Environ Resource Econ (2011) 48:1–23
DOI 10.1007/s10640-010-9394-1

# External Tests of Scope and Embedding in Stated Preference Choice Experiments: An Application to Endangered Species Valuation

**Daniel K. Lew · Kristy Wallmo**

Accepted: 18 July 2010 / Published online: 7 August 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** A criticism often levied against stated preference (SP) valuation results is that they sometimes do not display sensitivity to differences in the magnitude or scope of the good being valued. In this study, we test the sensitivity of preferences for several proposed expanded protection programs that would protect up to three US Endangered Species Act-listed species: the Puget Sound Chinook salmon, the smalltooth sawfish, and the Hawaiian monk seal. An external scope test is employed via a split-sample SP choice experiment survey to evaluate whether there is a significant difference in willingness to pay (WTP) for protecting more species and/or achieving greater improvements in the status of the species. The majority of 46 scope tests indicate sensitivity to scope, and the pattern of scope test failures is consistent with diminishing marginal utility with respect to the amount of protection

The authors would like to thank Rita Curtis, Todd Lee, Vic Adamowicz, George Parsons, John Whitehead, Rob Hicks, Ron Felthoven, Michelle McGregor, John Horowitz, an anonymous reviewer, and participants at the 2009 Western Agricultural Economics Association meeting for useful comments on this work. All remaining errors, if any, are our own. This research was conducted under Pacific States Marine Fisheries Commission grant #NA04NMF4370384. The views and opinions expressed in this paper are the authors' own, and do not necessarily reflect those of the National Oceanic and Atmospheric Administration or the US Department of Commerce.

D. K. Lew
Resource Ecology and Fisheries Management Alaska Fisheries Science Center,
National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA 98115, USA

D. K. Lew (✉)
Department of Agricultural and Resource Economics, University of California,
One Shields Avenue, Davis, CA 95616, USA
e-mail: Dan.Lew@noaa.gov

K. Wallmo
Office of Science and Technology, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 1315 East-West Hwy, Silver Spring, MD 20910, USA

K. Wallmo
University of Maryland, Eastern Shore, Princess Anne, MD 21853, USA

to each species. Further tests suggest WTP may be proportional to the number of species valued.

**Keywords** Choice experiments · Endangered fish · US Endangered Species Act · Marine mammals · Scope tests · Stated preference methods · Willingness to pay

**Abbreviations**

| | |
|---|---|
| CV | Contingent valuation |
| SP | Stated preference |
| SPCE | Stated preference choice experiment |
| WTP | Willingness to pay |
| HMS | Hawaiian monk seal |
| SS | Smalltooth sawfish |
| PSC | Puget sound chinook |
| RPL | Random parameters logit |
| TEV | Type I extreme value |
| E | Endangered status |
| T | Threatened status |
| R | Recovered status |

## 1 Introduction

A criticism often levied against stated preference (SP) valuation results, particularly those arising from contingent valuation (CV) studies, is that the results do not display sensitivity to differences in the magnitude or scope of the good being valued as one would expect from economic theory. The National Oceanic and Atmospheric Administration Panel on Contingent Valuation (Arrow et al. 1993) emphasized the need for researchers to test for responsiveness to scope in stated preference studies, CV particularly, to ensure the reliability of the results. The Panel viewed scope insensitivity, which they referred to as "the embedding phenomenon", as the "most important internal argument against the reliability of the CV approach." (p. 25)

Concern over scope in SP studies can be traced back to Kahneman (1986), who called into question the reliability of CV-based estimates of value. His concern was based on a CV study that found the estimate of the value for preserving fish stocks in all lakes in Ontario, Canada, was not much larger than the estimate of the value of preserving fish stocks in the lakes of one small province of Ontario. He argued that this provided evidence that respondents were indifferent to the size or amount of the non-market good being valued. This "scope insensitivity" hypothesis was furthered by Kahneman and Knetsch (1992) who proposed that respondents to CV questions were not purchasing the good intended to be valued, but were expressing the "moral satisfaction" associated with contributing to the provision of the good or what it represents (its "symbolic value"). Several studies in the early 1990s provided some empirical evidence for the scope insensitivity hypothesis (Diamond et al. 1993; Desvousges et al. 1993 and Schkade and Payne 1994).

Proponents of SP methods argued that the conclusions of most of these early tests of scope insensitivity should be viewed skeptically (Carson 1997; Carson and Mitchell 1993; Smith 1992), drawing in large part upon survey design and implementation shortcomings that weakened the scope insensitivity argument, or interpretative errors that could reverse the conclusions. Evidence from the empirical literature has also been presented to support the contention that scope insensitivity is more often the exception rather than the rule in carefully-conducted CV studies (Carson and Mitchell 1995; Carson 1997; Smith and Osborne 1996).

Moreover, there is an explanation for observed scope insensitivity in economic theory. Rollins and Lyke (1998) showed that in the context of CV, scope tests may find scope insensitivity due to diminishing marginal utility. They argue that scope tests conducted over the upper range of a public good being valued may fail to identify scope effects because the comparisons are made over the range of the good where marginal utility approaches zero. Large sample sizes would be required to identify significant differences over this range, something most samples used in scope tests cannot handle. Several CV-based scope tests have provided evidence of diminishing marginal values for public goods (e.g., Bateman et al. 2005). Nevertheless, concerns over scope insensitivity in SP studies remain. As alternatives to CV, like stated preference choice experiments and contingent ranking and rating methods, are used to value non-market goods, it is important to evaluate scope insensitivity in SP results using these methods.

