (Observations)	962	974	1509	1134
Log likelihood	−584.71	−602.64	−962.15	−679.10
McFadden pseudo R^2 (%)	44.7	43.7	41.9	45.5
χ^2 (degrees of freedom)	944.30*** (31)	934.81*** (31)	1391.31*** (31)	1133.45*** (31)

Notes: Standard errors in parentheses; single (*), double (**) and triple (***) asterisks denote significance at the 10%, 5% and 1% levels respectively.

greenhouse gases emissions and health. As well, respondents' lack of knowledge about the issues may be a contributing factor that increases heterogeneity in respondents' choices.

Table 3 reports the estimates of mean WTP and Compensating Surplus (CS) derived from the models.[20] There are two main approaches to BT in this study: marginal value transfer and function transfer. The mean WTP for all the attributes can be considered as marginal value transfer (unadjusted WTP), and this assumes that the welfare change experienced by the average person in the study site is the same as that experienced by the average person in the policy site. The function transfers encompass the transfer of a benefit function such as CS from a study site to a policy site that involves combination of multiple attributes using utility models of respondent choice behaviour. This will help to identify the extent to which benefit functions can be transferred between sites. It includes the calculation of the benefits that respondents receive from the environmental condition of winegrowing regions both before and after the change in alternative management that is being proposed. In this study, the function transfer was done without adapting the function to fit the specifics of the policy site such as socioeconomic characteristics. In other words, it is transferred with an absolute term (unadjusted CS) in the first place to test their equivalence between study site and policy site and, subsequently, can be used to forecast a benefit measure for the policy site by adjusting it with socioeconomic characteristics (adjusted CS).

The estimated values are mean WTP annually for a period of five years for a change (improvement) in the ES attributes concerned, *ceteris paribus*. The mean WTP for all the attributes are positive, implying that on average, respondents have positive utilities (well being) for increases in the quality or quantity of each attribute regardless of whether these improvements occur in their region of residence or in a geographically distant region. The WTP for reduced chemical residues in organically produced wine (RESORG) is not statistically different from zero in all models. Notice that for all the samples, the risks of contamination of groundwater quality are highly valued and are viewed to be the most important attribute.

Values are much higher for the MARL sample even though the mean household income for this sample is not much different in comparison to other samples. A possible reason could be that Marlborough region has experienced very rapid rate of change in rural land use from sheep and crops to vineyards while becoming the largest wine growing region in New Zealand. The fast expansion of

[20] WTP is the amount of money respondent is willing to pay in order to receive one unit more of the non-marketed environmental attribute. It is calculated as a ratio of the attribute coefficient to the cost coefficient. Mean WTP is estimated for the entire sample as an average respondent willingness to pay. CS is the overall WTP for a change from the status quo and is calculated as a ratio of the utility differences between the status quo and change scenarios to the cost coefficient.

Table 3
Mean annual WTP per household for the HBPOP, HB, MARLPOP and MARL attributes.

Attribute	HBPOP	HB	MARLPOP	MARL
RESORG	7.26#	0.75#	2.33#	4.40#
	(−9, 24)	(−9, 11)	(−16, 20)	(−19, 26)
RESZERO	15.47	10.69	11.91	19.56
	(1, 31)	(−1, 23)	(−4, 29)	(−3, 47)
WATLOW	80.12	64.98	91.04	132.44
	(51, 118)	(45, 88)	(62, 134)	(84, 213)
WATNO	83.17	67.11	97.82	145.29
	(54, 120)	(48, 90)	(66, 145)	(87, 237)
GHG30	29.99	28.40	28.36	39.37
	(12, 53)	(15, 43)	(13, 50)	(14, 75)
GHGZERO	34.34	019.68#	12#	48.59
	(14, 62)	(6, 35)	(−5, 31)	(24, 89)
NAT10	21.30	10.54#	24.59	55.13
	(4, 42)	(−0.5, 22)	(9, 45)	(27, 101)
NAT30	24.64	24.53	31.75	47.81
	(8, 44)	(11, 39)	(11, 56)	(19, 88)
CS1	156.17	147.35	227.90	287.54
	(103, 228)	(111, 193)	(156, 336)	(183, 466)
CS2	175.13	164.69	235.04	317.44
	(118, 254)	(125, 216)	(162, 341)	(203, 518)
CS3	134.87	136.81	203.30	232.41
	(86, 198)	(103, 183)	(139, 300)	(150, 373)
CS4	104.88	108.41	174.94	193.03
	(68, 155)	(82, 146)	(121, 261)	(126, 304)

Notes: Confidence intervals (CIs) in parentheses at 95% level; the unconditional mean WTPs and CIs are calculated following the simulation procedure proposed by Hu et al. (2005); single (*), double (**) and triple (***) asterisks denote significance at the 10%, 5% and 1% levels respectively; # — non-significant coefficients.

vineyards and their environmental impacts may be a nuisance and of concern to many residents. On the other hand, the Hawke's Bay region is the oldest winegrowing region in New Zealand, has experienced slower growth rate in area of vineyards, and vineyards are more dispersed in the region. Thus, the impacts on ES are likely to be less intensive. This notion can be confirmed when comparing the WTP for the selected attributes between regions. It is interesting to observe that the mean WTP of the Hawke's Bay residents valuing their own region are lower than the mean WTP when valuing the Marlborough region. However, there is a completely diametrical point of view in the Marlborough region. The mean WTP for increasing environmental quality in the vineyards of the region are higher than the mean WTP for increasing environmental quality in vineyards of the geographically distant Hawke's Bay region. Another reason could be due to the scale effect where the estimated values may be sensitive to the ways in which the issues are presented to the respondents (Bennett, 2006). For example, values for the groundwater quality attributes in Hawke's Bay (4665 ha of vineyards) and Marlborough (13187 ha of vineyards) regions may vary if respondents were aware of the scale differences in terms of the area of vineyards in each region and the intensity of the issues in each region.