We used results from a stated preference choice experiment (SPCE) to test for scope insensitivity concerning the number of species being protected and the amount of protection afforded to each species. The hypothetical scenarios in the SPCE focus on expanded protection programs for three US Endangered Species Act (ESA)-listed marine species: the threatened Puget Sound Chinook salmon (*Oncorhynchus tshawytscha*), the endangered smalltooth sawfish (*Pristis pectinata*), and the endangered Hawaiian monk seal (*Monachus schauinslandi*).[1] Within the SPCE we conducted a split-sample test using two survey versions, one that presents information about and asks respondents to value protecting up to two species (the smalltooth sawfish and Hawaiian monk seal), and another survey that includes all three species. The tests employed were intended to evaluate whether the willingness to pay (WTP) for protecting more species, in some cases with greater levels of protection, from the three-species survey is greater than the WTP for protecting fewer species receiving equal or less protection from the two-species survey. Thus, the WTP comparisons are between protection alternatives in the two-species version and clearly dominating alternatives in the three-species version. Tests to evaluate whether the estimated two-species and three-species models lead to commensurate welfare measures for equivalent scenarios were also conducted and support consistency in preferences being measured.

Stated preference choice experiments are increasingly being used in non-market valuation applications, including valuation of threatened and endangered species protection (Layton and Levine 2005; Olar et al. 2007; Rudd 2009; Lew et al. 2010). Given the increased usage of the method to value non-market goods and services, it seems prudent to explore scope insensitivity issues in the context of SPCE results. From a policy perspective, one advantage of using a SPCE approach is its ability to generate numerous welfare comparisons since the estimated utilities are functions of the good for which sensitivity to scope is being evaluated. In this application, we exploit this feature and conduct 46 evaluations that test for sensitivity to scope with respect to both the number of species whose protection is valued and the level of protection achieved for each species. Thus, we evaluate scope effects along a significant portion of the multi-dimensional utility surface, which lends itself to an evaluation of the

[1] The US Endangered Species Act of 1973 defines two types of at-risk species. An "endangered" species is "any species which is in danger of extinction throughout all or a significant portion of its range" (Section 3.6). A "threatened" species is "any species which is likely to become an endangered species within the foreseeable future throughout all or a significant portion of its range" (Section 3.20).

argument for scope insensitivity being observed as a result of diminishing marginal utility made by Rollins and Lyke (1998).

Bateman et al. (2002) distinguish between scope tests, which measure differences in WTP resulting from a change in one argument of the utility function, and embedding tests, which measure differences in WTP resulting from a change in multiple arguments of the utility function. In this study, we conduct both scope and embedding tests since we measure (1) differences in WTP across samples for changes to solely species improvement levels (scope tests), (2) only the number of species (scope tests), and (3) both simultaneously (embedding tests), with the last ones being comparisons of perfectly embedded goods (Carson and Mitchell 1995). Henceforward, we refer to both embedding and scope tests as "scope tests" for simplicity.

This paper makes several significant contributions to the valuation literature. To our knowledge, it is the first study to use split-sample based scope tests with SPCE methods. It is also the first split-sample scope test involving threatened and endangered species to assess scope along multiple dimensions; in this case the number of species and the amount of protection for each species. In addition, our scope tests are based on estimated welfare estimates from random parameters logit, or mixed logit, models (Train 2003). In recent years, researchers have increasingly used these models in non-market valuation applications due to the advantages they have over the conditional logit models that are most often turned to for estimating discrete choice data (e.g., Train 1998; Layton 2000; Siikamaki and Layton 2007). In particular, mixed logit probabilities do not exhibit the restrictive Independence from Irrelevant Alternatives property, preferences are not restricted to be the same for everyone in the population, and with increased computing power the model is straightforward to implement for large choice sets using simulated maximum likelihood techniques. This study appears to offer the first scope test conducted using mixed logit model results. Finally, although SPCE methods have been increasingly applied in non-market valuation since its first application in the area by Adamowicz et al. (1994),[2] only recently have they been utilized to value protecting threatened or endangered species. In fact, to our knowledge, only four studies employ SPCE methods to value threatened or endangered species.[3] Layton and Levine (2005) estimate the value of protecting the northern spotted owl using data from a sample of Seattle residents, Olar et al. (2007) and Rudd (2009) both estimate the value of several at-risk aquatic species in Canada using data collected via Internet methods, and Lew et al. (2010) value the provision of additional protection to the Steller sea lion using data from a national mail-based survey. Our work adds to this important yet currently limited field of the literature.

The remainder of the paper is organized as follows. In the next section, we discuss several recent scope tests related to the valuation of protecting threatened, endangered, or other at-risk species. This is followed by a description of the survey data and methods used. The model results are then presented. The subsequent section describes the protection scenarios for which willingness to pay estimates are generated and presents the estimates themselves. A test to evaluate the consistency of the two-species model welfare estimates with the three-species model welfare estimates is presented and followed by the formal scope test results. Then, we present a stronger external scope test that assesses scope effects with respect to only the number of species improved. Finally, a discussion of the results of all scope tests concludes the article.

[2] See, for example, Hanley et al. (1998); Adamowicz et al. (1998) and Alpizar et al. (2003).

[3] In a meta-analysis of threatened and endangered species non-market valuation studies, Richardson and Loomis (2009) note that only one study (Layton et al. 2001) did not use the CV method.

## 2 Scope Tests Involving Valuation of Threatened and Endangered Species Conservation

There are two types of scope tests: internal (or within-sample) and external (or between-sample). The former involves tests comparing multiple WTP estimates from SP responses collected from the same respondents, while the latter employs split-sample designs to compare WTP estimates across samples from the same population. We focus primarily on external scope tests.

Threatened and endangered species-based scope tests have primarily been concerned with evaluating scope effects in relation to the number of species preserved or protected within CV applications. These tests are exemplified by Berrens et al. (1996, 2000); Loomis and Ekstrand (1997); Giraud et al. (1999) and Veisten et al. (2004). Berrens et al. (1996) value the protection of a single endangered species (the silvery minnow) and the protection of an aggregation of 11 threatened and endangered fish species found in New Mexico (including the silvery minnow) using referendum CV data. Their external scope test suggests a higher willingness to pay for the protection of 11 species compared to the WTP for only protecting the silvery minnow. In a follow-up analysis using additional data collected with the same valuation questions a year later, Berrens et al. (2000) reconfirmed the presence of external scope sensitivity. Loomis and Ekstrand (1997) used multiple-bounded CV data to evaluate whether the WTP for protecting the Mexican spotted owl is less than the WTP for protecting 62 threatened and endangered species including the Mexican spotted owl. Giraud et al. (1999) performed a parallel analysis to Loomis and Ekstrand (1997) using referendum CV questions. In both analyses, the data were found to support external scope sensitivity, with higher WTP values for protecting the 62 species compared to protecting only the single species. However, in the Giraud et al. study, an internal scope evaluation based on estimation of bivariate probit models that explicitly account for correlation between two CV question responses could not reject scope insensitivity. Given the survey was administered by mail, and respondents had the ability to go back and change their responses to make them consistent, it is surprising external scope insensitivity was rejected while internal scope insensitivity was not. As a possible explanation for this result, the authors suggest sequencing effects (Carson 1997; Carson et al. 1998) may have affected the results, with respondents anchoring responses to the first CV question presented. Bateman et al. (2004) provide some empirical evidence that the ordering of CV questions can have an effect on the results of scope tests.

Veisten et al. (2004) conducted both external and internal scope tests using data from four surveys that contain either open-ended or payment card CV questions asking respondents to value one or more scenarios involving the protection of the following: the white-backed woodpecker in Norway, all endangered plants and animals in Norwegian forests, a subset of all endangered plants and animals in Norwegian forests consisting of endangered fungi, lichen, and mosses; and a comprehensive package of environmental projects that includes species protection. Their results with respect to internal and external scope insensitivity are mixed—three of five internal scope tests reject scope insensitivity and three of four external scope tests reject scope insensitivity. However, there is some evidence that suggests the test results are sensitive to CV elicitation format, with payment card-based CV results indicating rejection of scope insensitivity and open-ended CV results failing to reject scope insensitivity. The referendum CV format tends to be favored over open-ended and payment card CV questions due to incentive compatibility concerns (e.g., Arrow et al. 1993; Carson and Groves 2007). Additionally, in a health risk application, Bateman and Brouwer (2006) provide evidence that opposite conclusions about scope insensitivity can be reached resulting from the use of referendum CV data compared with open-ended CV data, suggesting that response biases, such as those associated with open-ended and payment card question formats, may confound scope test results.

## 3 Methods

Note that the threatened and endangered species studies discussed above focus solely on testing for scope effects related to the number of species being protected and thus ignore consideration of the amount of protection being provided to the species, treating this dimension as achieving some uniform protection level, such as recovery. However, people may also have preferences for differing amounts of improvements to each threatened and endangered species' population size (Loomis and Larson 1994), ESA status (Lew et al. 2010), or chances of survival (Reaves et al. 1999).

In this study, the SPCE responses from the two-species survey that includes the endangered Hawaiian monk seal and endangered smalltooth sawfish, and the three-species survey that includes both these species plus the threatened Puget Sound Chinook salmon, are used to estimate a preference function that depends upon the levels of policy-relevant attributes (ESA status levels in this case) for each species, and thus provides a more flexible tool for decision makers compared to most CV-based welfare estimates associated with a specific good being valued (e.g., a specific protection program). In this study, welfare estimates and associated confidence intervals are calculated for several species status improvements. $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y)$ denotes the willingness to pay calculated using the two-species survey version model and data for a protection alternative that leads to the Hawaiian monk seal (HMS) achieving an ESA status of $x$ and the smalltooth sawfish (SS) achieving an ESA status of $y$, where $x, y$ = endangered (E), threatened (T), or recovered (R).[4] Since the two-species version of the survey does not contain specific information about the Puget Sound Chinook salmon (PSC), $\mathrm{WTP}_2(\cdot)$ implicitly assumes PSC remains threatened. $\mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = z)$ is the estimate of willingness to pay from the three-species version for HMS and SS achieving ESA status levels of $x$ and $y$, respectively, and PSC reaching an ESA status level of $z$ in the time period, where $z$ = threatened or endangered. Differences between two-species and three-species welfare estimates ($\mathrm{WTP}_2 - \mathrm{WTP}_3$) are assessed by calculating the confidence intervals around this difference using the method of convolutions approach (Poe et al. 2005). Insensitivity to scope, specifically, is tested for three types of WTP comparisons.

The first type of comparison, Type 1, compares the differences between the WTP associated with improving the status of the SS and HMS in the two-species survey and the WTP associated with improving the status of the SS and HMS in the three-species version by the same amount, but also improving the PSC salmon's status. That is, we examine the difference between $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y)$ and $\mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = \mathrm{R})$. The only difference between the two welfare measures is that the improvement to the PSC (to a recovered status) is accounted for in the three-species version WTP estimates. Type 1 comparisons test for scope insensitivity in relation to the number of species being valued, holding the amount of protection afforded the species common to both welfare estimates constant. There are a total of 8 Type 1 comparisons.

Type 2 comparisons test for scope insensitivity with respect to the amount of protection afforded to each species included in the survey. Thus, the comparison is between the WTP

[4] For the purposes of the survey, a "recovered" species was defined as a species whose population increases sufficiently to be removed from the list of threatened and endangered species.

for protecting the HMS and/or the SS in the two-species version and the WTP for a dominant protection alternative with respect to HMS and/or SS in the three-species version holding the PSC at a threatened level, which is its implied ESA status level in the two-species version. In other words, a Type 2 comparison is between $\text{WTP}_2(\text{HMS} = x^0, \text{SS} = y^0)$ and $\text{WTP}_3(\text{HMS} = x^1, \text{SS} = y^1, \text{PSC} = \text{T})$, where $x^1$ and $y^1$ are ESA status levels equal to or better than $x^0$ and $y^0$, respectively. There are 19 Type 2 comparisons.