As for the CS estimations, four options were created for policy analysis relative to a baseline of current conditions.[21] The first policy option calculated (CS1) includes lower levels of improvements (wine residue — organic, water quality — low risk, GHG reduction — 30%, Native species increase — 10%). The second policy option includes the best levels of improvements. As expected, the CS increases if there is improvement over the current (deteriorating) ES towards better environmental conditions in winegrowing. For a change from current conditions to improved conditions as in Policy 1, on average, respondents in HB are willing to pay NZ$147.35 each year over five years for the specified ES improvements. In contrast, greater improvements under Policy 2 increases the mean WTP to NZ$164.69. In addition, the results also indicate the importance of attribute tradeoffs when calculating CS for environmental improvements. For instance, Policy 1 and Policy 3 differ only in terms of native wildlife effects (with and without native wildlife improvement). The 'without native wildlife' effect reduces WTP by about 7.2% for Policy 3 compared to Policy 1. Comparing Policy 2 to Policy 4, trading off GHG reduction and native wildlife attributes reduces WTP by about 34%. Overall the respondents on average not only experience positive marginal utility for improvement in the selected ES attributes but also are willing to pay more for higher levels of environmental enhancement.

Based on these estimated values, it is possible to calculate the typical transfer errors of unadjusted mean WTP (simple absolute value difference) as presented in Table 4. In general, an *a priori* expectation is that study and policy site populations are similar so the errors associated with value (function) transfer should be relatively small. Conversely, where the preferences of study populations differ substantially from those at the policy site both function transfers and absolute value difference transfers may well produce relatively large errors. It is important to observe that attributes RESORG, GHGZERO and NAT10, which had insignificant coefficients in the models, have larger transfer errors, given that the 'true' value is close to zero.[22] Another feature of the results is that, the low average transfer errors in BT Type 4 for HBPOP and MARLPOP samples indicate that these samples are very close in terms of implied WTP for an improvement. This suggests either that people view these regions as close substitute for each other in terms of how they value their own regions, or that they are equally good indicators of value in other areas. Overall, policymakers would probably find the average levels of transfer error in Table 4 (within the range of 30–80%) quite acceptable for cost–benefit analysis purposes especially when the benefits outweigh the costs (Colombo et al., 2007).[23] However, it is worth mentioning that the calculation of the transfer errors typically used in the literature only takes into account the mean WTP values of each attribute. If either a large sampling variance or preference heterogeneity is present, it is at least necessary to take into consideration the spread of the empirical WTP distribution. The following BT validity tests provide more details of this aspect.

3.5. Benefit transfer (BT) tests

The validity of marginal value and benefit function (CS) transfer ought to be tested via statistical hypothesis concerning the equality of the benefit measure for policy and study sites. In other words, the underlying economic assumption is the equality of preferences across sites and populations. The study performs three types of statistical test in order to validate the BT analysis, given broad similarities in the sites and populations characteristics.

3.6. Testing if the model parameters are equivalent

A comparison of preference estimates between the two sites needs to allow for the fact that the estimated parameters are confounded with a scale parameter which is inversely proportional to the variance of the random term. The study thus performs a grid search technique as proposed by Swait and Louviere (1993).[24] The

[21] The CS estimation is based on unadjusted CS without adjusting with socio-economic variables of the policy site.

[22] This is because the "true" value, which is assumed to be the value estimated at a particular site using an original study, enters in the formula for estimating the transfer error at the denominator $((|\text{predicted WTP}_{\text{site A}} - \text{observed WTP}_{\text{site A}}|)/\text{observed WTP}_{\text{site A}})$.

[23] There is no general rule on "acceptable BT error" as it all depends on the decision context: 30–80% would be acceptable in a Cost–Benefit Analysis where costs and benefits are not very close (as typically happens in environmental BT application where the benefits clearly outweigh the costs), but not acceptable if it is used as basis for a compensation claim after pollution incidents (Natural Resource Damage Assessment in the US). The authors thank Stale Navrud for this comment.

[24] The procedure tests if the differences in the model parameters of the two datasets are due to scale parameter. It stacks the two datasets and then rescales one dataset relative to another dataset in order to cancel out the differences.

Table 4
Transfer error (%) for unadjusted value transfer WTP (absolute value difference).

Models	RESORG	RESZERO	WATLOW	WATNO	GHG30	GHGZERO	NAT10	NAT30	AVERAGE
Type 1 BT									
(HBPOP − MARL) / MARL	**65**	21	40	43	24	29	61	48	38
(HBPOP − MARL) / HBPOP	**39**	26	65	75	31	41	159	94	70
(HB − MARLPOP) / MARLPOP	**68**	10	29	31	0.1	**64**	**57**	23	19
(HB − MARLPOP) / HB	**201**	11	40	46	0.1	**39**	**133**	29	25
Type 2 BT									
(HB − HBPOP) / HBPOP	**90**	31	19	19	5	**43**	**51**	0.4	15
(HB − HBPOP) / HB	**868**	45	23	24	6	**75**	**102**	0.4	20
(MARLPOP − MARL) / MARL	**47**	39	31	33	28	**75**	55	34	37
(MARLPOP − MARL) / MARLPOP	**89**	64	46	49	39	**305**	124	51	62
Type 3 BT									
(HB − MARL) / MARL	**83**	45	51	54	28	**60**	**81**	49	46
(HB − MARL) / HB	**487**	83	104	116	39	**147**	**423**	95	87
Type 4 BT									
(HBPOP − MARLPOP) / MARLPOP	**212**	30	12	15	6	**186**	13	22	16
(HBPOP − MARLPOP) / HBPOP	**68**	23	14	18	5	**65**	15	29	17
AVERAGE	175	32	36	39	16	28	59	35	

Note: Bolded transfer errors have insignificant coefficients and are not included in the average.

test is performed by estimating two models separately and then a combined model. A likelihood ratio test is then carried out using the log-likelihoods from each model. The test results for each type of BT are shown in Table 5.