The final comparisons, Type 3, test for scope insensitivity along both dimensions. That is, these 19 comparisons compare each welfare estimate from the three-species version that has a dominating alternative both in terms of status improvements to individual species and the number of species protected relative to the two-species version WTP. Thus, the Type 3 comparison looks at the difference between $\text{WTP}_2(\text{HMS} = x^0, \text{SS} = y^0)$ and $\text{WTP}_3(\text{HMS} = x^1, \text{SS} = y^1, \text{PSC} = \text{R})$.

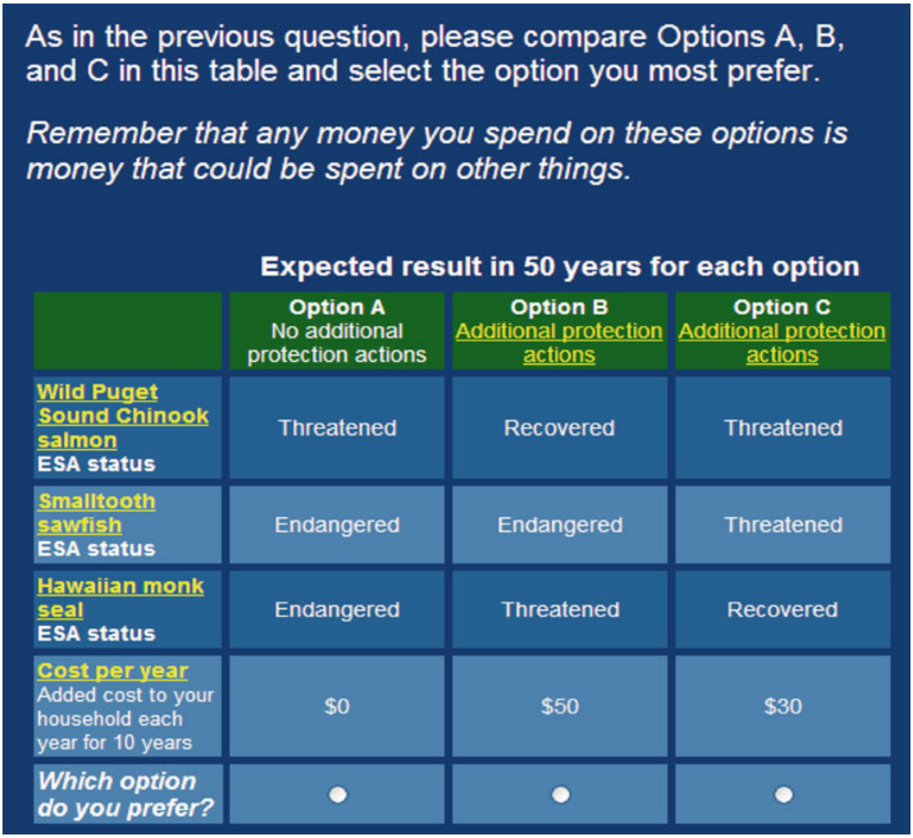
## 4 Data

Scope test results are dependent in large measure on the quality and administration of the particular SP survey instrument used to elicit preferences. Survey design, choice of elicitation format, model specification and analysis, and survey administration play important roles in the generation of defensible SP value estimates, which are precursors to valid scope tests. Carson and Mitchell (1995) argue that poor survey design, and amenity misspecification specifically, represents a major reason for questioning the results of many empirical scope tests. In species-related scope studies, a concern related to survey design is the amount and balance of information provided about each species or group of species being valued, which is part of the specification of the amenity. In all of the species valuation scope studies discussed above, it is not clear that respondents are provided a balanced amount of information about each amenity to be valued.

With this in mind, the two-species and three-species survey instruments used in this study both ask questions about and provide background information on the US Endangered Species Act and two pages of information about each species included in the survey, including the historical and current population trends and habitat, as well as past, present, and potential future protection actions. Biologists studying the species reviewed the information for accuracy. Following the choice question instructions were a budget reminder, a “cheap talk” script to deflate any potential hypothetical bias (Cummings and Taylor 1999),[5] and three SPCE questions. A series of focus groups and cognitive interviews held in cities across the United States were used to thoroughly test and refine the survey instruments.

In each choice question, respondents are asked to choose between three protection options (see Fig. 1). Each option is described in terms of the results the protection would have on the ESA status of the species included in the survey in 50 years, and the increased cost to the respondents’ household of providing the protection. This payment vehicle is described in terms of additional costs resulting from taxes and increased costs of goods and services

[5] The specific cheap talk script is the following: “For hypothetical questions like these, studies have shown that many people say they are willing to pay more for protecting threatened and endangered species than they actually would pay out of their pockets. We believe this happens because people do not really consider how big an impact an extra cost actually has to their family’s budget when answering these types of questions. It is easy to be generous when you do not really need to open your wallet. To avoid this, as you consider each question, please imagine your household actually paying the cost of the choice you select out of your household’s budget.”


As in the previous question, please compare Options A, B, and C in this table and select the option you most prefer.

*Remember that any money you spend on these options is money that could be spent on other things.*

**Expected result in 50 years for each option**

| | **Option A** No additional protection actions | **Option B** Additional protection actions | **Option C** Additional protection actions |
|---|---|---|---|
| **Wild Puget Sound Chinook salmon ESA status** | Threatened | Recovered | Threatened |
| **Smalltooth sawfish ESA status** | Endangered | Endangered | Threatened |
| **Hawaiian monk seal ESA status** | Endangered | Threatened | Recovered |
| **Cost per year** Added cost to your household each year for 10 years | $0 | $50 | $30 |
| ***Which option do you prefer?*** | ○ | ○ | ○ |

**Fig. 1** Stated preference choice experiment question

that the households buy affected by protection actions.[6] In each of the three SPCE questions in the surveys, one option is always the status quo level of protection, which results in no improvements to the ESA status of any of the species. The other two options do more to protect the species, but cost more as well.