For example, in Type 3 BT, the estimated variance–scale ratio was found to be 1.0 after stacking both HB and MARL datasets, then rescaling the HB data relative to MARL data which implies that the MARL sample has on average the same response variability as the HB sample. The likelihood ratio test statistic for a comparison of the choice model parameters between the HB and MARL is 142. The critical Chi-square value of 45 at the 5% significance level (31 degrees of freedom), is well below the calculated value. Therefore, it can be concluded that a significant difference does exist between the two sites and we can reject the null hypothesis, even after taking into account the scale differences. This means that using the whole set of models parameters for BT would be inaccurate or biased. However,

Table 5
Swait and Louviere (1993) procedure results for testing equivalent of model parameters.

Models	Variance–scale ratio	Log likelihood ratio $\chi^2 = -2(LL_{Model1+Model2} - (LL_{Model1} + LL_{Model2}))$	Critical value $\left(\chi^2_{(\alpha, K_1+K_2-K_{Pool})}\right)$	Reject H_0
Type 1 BT				
HBPOP & MARL	$\mu = 1.2$	$\chi^2 = -2(-1601.28 - (-604.64 - 962.15)) = 69$	$\chi^2_{(0.05,32)} = 46$	Yes
HB & MARLPOP	$\mu = 1.2$	$\chi^2 = -2(-1324.63 - (-584.71 - 679.10)) = 122$	$\chi^2_{(0.05,32)} = 46$	Yes
Type 2 BT				
HB & HBPOP	$\mu = 1.2$	$\chi^2 = -2(-1230.38 - (-584.71 - 604.64)) = 82$	$\chi^2_{(0.05,32)} = 46$	Yes
MARLPOP & MARL	$\mu = 1.2$	$\chi^2 = -2(-1684.05 - (-679.10 - 962.15)) = 86$	$\chi^2_{(0.05,32)} = 46$	Yes
Type 3 BT				
HB & MARL	$\mu = 1.0$	$\chi^2 = -2(-1617.68 - (-584.71 - 962.15)) = 142$	$\chi^2_{(0.05,31)} = 45$	Yes
Type 4 BT				
HBPOP & MARLPOP	$\mu = 1.3$	$\chi^2 = -2(-1320.06 - (-604.64 - 679.10)) = 73$	$\chi^2_{(0.05,31)} = 45$	Yes

Table 6
Poe et al. (2005) test results.

Models	RESORG	RESZERO	WATLOW	WATNO	GHG30	GHGZERO	NAT10	NAT30	CS1	CS2	CS3	CS4
Type 1 BT												
HBPOP & MARL	0.5836	0.3902	**0.0558**	**0.0429**	0.3129	0.2349	**0.0295**	**0.1068**	**0.0225**	**0.0237**	**0.0377**	**0.0204**
HB & MARLPOP	0.4384	0.4589	**0.1049**	**0.0728**	0.5214	0.7579	**0.0855**	0.3057	**0.03 52**	**0.0751**	**0.0536**	**0.0245**
Type 2 BT												
HB & HBPOP	0.2442	0.3069	0.2313	0.2111	0.4631	0.1502	0.1516	0.5055	0.4253	0.4168	0.5413	0.5708
MARLPOP & MARL	0.4370	0.2959	0.1170	0.1127	0.2713	0.0131	0.0467	0.2196	0.2264	0.1658	0.3357	0.3743
Type 3 BT												
H B & MARL	0.3730	0.2519	**0.0063**	**0.0055**	0.2679	**0.0311**	**0.0012**	**0.0915**	**0.0059**	**0.0066**	**0.0249**	**0.0145**
Type 4 BT												
HBPOP & MARLPOP	0.6628	0.6352	0.3348	0.2843	0.5501	**0.9368**	0.3997	0.3176	**0.0885**	0.1457	**0.0728**	**0.0329**

Note: Bolded denote as significance level at p-values lower than 0.10 or greater than 0.90 (i.e., Reject the null hypothesis that WTPs or CSs are equivalent).

Table 7
TOSC equivalence test results for unadjusted annual mean WTP and CS for TL of 50%.