The experimental designs for the two-species and three-species survey versions—i.e., the attribute levels seen in each survey—were statistically-based, account for main effects, and maximize a D-efficiency criterion (see Louviere et al. 2000).[7] For the experimental design, there are a total of six possible non-zero costs (plus zero). The Hawaiian monk seal (HMS) and smalltooth sawfish (SS) can both take on one of three ESA status levels—endangered, threatened, or recovered—since at present they are both endangered and only improvements

[6] Specifically, the payment vehicle is described as an increase in costs to the household resulting from "higher prices for products and services affected by the protection actions" and "increases in taxes." Respondents are told the annual household costs will occur for the "first 10 years when new regulations and restrictions are put into place, government funds are spent on research and purchases like land for wildlife refuges, and those directly affected by restrictions adjust to the changes." The payment vehicle was tested extensively in focus groups.

[7] The designs were selected using a D-efficiency based algorithm written in GAUSS that assumes a main effects specification and an underlying conditional logit model. The two-species and three-species survey designs have D-efficiency levels of 97.4 and 98.3%, respectively.

**Table 1** Sample statistics

| Characteristic (Census) | Two-species version | | | Three-species version | | |
|---|---|---|---|---|---|---|
| | Mean | Median | SD | Mean | Median | SD |
| Percent male (49%) | 50.9% | | | 47.8% | | |
| Age (36.4 yrs) | 48.2 | 51 | 16.8 | 48.0 | 49 | 16.7 |
| Household size (2.6) | 2.6 | 2 | 1.4 | 2.7 | 2 | 1.4 |
| % White/Caucasian (73.9%) | 72.1% | | | 73.6% | | |
| Household income ($48,201) | $60,740 | $55,000 | $45,934 | $59,519 | $55,000 | $39,910 |

*SD* Standard deviation

to their status are considered in the choice questions. The Puget Sound Chinook (PSC) salmon, on the other hand, is presently at a threatened level, so it can only take one of two possible ESA status levels, threatened or endangered.

Following the choice questions, several debriefing questions were asked to identify protest respondents and gauge each respondent's level of confidence about how they answered the choice questions.

A total of 9 two-species and 18 three-species survey versions were implemented using an Internet-based survey approach. The survey was implemented in December 2008 by Knowledge Networks (KN) utilizing a random sample of the KN web panel of US households and followed a Dillman Tailored Design Method approach (Dillman 2000).[8] A total of 1,120 surveys were fielded—745 three-species surveys and 375 two-species surveys. The cooperation rate for the overall survey was 62% for a total of 699 completed surveys (234 completes from the two-species and 465 from the three-species). The two samples are very similar with respect to key demographics (see Table 1) including age, gender, household size, income, and ethnicity distributions. Compared to the US population, however, both samples tended to be older and wealthier.

## 5 Estimation Model

We analyze the stated preference choice experiment data using random utility maximization-based discrete choice econometric models, wherein the conditional indirect utility of the $j$th choice alternative ($U_j$) is assumed to be composed of an observable deterministic component ($V_j$) and a stochastic component ($\varepsilon_j$) that is known to the individual but not the researcher. In this application, $V_j$ is assumed to be a linear function of the cost of the protection option and each species' ESA status levels achieved with the option (endangered, threatened, or recovered), which are represented by effects-coded variables instead of dummy variables (Louviere et al. 2000). The advantage of effects-coded variables over dummies is the ability to recover the marginal utility (in a main effects model) of the baseline level that would otherwise be dropped using dummy variables. Our assumed baseline level for the ESA status variables is the status quo ESA level, which is endangered for the SS and HMS, and threatened for the PSC. Thus, there are two effects-coded variables for the SS and HMS, but

[8] Knowledge Networks maintains a large probability-based web panel of US households. Our sample was randomly drawn from this panel. Respondents were first notified by an e-mail letting them know a survey was available for them to complete and where to find the survey (i.e., a web link). Follow-up contacts included e-mail reminders and a telephone reminder. For more information on Knowledge Networks, its panel, and survey protocols, see http://www.knowledgenetworks.com/ganp/index.html.

only one for the PSC. As a result, we expect that, *a priori*, the utility parameters associated with the ESA status levels will be positive, given we assume improvements to the species yield positive marginal utility, all else being equal.

For a given choice question, individuals choose the alternative that yields the highest utility from among the $J$ choices in the choice set. Here, $J = 3$ with corresponding choice alternatives A, B, and C, so we can model the probability that the individual chooses the $j$th alternative as Pr[choose $j$] = $\Pr[\mathrm{V}_j \geq \max\{\mathrm{V_A}, \mathrm{V_B}, \mathrm{V_C}\}]$. A common assumption in analyzing SPCE data is to assume $\varepsilon_j$ is independent and identically distributed Type I extreme value (TEV).[9] This leads to the familiar conditional logit model and probabilities of the form:

$$\Pr[\text{choose } j] = \exp\left(\mathrm{V}_j\right) / \left[\exp\left(\mathrm{V_A}\right) + \exp\left(\mathrm{V_B}\right) + \exp\left(\mathrm{V_C}\right)\right] \qquad j = \mathrm{A, B, C}. \qquad (1)$$

The logit probabilities in Eq. (1) exhibit the Independence from Irrelevant Alternatives property, which restricts the substitution patterns. To relax this property, we employ the mixed, or random parameters, logit (RPL) model (Train 2003). As in the conditional logit model, the unobserved component of conditional indirect utility is assumed to be distributed as a TEV error term in the RPL model. However, this model assumes utility parameters are distributed continuously over the population instead of being fixed over the population as they are assumed to be in the conditional logit. The probabilities in a RPL model are thus evaluated over the parameter distributions and do not display the IIA property:

$$\Pr[\text{choose } j] = \pi_j = \int \left\{\exp(\mathrm{V}_j(\boldsymbol{\beta}))/\Sigma_i \exp(\mathrm{V}_i(\boldsymbol{\beta}))\right\} f(\boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}, \qquad (2)$$

for all $j$ and where $f(\boldsymbol{\beta})$ is the probability distribution of the utility parameters $\boldsymbol{\beta}$.