Models	RESORG	RESZERO	WATLOW	WATNO	GHG30	GHGZERO	NAT10	NAT30	CS1	CS2	CS3	CS4
Type 1 BT												
HBPOP vs MARL – P_L	0.3146	0.3831	0.6050	0.6731	0.3547	0.4196	0.8915	0.6983	0.9741	0.9689	0.8346	0.8416
P_U	0.4749	0.1942	0.0029	0.0019	0.0837	0.0506	0.0059	0.0233	0.0196	0.0181	0.0069	0.0022
MARL vs HBPOP – P_L	0.3534	0.3273	0.3099	0.3477	0.2663	0.2839	0.5942	0.4501	0.9603	0.9572	0.9413	0.9596
P_U	0.5179	0.1552	0.0005	0.0001	0.0489	0.0238	0.0002	0.0034	0.0126	0.0127	0.0238	0.0100
HB vs MARLPOP – P_L	0.5463	0.3317	0.3505	0.4161	**0.1081**	0.0648	0.7971	0.3382	0.9641	0.9051	0.7633	0.8033
P_U	0.4232	0.2587	0.0013	0.0006	**0.1062**	0.4302	0.0281	0.0693	0.0344	0.0585	0.0065	0.0012
MARLPOP vs HB – P_L	0.5137	0.3102	0.1760	0.1959	**0.1083**	0.1135	0.5494	0.2504	0.9251	0.8862	0.9284	0.9401
P_U	0.3914	0.2393	0.0000	0.0000	**0.1065**	0.5676	0.0040	0.0381	0.0145	0.0468	0.0394	0.0088
Type 2 BT												
HB vs HBPOP – P_L	0.7432	0.4729	0.1886	0.1819	0.1524	0.6193	0.6945	0.1431	0.5707	0.5303	0.1512	0.0843
P_U	0.2320	0.1416	0.0061	0.0043	0.0993	0.0357	0.0619	0.1373	0.4213	0.3653	0.1839	0.1243
HBPOP vs HB – P_L	0.6209	0.3726	0.1118	**0.1019**	0.1387	0.4109	0.4927	0.1419	0.4233	0.4866	0.3971	0.2615
P_U	0.1418	0.0933	0.0018	**0.0011**	0.0881	0.0092	0.0209	0.1361	0.2802	0.3235	0.4777	0.3799
MARLPOP vs MARL – P_L	0.5289	0.5409	0.4168	0.4434	0.4022	0.9683	0.8259	0.4804	0.7687	0.8141	0.3789	0.2679
P_U	0.4038	0.1679	0.0070	0.0062	0.0678	0.0053	0.0083	0.0533	0.2216	0.1472	0.1237	0.1049
MARL vs MARLPOP – P_L	0.4986	0.4316	0.2249	0.2414	0.2882	0.7474	0.5246	0.3304	0.7093	0.8139	0.5881	0.5019
P_U	0.3744	0.1097	0.0025	0.0017	0.0346	0.0008	0.0008	0.0214	0.1716	0.1470	0.2662	0.2632
Type 3 BT												
HB vs MARL – P_L	0.6148	0.5956	0.8664	0.9009	0.3865	0.8736	0.9955	0.7075	0.9939	0.9912	0.8788	0.8825
P_U	0.3606	0.1376	0.0000	0.0000	0.0529	0.0049	0.0002	0.0148	0.0058	0.0049	0.0033	0.0008
MARL vs HB – P_L	0.5517	0.4574	0.4566	0.4993	0.2731	0.5751	0.8247	0.4441	0.9864	0.9889	0.9664	0.9653
P_U	0.3027	0.0761	0.0000	0.0000	0.0223	0.0003	0.0000	0.0011	0.0025	0.0039	0.0183	0.0054
Type 4 BT												
HBPOP vs MARLPOP – P_L	0.2343	0.1472	0.1224	0.1441	0.1089	0.0051	0.2742	0.3501	0.8993	0.8187	0.7077	0.7510
P_U	0.5451	0.3509	0.0222	0.0157	0.1613	0.6276	0.1342	0.0899	0.0777	0.1150	0.0118	0.0033
MARLPOP vs HBPOP – P_L	0.3019	0.1866	**0.0875**	**0.0937**	0.1201	0.0269	0.2333	0.2663	0.8533	0.7678	0.8882	0.9339
P_U	0.6234	0.4150	**0.0139**	**0.0086**	0.1760	0.8657	0.1084	0.0563	0.0504	0.0843	0.0457	0.0157

p-values lower than 0.10 indicate no differences in the two distributions (bolded).

interest may be focused on transferring the marginal WTP of only some attribute. It may happen that the transfer error of marginal WTP lies within an "acceptable" interval from the policymaker's point of view. This is what we test in the following analysis.

3.7. Testing if the mean WTP and CS are equivalent

The complete combinatorial method proposed by Poe et al. (2005) has been carried out to identify where differences in WTP might be occurring, given that a significant difference exists in the model's parameters. This test explicitly consider the empirical WTP distribution and seeks to determine whether the distribution of the WTP at the study site and at the policy site overlap significantly. If they do, it considers that the two WTP measures are not statistically different and can be considered suitable for transfer. Results are reported in Table 6.

The test suggests that there are no significant differences in WTP and CS between the HB and HBPOP (Type 2 BT). This means that both marginal value and CS (benefit functions) are equivalent between these samples and suggests that BT would be appropriate. The results indicate that attributes RESZERO and GHG30 are shown to be not significantly different in all the models and hence, it can be implied that marginal value transfer is valid for these attributes for all types of BT across sites and populations. All the CS estimates in Types 1 and 3 BT indicate that there are significant differences suggesting that benefit function transfer is not applicable for these models. However, Kristofferson and Navrud (2005) illustrated that the above results may provide a Type II Error null hypothesis (i.e., failing to reject the null hypothesis that WTPs are the same when indeed, it is false). The authors further comment from a policy perspective that it is important for the analyst to choose a tolerance limit when testing for the transferability of the welfare measures. The following section illustrates this test.

3.8. Testing if the mean WTP and CS are transferable

Kristofferson and Navrud (2005) reverse the traditional null and alternative hypotheses, and the burden of proof. In other words, welfare estimates are assumed different unless hypothesis tests can demonstrate, with a chosen probability level, that the difference is smaller than a specified tolerance limit within which values are considered equivalent. They called this test the equivalence test. On the other hand, Johnston and Duke (2008) argue that the equivalence test is only valid if the WTP distributions can be assumed normal. Given this is not the case in many empirical studies they proposed an extension of the equivalence test, denoted as the two one-sided convolutions (TOSC) test which is statistically valid regardless of the empirical distribution of welfare estimates.[25] This test incorporates the complete combinatorial convolutions approach of Poe et al. (2005) as well as the Kristofferson and Navrud (2005) equivalence test with a null hypothesis of WTP divergence (i.e., H_0: $WTP_{HB} - WTP_{MARL} \neq 0$). In order to implement the test, analysts should choose the tolerance limit of difference between the welfare measures they are willing to accept and calculate the interval of tolerance. This interval is typically set as a percentage of the WTP measured at the study site. Subsequently, the analyst must calculate the differences of the complete combinatorial of the two WTP distributions and test if the resulting difference falls inside or outside the tolerance interval.

Tables 7 and 8 show the TOSC equivalence test results for unadjusted annual mean WTP and CS at $\alpha = 0.10$ using two different tolerance limits (TL) of 50% and 80% for the HB, MARL, HBPOP and MARLPOP samples and treating each as both a study site and as a

[25] In this study, the empirical distributions of welfare estimates are non-normal. The standard "two one-sided *t*-test" (TOST) equivalence test may provide erroneous inference and thus, it is inappropriate (Johnston and Duke, 2008).

Table 8
TOSC equivalence test results for unadjusted annual mean WTP and CS for TL of 80%.

Models	RESORG	RESZERO	WATLOW	WATNO	GHG30	GHGZERO	NAT10	NAT30	CS1	CS2	CS3	CS4
Type 1 BT												
HBPOP vs MARL – P_L	0.2601	0.2635	0.3302	0.4054	0.1982	0.2328	0.8013	0.5406	0.9719	0.9636	0.6906	0.6466
P_U	0.4100	0.1139	0.0006	0.0003	0.0290	0.0170	0.0020	0.0073	0.0179	0.0153	0.0029	0.0008
MARL vs HBPOP – P_L	0.3174	0.1948	**0.0799**	**0.0997**	0.1111	**0.1044**	0.2687	0.2065	0.9453	0.9402	0.9248	0.9406
P_U	0.4781	0.0749	**0.0000**	**0.0000**	0.0114	**0.0054**	0.0000	0.0001	0.0089	0.0087	0.0179	0.0066
HB vs MARLPOP – P_L	0.5371	0.2272	0.1192	0.1485	**0.0332**	0.0254	0.6953	0.1757	0.9636	0.8919	0.5868	0.5912
P_U	0.4142	0.1656	0.0000	0.0000	**0.0228**	0.2393	0.0126	0.0186	0.0340	0.0499	0.0019	0.0001
MARLPOP vs HB – P_L	0.4849	0.1998	**0.0274**	**0.0286**	**0.0333**	0.0671	0.2789	**0.0936**	0.8890	0.8571	0.9155	0.9048
P_U	0.3636	0.1427	**0.0000**	**0.0000**	**0.0229**	0.4388	0.0003	**0.0049**	0.0083	0.0344	0.0326	0.0046
Type 2 BT												
HB vs HBPOP – P_L	0.7355	0.3401	**0.0434**	**0.0381**	**0.0505**	0.4511	0.5785	**0.0495**	0.5684	0.4991	**0.0631**	**0.0237**
P_U	0.2249	0.0802	**0.0002**	**0.0000**	**0.0234**	0.0122	0.0335	**0.0398**	0.4189	0.3352	**0.0664**	**0.0275**
HBPOP vs HB – P_L	0.5299	0.2072	**0.0147**	**0.0128**	**0.0424**	0.1742	0.2648	**0.0488**	0.3388	0.4301	0.3617	0.1832
P_U	0.0974	0.0368	**0.0000**	**0.0000**	**0.0183**	0.0011	0.0045	**0.0390**	0.2059	0.2713	0.4392	0.2741
MARLPOP vs MARL – P_L	0.5085	0.4387	0.1801	0.2062	0.2359	0.9471	0.6930	0.3056	0.7657	0.8012	0.2380	0.1341
P_U	0.3838	0.1130	0.0019	0.0013	0.0236	0.0033	0.0028	0.0178	0.2187	0.1368	0.0596	0.0411
MARL vs MARLPOP – P_L	0.4600	0.2788	**0.0558**	**0.0656**	0.1204	0.4142	0.2205	0.1425	0.6672	0.8009	0.5407	0.4279
P_U	0.3385	0.0527	**0.0006**	**0.0003**	0.0082	0.0002	0.0000	0.0040	0.1435	0.1365	0.2279	0.2068
Type 3 BT												
HB vs MARL – P_L	0.6072	0.4954	0.6431	0.7213	0.2208	0.7706	0.9912	0.5407	0.9938	0.9896	0.7658	0.7333
P_U	0.3534	0.0903	0.0000	0.0000	0.0133	0.0015	0.0000	0.0033	0.0057	0.0041	0.0009	0.0001
MARL vs HB – P_L	0.5055	0.2904	0.1243	0.1524	**0.1087**	0.2473	0.4586	0.1988	0.9786	0.9850	0.9599	0.9447
P_U	0.2638	0.0312	0.0000	0.0000	**0.0027**	0.0000	0.0000	0.0000	0.0014	0.0028	0.0151	0.0028
Type 4 BT												
HBPOP vs MARLPOP – P_L	0.1828	0.0746	**0.0245**	**0.0292**	**0.0336**	0.0016	0.1383	0.1899	0.8915	0.7946	0.5050	0.4975
P_U	0.4711	0.2070	**0.0035**	**0.0023**	**0.0526**	0.3518	0.0558	0.0320	0.0716	0.0990	0.0047	0.0013
MARLPOP vs HBPOP – P_L	0.2816	0.1144	**0.0132**	**0.0129**	**0.0401**	0.0157	**0.1002**	**0.1065**	0.8075	0.7043	0.8584	0.9029
P_U	0.6039	0.2908	**0.0023**	**0.0009**	**0.0629**	0.8032	**0.0375**	**0.0127**	0.0351	0.0587	0.0343	0.0101

p-values lower than 0.10 indicate no differences in the two distributions (bolded).