These probabilities are approximated through simulation as follows: $R$ draws of $\boldsymbol{\beta}$ are taken from $f(\boldsymbol{\beta})$, and the conditional choice probabilities are evaluated at each draw. The simulated probability of choosing the $j$th alternative ($\pi_j^s$) is the mean over the $R$ draws:

$$\pi_j^s = (R^{-1}) \cdot \sum_{r=1}^{R} \frac{\exp(V_j(\beta^r))}{\sum_k \exp(V_k(\beta^r))} \qquad (3)$$

where $\boldsymbol{\beta}^r$ is the $r$th coefficient vector draw from the mixing distribution, $f(\boldsymbol{\beta})$. In this application, we assume the mixing distribution is a multivariate normal distribution.

Given the fact that respondents are faced with three choice questions in the survey, we model the joint probability of observing the sequence of choices an individual makes as the product of individual choice probabilities (e.g., Morey et al. 1993). This assumes the TEV errors are independent across the repeated choices.

$$\Pr[j, k, l] = \pi_j^s \cdot \pi_k^s \cdot \pi_l^s, \qquad (4)$$

where $j$ is selected in the first question, $k$ is selected in the second question, and $l$ is selected in the third, and $j$, $k$, and $l \in \{\mathrm{A, B, C}\}$. This repeated random parameters logit model is separately estimated for the two-species version data and the three-species version data using simulated maximum likelihood with 150 randomized Halton draws (Train 2003) in GAUSS.

9 Choice-based experimental designs are often chosen to maximize D-efficiency assuming an underlying conditional or multinomial logit model (see Huber and Zwerina 1996, for example), which provides one possible reason for selecting it for estimation.

## 6 Estimation Results

Two models were estimated: one using the two-species version data and the other using the three-species version data.[10] For estimation, individuals not answering any choice questions, protest respondents, and those stating they are "not at all confident" in their responses to the choice questions were excluded. This resulted in estimation sample sizes of 169 for the two-species model and 364 for the three-species model. Parameter estimates and model fit statistics for the two-species and three-species repeated random parameters logit models are presented in Table 2. Both models have a likelihood ratio index, a measure of goodness-of-fit, of approximately 0.265.

In these models, all ESA status level-related effects-coded variables are assumed to have random parameters associated with them, which imply estimated mean and standard deviation parameters. The full set of Choleski off-diagonal parameters are reported for each model, as the random parameters are allowed to be correlated.[11] Statistical significance of several off-diagonal parameters suggests correlation between some random parameters. Mean parameters associated with the threatened and recovered status levels for the two endangered species (SS and HMS) are positive and statistically significant at the 5% level. The signs of these parameters conform to our expectation that individuals (on average) yield positive marginal utility from improvements in status level from the status quo endangered level and that improvements to a recovered status are worth more than improvements to a threatened level, *ceteris paribus*. Preferences for the SS appear to vary significantly over the population, as the standard deviations of the SS parameters are statistically significant (and large) in both models. The magnitude of the variation suggests some individuals may have a disutility from improving the SS ESA status, while others get a much larger utility gain from the same improvement. In both the two-species and three-species models, preferences do not seem to vary across the population for the HMS. The parameter associated with a recovered level for the PSC in the three-species model is positive and statistically significant indicating that, all else being equal, individuals get more utility when the PSC is recovered compared to when it is at its current threatened status. Like for the SS, preferences for improving the PSC appear to vary considerably over the population. For both models, the cost parameters are negative and statistically significant at conventional levels, which is consistent with our expectations from economic theory.

Although not the focus of this paper, it is worth noting that internal scope tests on each of these two models were conducted to test whether the marginal utility of improving the HMS or SS from an endangered to threatened level is the same as the marginal utility of improving the same species from endangered to recovered. Likelihood ratio tests indicate the null hypothesis of equal marginal utilities (internal scope insensitivity) can be rejected at all conventional levels of significance, and thus support internal scope sensitivity.

[10] Additional models that pooled the data while accounting for scale differences were estimated, as well as ones accounting for potential demographic differences in preferences. The pooled models indicated the relative scale factor is significant and different from one while the demographic-based models indicated little variation in preferences across different demographic types. The latter result suggests an analysis along the lines of Leiter and Pruckner (2009) may not be warranted.

[11] Random parameters logit models that assume univariate random parameters were also estimated and are qualitatively similar. Those results are available upon request from the authors.

**Table 2** Estimation results

| Parameter | Repeated random parameters logit model | | | |
|---|---|---|---|---|
| | Three-species model | | Two-species model | |
| | Estimate | *T*-value | Estimate | *T*-value |
| RECPSC (PSC recovered) | 0.9114 | 7.2841 | | |
| THRSS (SS threatened) | 0.1910 | 1.9095 | 0.5876 | 3.1398 |
| RECSS (SS recovered) | 0.9424 | 5.8641 | 1.3203 | 3.2872 |
| THRHMS (HMS threatened) | 0.2507 | 2.3334 | 0.09993 | 0.5343 |
| RECHMS (HMS recovered) | 1.2028 | 6.4989 | 2.07276 | 4.6757 |
| COST | −0.0390 | −8.2688 | −0.04958 | −5.4576 |
| SD(PSC3) | 0.8816 | 5.4772 | | |
| CHOL(PSC3,SS2) | 0.009157 | 0.04777 | | |
| SD(SS2) | 0.6108 | 2.6652 | 0.7466 | 2.0017 |
| CHOL(PSC3,SS3) | 0.2521 | 1.0313 | | |
| CHOL(SS2,SS3) | −0.4140 | −1.3537 | −0.1705 | −0.2233 |
| SD(SS3) | 0.8960 | 4.7435 | 2.3830 | 4.6243 |
| CHOL(PSC3,HMS2) | −0.1637 | −0.7925 | | |
| CHOL(SS2,HMS2) | 0.5151 | 2.3582 | 0.9530 | 1.9024 |
| CHOL(SS3,HMS2) | −0.1865 | −0.7494 | −0.5436 | −1.3866 |
| SD(HMS2) | 0.4276 | 1.4791 | −0.5138 | −0.7591 |
| CHOL(PSC3,HMS3) | 0.3310 | 1.1428 | | |
| CHOL(SS2,HMS3) | 0.02918 | 0.08874 | −0.05152 | −0.07078 |
| CHOL(SS3,HMS3) | 1.0209 | 3.3813 | 1.5894 | 4.2459 |
| CHOL(HMS2,HMS3) | −0.8836 | −1.9005 | −0.5274 | −0.5023 |
| SD(HMS3) | −0.6861 | −1.4872 | 0.8310 | 1.02759 |
| LRI | 0.2655 | | 0.2624 | |
| AICc | 1822.2 | | 888. 5 | |
| BIC | 1886.2 | | 898.7 | |
| Sample size | 364 | | 169 | |