policy site. For example, if the policymaker is willing to tolerate a 50% difference between the WTP estimated at HB and MARL (Type 3 BT), there is a failure to reject the null hypothesis of different WTP, implying that the two measures cannot be shown to be equivalent, and therefore, transfer is presumed invalid for both marginal value and benefit function transfers. In contrast, for Type 4 BT, it can be noted that marginal values for groundwater quality can be transferred between MARLPOP and HBPOP samples if the policymaker is willing to tolerate a 50% error.

Table 8 shows the TOSC equivalence test results when the policymaker chooses an 80% TL. It is easily seen that as the tolerance limit increases to 80%, more WTP attributes seems to be equivalent; suggesting marginal values transfer is warranted across sites and populations. A further finding of the results indicates that only CS3 and CS4 distributions (Type 2 BT between HB and HBPOP samples) are revealed to be equivalent and suggestive of benefit function transfer.

In summary, the empirical findings from this study suggest that the validity of BT depends on which test analysts are willing to use. There are four types of analyses done in the study: standard transfer error calculation that only takes into account the mean WTP values (as in Table 4), the Swait and Louviere (1993) test (as in Table 5), the Poe et al. (2005) test (as in Table 6) and lastly, the TOSC test (as in Tables 7 and 8).These tests depend on the distribution of the WTP. If the WTP distributions are tight (either because the sampling variance is low or preferences are homogeneous in the sample), then the above tests may provide similar results. If the WTP distributions have a larger spread, then the analyst has to decide which criteria best suit the data at hand and the objectives of the study. For instance, if the objective of a BT study is to determine the transferability of the mean WTP (generally aggregated over the population) with no or little interest in the heterogeneity in the sample, then the standard transfer error calculation can be used.

The Poe et al. (2005) tests may be useful if the analyst is interested in knowing whether the empirical WTP distributions at the study and policy sites have some similarities given that a significant difference exists in the model's parameters. However, it is important to note that the Poe et al. (2005) test can lead to Type II Error null hypothesis, especially if one of the empirical WTP distributions has a large spread. Analysts should be aware of this caveat when accepting the results of Poe et al. (2005) test in Table 6.

On the other hand, the TOSC test is very restrictive and requires the differences between the WTP distributions at the study and policy sites to be fully within certain specified bounds. Although different bounds can be set this test penalizes WTP distributions that have significant spread without recognizing the causes of the spread. In our opinion, this test is more suitable for large sampling variability but not for large variability caused by preference heterogeneity, a fact which seems to be common when valuing environmental impacts. An interesting point to note in this study is that it is able to explain the behavioral and statistical reasons for choosing which BT exercise to employ. From a statistical point of view, it is important to recognize what causes the large spread of the WTP distribution: Is it mainly due to high individual preference heterogeneity or to large sampling variability? If sampling variability exists, a stricter test such as TOSC can be applied to show the limit of tolerance which would be required to accept the transferability of the welfare measures. If the WTP distribution has a wide spread due to individual preference heterogeneity, as is often the case for environmental projects (for instance reversible changes), the use of the standard transfer error criterion (Table 4) may be appropriate. However, analysts need to be cautious about use of the standard criterion if environmental changes are not reversible or when the public health and safety are involved.

3.9. Policy implications

The successful application of benefit transfer remains a challenge. Navrud and Ready (2007), comment that the statistical tests of equivalence (validity tests) may not be the most important criterion for deciding the usefulness of benefit transfer and most researchers are now focusing on the relative size of transfer error. For example, in this study, the TOSC test is applied to examine the tolerance levels of transfer error in a policy context. Although not reported in the study, a 30% TL did not indicate any transferability across sites and populations. As shown in Tables 7 and 8, 50% and 80% TL may be too large to convince policymakers about the merit of BT. On the other hand, the results in Table 4 indicate that for the majority of the attributes the average absolute transfer errors of the mean values of the welfare measures are in the range of 10–40%.

Are these transfer errors small enough for the purposes of the policy analyst? Transferring values from a study site to a policy site necessarily increases the errors in those values. Analysts must judge how to provide policy advice in a timely manner, subject to the resource constraints they face. The levels of transfer error acceptable to the analysts will vary with their experience and professional judgment regarding the advisability of a policy or project. Analysts should compare the cost of doing a new, original valuation study with the potential loss from making the wrong decision when based upon transferred estimates.