Chol$(x, y)$ is the off-diagonal element of the lower diagonal Cholesky matrix associated with the $x$th and $y$th species' status dummy variable. *PSC* Puget Sound Chinook salmon, *SS* Smalltooth sawfish, and *HMS* Hawaiian monk seal

## 7 Welfare Estimates and Confidence Interval Tests

When the marginal utility of money is constant[12] and the compensating variation of a change from the status quo to an alternative state of the world is desired, the expected WTP is measured by $E[WTP] = (-1/\gamma)\cdot(V^1 - V^0)$, where $\gamma$ is the cost parameter, $V^0$ is the conditional indirect utility evaluated at the original (status quo) levels and $V^1$ is the conditional indirect utility under the alternative (improved) state of the world. Given $V^0$ and $V^1$ are functions of random parameters, WTP is calculated over the distribution of parameters in parallel fashion

[12] Though clearly a simplification, constant marginal utility is typically assumed for discrete choice behavioral models to ease welfare calculation. Relaxing this assumption is beyond the scope of the current article, but see Herriges and Kling (1999) for a treatment that involves non-linear income effects.

**Table 3** Mean willingness to pay estimates and 95% confidence intervals for welfare scenarios

| Hawaiian monk seal status | Small-tooth sawfish status | Column I Two-species model (PSC threatened*) | Column II Three-species model (PSC threatened) | Column III Three-species model (PSC recovered) |
|---|---|---|---|---|
| E | T | \$50.45 (\$33.77, \$67.42) | \$33.96 (\$25.01, \$42.92) | \$81.00 (\$68.83, \$94.85) |
| E | R | \$64.50 (\$41.09, \$85.06) | \$53.39 (\$43.27, \$63.95) | \$100.20 (\$86.55, \$114.31) |
| T | E | \$46.49 (\$31.51, \$63.48) | \$43.72 (\$34.60, \$52.91) | \$90.60 (\$77.89, \$104.59) |
| T | T | \$96.89 (\$72.41, \$125.49) | \$77.83 (\$64.79, \$91.19) | \$124.37 (\$108.70, \$142.40) |
| T | R | \$136.20 (\$106.96, \$168.16) | \$102.35 (\$85.53, \$119.06) | \$143.44 (\$126.47, \$163.41) |
| R | E | \$85.66 (\$68.25, \$102.96) | \$68.12 (\$55.29, \$81.00) | \$114.50 (\$99.08, \$130.79) |
| R | T | \$110.76 (\$84.59, \$138.00) | \$96.95 (\$83.47, \$111.03) | \$148.77 (\$130.02, \$168.48) |
| R | R | \$151.90 (\$117.84, \$183.72) | \$121.68 (\$103.59, \$138.24) | \$168.87 (\$148.67, \$190.56) |

Confidence intervals in parentheses
*E* Endangered, *T* Threatened, *R* Recovered

to the simulation-based estimation procedure. Table 3 contains welfare estimates calculated from the three-species and two-species models, as well as the 95% Krinsky-Robb confidence intervals.

Columns I, II, and III in Table 3 each present eight annual household WTP estimates associated with improvements from the status quo ESA levels for the HMS and SS (both originally endangered). Column I presents the welfare estimates using the two-species model results, which implicitly assumes the PSC is at its status quo level of threatened since it is excluded from the survey altogether. Columns II and III display the WTP estimates for the three-species model under differing explicit assumptions made about the PSC. In Column II, WTP is calculated assuming the PSC remains threatened in 50 years, while the estimates in Column III assume the PSC is recovered by then. For the set of scenarios considered, which represent improvement from the status quo ESA levels for one or more species, mean annual household WTP estimates range from \$46.49 to \$151.90 using the two-species model estimates, from \$33.96 to \$121.68 for the three-species model that assumes the PSC is threatened, and from \$81.00 to \$168.87 for the three-species model that assumes the PSC is recovered. All welfare estimates are statistically different from zero.[13]

To formally compare the WTP estimates from the two-species and three-species surveys, we cannot rely on a direct comparison of 95% confidence intervals for the mean WTP estimates from the two surveys. Poe et al. (1994) have shown that non-overlapping confidence intervals are biased indicators of the significance of differences in estimated means. To formally test for WTP differences between survey versions, we utilize the approach suggested by Poe et al. (2005) to develop precise confidence bounds for the difference between the mean WTP estimates for the two baseline surveys. We use a complete combinatorial convolution approach that involves empirically estimating the confidence interval around the difference of the mean WTP values. This is accomplished by computing the empirical distribution of the difference from calculating every possible difference between the WTP values in each iteration of the Krinsky-Robb simulation used to empirically simulate the welfare measures.

[13] Note that a CV-based estimate of the value of protecting the Hawaiian monk seal by Samples and Hollyer (1990) in 2006 dollars is more than three times the magnitude (\$165.80) of a similar estimate of the public's WTP for recovering the species.

This is a computationally-intensive, but precise, method for estimating the difference between two (independent) WTP distributions.