In light of this study, and the significance of the external costs of winegrowing, estimates of the benefits likely to be delivered by policy implementation should be directly valued and weighted against policy costs. The results in Table 4 may be particularly useful in policy contexts when average benefits are needed as inputs to an environmental cost benefit analysis. To make this concrete, consider a policy that controls residue content in wine. The GrapeFutures (GF) project is a joint industry and research initiative between Sustainable Winegrowing New Zealand®, New Zealand Winegrowers, Ministry of Agriculture and Forestry Sustainable Farming Fund (SFF), grapegrowers and HortResearch. It aims to address environmental and economic concerns surrounding agrichemical use and to develop pest and disease management strategies to minimize chemical inputs, thereby ensuring that any subsequent wine production is free of detectable pesticide residues. The project costs $938,132 for a total of 3 years (July, 2007–June, 2010) and focuses on Marlborough and Hawke's Bay regions (Ministry of Agriculture and Forestry, MAF, 2009). The estimated WTP values for reducing the wine residues content to zero (RESZERO) can be used to calculate the aggregate benefits of the project to Marlborough and Hawke's Bay households. The mean WTP for RESZERO for MARL and HB are $20 and $11 respectively. For illustrative purposes, we estimate aggregate benefit for Marlborough and Hawke's Bay households for a three year period (assuming no discounting). The aggregate benefits for a total of 3 years to MARL and HB are $986,220 (3 × $20 × 16,437 households) and $1,802,394 (3 × $11 × 54,618 households) respectively, which gives a total of $2,788,614. This amount can be used when judging whether the costs of implementing GF are likely to outweigh the benefits to Marlborough and Hawke's Bay households of reduced residues. Thus, the net social benefit is the sum of the present value of the benefits less the present value of the costs, which is $2,788,614 − $938,132 = $1.9 million.

Suppose there is no original study on WTP for RESZERO in MARL and based on Table 4 results, the analyst is willing to tolerate 45% transfer error from HB to MARL. The aggregate benefit for a total of 3 years to MARL is $542,421 (3 × $11 × 16,437 households) plus $1,802,394 from HB, which gives a total of $2,344,815. The net social benefit for 45% TE is $2,344,815 − $938,132 = $1.4 million. Similarly, if the analyst willing to accept 83% TE from MARL to HB (Table 4), the aggregate benefit for 3 years to HB is $3,277,080 (3 × $20 × 54,618 households) plus $986,220 from MARL, which gives a total of $4,263,300 and hence, the net social benefit is $4,263,300 − $938,132 = $3.3 million.

Nevertheless, caution is advisable before using transferred values for major government policy decisions, in particular concerning public health and safety which effect the general population. In those cases, respondents' heterogeneity does matter and should not be ignored. For example, the heterogeneity of preferences for groundwater quality is considered the most important attribute in both winegrowing regions. Policy decisions such as implementing Public Health Risk Management Plans (Ministry of Health, MOH, New Zealand, 2005) to improve groundwater quality are crucial and in need of urgent attention. Hence careful attention should be given to determining whether to conduct a new study based on primary data or using a benefit transfer approach given that the groundwater quality directly affects public health.

If analysts perform a cost benefit analysis of the policy using BT Type 3 from Table 4 results, they may tend to overestimate (transferring values from MARL to HB) or underestimate (transferring values from HB to MARL) the net present value of the policy. Suppose the estimated groundwater quality values are transferred from Hawke's Bay (study site) to Marlborough (policy site), Hawke's Bay populations' lower WTP estimates may result in underestimation of net benefits from policy change in the Marlborough region. This may lead to large welfare loss and misdirected resources because of the wrong policy decision. In addition, comparing the standard transfer error measures in Table 4 and the TOSC tests (Tables 7 and 8), the policy analyst may find it difficult to make decisions due to ambiguous results. The presence of high preference heterogeneity in the empirical WTP distribution "penalizes" the acceptance of transferability under the TOSC tests. The TOSC tests for BT Type 3 indicate that groundwater WTP are inappropriate to be transferred across these sites and populations, whereas typical transfer errors within "acceptability range" from Table 4 would suggest otherwise. In this case, the BT approach may not be preferred and it is advisable to conduct a new primary study. Indeed, the costs of a new study are likely to be much smaller than the cost of a wrong decision.

In short, when performing a Cost Benefit Analysis (CBA) of a new project or policy, the estimated value of present benefits (costs) should be compared with the corresponding present value of costs (benefits). The effect on total annual benefits (costs) due to an expected general transfer error of 30–80% should be calculated in order to see if this reduces the benefits (increases the costs) to a critical level where the net benefits becomes negative (from positive). If this is so, the transfer errors are large enough to change the outcome of the CBA, and we should try to increase the accuracy of the estimate (either by conducting a full primary study or calibrating the transferred value by conducting a small scale pilot study). If the analyst finds it difficult to decide and is unsure of the BT approach, then there is a need to determine whether the costs of a new primary study is worth the benefits in terms of lower probability of making the wrong decision. These decision rules could be used as a rough test of whether benefit transfer is acceptable or not (Navrud and Brouwer, 2007).

4. Conclusion

This paper focuses on the environmental impact of winegrowing. The specific contribution of this paper is in employing data and results from a study that was designed with BT validity in mind and comparing different ways of testing the validity of BT. This study found a significant portion of the general public lack understanding about winemaking and there is heterogeneity in people's preferences regarding ES that are linked to winegrowing. Despite those limitations, the study found that respondents' value programs which result in a significant total reduction in toxic chemical residue content in wine, a reduction of the risk of toxic substances reaching groundwater, a reduction of greenhouse gas emissions and an increase of natural environment and native wildlife populations in vineyards. The welfare estimation results show that respondents not only experience greater marginal satisfaction from improvements in these selected ES attributes but also are willing to pay more for higher levels of environmental enhancement.

The second purpose of this paper was to validate the transferability of unadjusted value transfer (WTP) and function transfer (CS) using different types of BT tests. Although the statistical tests showed conflicting results of whether to follow the transfer process, the study suggests that assessing how well benefit transfers can predict values at new sites and under which conditions they perform best, depend on the professional judgement of the analyst, the data at hand and the goals of the study. That is, if the objective of a BT study is to gain more knowledge about some benefit at a policy site or provide an initial assessment of the value of policy options (i.e., scoping/screening) then it may be that a relatively low level of accuracy is acceptable. If the analyst moves towards undertaking a transfer study to inform a major government policy decision, in particular concerning public health and safety or compensation litigation then a greater degree of accuracy is arguably desirable. In such cases, either compelling evidence for the validity of benefits transfer needs to exist or an original valuation may be warranted. Therefore, from a practical point of view, the policy analyst may need to consider the trade-off in using BT estimates between the risks of (i) under/over-estimated WTP values across sites and populations, (ii) saving in time and resources of conducting a primary study, and finally, (iii) making costly policy decision mistakes.

Acknowledgements

This research was funded by the New Zealand Foundation for Research, Science and Technology (FRST) Ecosystems project (LINX0303). The authors thank Steve Wratten for his assistance in defining ES attributes and survey design. The authors thank two anonymous reviewers and the editors for their comments and constructive suggestions. Any errors in this article are the authors' responsibility.

References

Bech, M., Gyrd-Hansen, D., 2005. Effects coding in discrete choice experiments. Health Econ. 14, 1079–1083.

Bennett, J., 2006. Choice modelling and the transfer of environmental values. In: Rolfe, J., Bennett, J. (Eds.), Choice Modelling and the Transfer of Environmental Values. Edward Elgar, USA, pp. 1–9.

Colombo, S., Calatrava-Requena, J., Hanley, N., 2007. Testing choice experiment for benefit transfer with preference heterogeneity. American Journal of Agricultural Economics 89 (1), 135–151.

Desvousges, W., Naughton, M., Parsons, G., 1992. Benefit transfer: conceptual problems in estimating water quality benefits using existing studies. Water Resour. Res. 28 (3), 675–683.

Domencich, T., McFadden, D.L., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Company.

Hanemann, W.M., 1984. Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics 66 (3), 332–341.

Hanley, N., Colombo, S., Tinch, D., Black, A., Aftab, A., 2006. Estimating the benefits of water quality improvements under the Water Framework Directive: are benefits transferable? European Review of Agricultural Economics 33 (3), 391–413.

Hensher, D.A., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press, United Kingdom.

Hu, W., Veeman, M., Adamowicz, W., 2005. Labelling genetically modified food: heterogeneous consumer preferences and the value of information. Canadian Journal of Agricultural Economics 53 (1), 79–99.

Johnston, R.J., Duke, J.M., 2008. Benefit transfer equivalence tests with non-normal distributions. Environmental and Resource Economics 41, 1–23.

Kristofferson, D., Navrud, S., 2005. Validity tests of benefit transfer: are we performing the wrong tests? Environmental and Resource Economics 30 (3), 279–286.

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Application. Cambridge University Press, Cambridge, United Kingdom.

Ministry of Agriculture and Forestry (MAF), 2009. Implementing Ultra Low Residue Wine Grape Production (GrapeSafe), Sustainable Farming Fund (Grant No.: 07/123). Retrieved at http://www.maf.govt.nz/sff/about-projects/search/07-123/index.htm#update.

Ministry of Health (MOH), NewZealand, 2005. Public Health Risk Management Plans. Retrieved at http://www.moh.govt.nz.

Morrison, M., Bergland, O., 2006. Prospects for the use of choice modelling for benefit transfer. Ecological Economics 60, 420–428.

Meyerhoff, J., Liebe, U., 2008. Do protest responses to a contingent valuation question and a choice experiment differ? Environmental and Resource Economics 39, 433–446.

Navrud, S., Ready, R., 2007. Lessons learned for environmental value transfer. In: Ready, Navrud (Eds.), Environmental Value Transfer: Issues and Methods. Springer, The Netherlands, pp. 283–290.

Navrud, S., Brouwer, R., 2007. Good practice guidelines in benefit transfer of forest externalities. COST Action E45, European Forest Externalities (EUROFOREX). Retrieved at http://www.medforex.net/E45/Guidelines/BT%20guidelines%20281107.pdf.

Ruud, P., 1996. Simulation of the multinomial probit model: an analysis of covariance matrix estimation. Working Paper. Department of Economics, University of California.

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational methods for measuring the difference of empirical distributions. American Journal of Agricultural Economics 89 (2), 353–365.

Street, D.J., Burgess, L., Louviere, J.J., 2005. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing 22, 459–470.

Street, D.J., Burgess, L., 2007. The construction of optimal stated choice experiments, theory and methods. Wiley Interscience, New Jersey, USA.

Swait, J., Louviere, J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 30, 305–314.

Swait, J., 2007. Advance choice models. In: Kanninen, B.J. (Ed.), Valuing Environmental Amenities Using Stated Choice Studies. Springer, The Netherlands, pp. 229–293.

Swinton, S.M., Lupi, F., Robertson, G.P., Hamilton, S.K., 2007. Ecosystem services and agriculture: cultivating agricultural ecosystems for diverse benefits. Ecological Economics 64, 245–252.

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

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, United Kingdom.

van Bueren, M., Bennett, J., 2006. Generalising environmental values: the case of the national land and water resources audit in Australia. In: Bennett, J., Rolfe, J. (Eds.), Choice Modelling and the Transfer of Environmental Values. Edward Elgar, USA, pp. 191–215.