## 8 External Tests of Validity and Scope

Method of convolutions-based confidence intervals are calculated to make comparisons between welfare measures from the two-species and three-species models. Prior to conducting the three types of external scope tests described above, we compare welfare measures contained in Columns I and II in Table 3, which can be viewed as an external validity, or consistency, test. The welfare estimates in these two columns, $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y)$ and $\mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = T)$, are measures of the same thing. Therefore, this comparison is made to ensure that the two sets of measures from the different surveys are leading to statistically indistinguishable welfare estimates. For all eight pairs of mean welfare estimates, the 95% confidence interval for the difference in means for each pair of welfare estimates contains zero, suggesting no statistical difference at the 5% level of significance between estimates from the two-species model and the three-species model for any of the welfare estimates calculated (see Table 4). Confirmation that both models lead to statistically identical welfare estimates for the same improvement in species status provides the foundation for making comparisons to identify scope effects when comparing welfare estimates for a clearly superior improvement to a lesser improvement. If the two models led to different estimates of the WTP for the same improvement, any scope effects (or lack thereof) could be questioned on the grounds of confounding effects caused by systematic differences in welfare estimates generated from the two datasets.

The remaining comparisons are the three types of external scope tests discussed above. The Type 1 scope test involves comparisons between WTP values in Columns I and III of Table 3 (comparison across each row between Columns I and III). As shown in Table 5, five of the eight tests reject, at the 5% level (and six of eight tests reject at the 10% level), the null hypothesis that the welfare estimates that improve only SS and/or HMS are the same as welfare estimates that improve SS and/or HMS, plus recovers PSC. This manifests as

**Table 4** Method of convolutions confidence intervals of differences in means for identical welfare measures, $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS}{=}y) - \mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS}{=}y, \mathrm{PSC} = T)$

| Two-species version and Three-species version ESA levels | | | 95% CI of difference in mean WTP | |
|---|---|---|---|---|
| HMS ($x$) | SS ($y$) | PSC ($z$) | Lower bound | Upper bound |
| R | R | T | −$14.64 | $22.47 |
| T | E | T | −$3.74 | $39.23 |
| R | E | T | −$2.07 | $34.85 |
| E | T | T | −$14.61 | $34.41 |
| E | R | T | −$8.98 | $49.55 |
| T | T | T | −$15.94 | $44.27 |
| R | T | T | −$0.23 | $69.37 |
| T | R | T | −$7.67 | $67.65 |

*E* Endangered, *T* Threatened, *R* Recovered

**Table 5** Method of convolutions confidence intervals of differences in means for welfare measures that differ only in the three-species version improving PSC to recovered, $WTP_2(HMS = x, SS = y) - WTP_3(HMS = x, SS = y, PSC = R)$

| Two-species version | | Three-species version | | 95% Confidence interval | | 90% Confidence interval | |
|---|---|---|---|---|---|---|---|
| HMS ($x$) | SS ($y$) | HMS ($x$) | SS ($y$) | Lower bound | Upper bound | Lower bound | Upper bound |
| *R* | *R* | *R* | *R* | *−$56.92* | *$22.23* | *−$50.18* | *$15.49* |
| T | E | T | E | −$64.46 | −$22.70 | −$61.10 | −$26.72 |
| R | E | R | E | −$52.44 | −$5.31 | −$48.55 | −$9.25 |
| E | T | E | T | −$51.72 | −$10.29 | −$47.99 | −$13.75 |
| E | R | E | R | −$63.08 | −$10.68 | −$58.13 | −$14.61 |
| T | T | T | T | −$57.67 | $4.17 | −$52.58 | −$1.22 |
| R | T | R | T | −$70.61 | −$5.05 | −$65.16 | −$10.51 |
| *T* | *R* | *T* | *R* | *−$42.82* | *$28.94* | *−$36.70* | *$21.91* |

*E* Endangered, *T* Threatened, *R* Recovered; *Italics* indicate comparisons that cannot reject scope insensitivity at the 5 and 10% levels

confidence intervals for the difference in means ($WTP^{2-\text{species}} - WTP^{3-\text{species}}$) being strictly in the negative range.

Three of the eight comparisons, however, had confidence intervals for the differences in means that contained zero, thus implying the null hypothesis could not be rejected at the 5% level of significance. At the 10% level, there are only two rejections of the null hypothesis. These WTP comparisons that fail to reject the null hypothesis of equality of welfare measures involve improvements to *both* the SS and HMS from their status quo levels. Four WTP comparisons that do reject the null hypothesis, and thus do not display scope insensitivity, involve scenarios where either the SS or HMS remains endangered. This suggests that there may be scope insensitivity to the number of species being improved (from two species to three species). We further investigate this in the next section.

The Type 2 external scope effect tests are between the WTP associated with an improvement from the status quo for the SS or/and HMS in the two-species version and the WTP for a clearly superior (in terms of improvement) scenario for the SS or/and HMS in the three-species version while holding the PSC at a threatened level. Out of the 19 Type 2 comparisons, 7 indicate zero is contained in the 95% (and 90%) method of convolutions-based confidence intervals, suggesting the null hypothesis of scope insensitivity cannot be rejected in these cases (Table 6). These cases have in common that the improvements to one or both of the HMS and SS are from threatened in the two-species version to recovered in the three-species version, suggesting that there may be some scope insensitivity on the upper end of ESA status-level improvements (from threatened to recovered). The lone exception is for the comparison between the WTP from the two-species version associated with improving the HMS from endangered to recovered, while the SS remains endangered and PSC is threatened, and the WTP from the three-species version associated with the same improvements, but the SS is improved to threatened. Together with the scope insensitivity shown for changes to SS from threatened to recovered while holding the changes to the HMS and PSC the same, this would seem to suggest an insensitivity to scope with respect to improvements to the SS when the HMS is recovered. However, this is countered by the fact that a comparison of WTP under the same conditions for HMS and PSC, but for SS from endangered in the two-species version to recovered in the three-species version, does not exhibit scope insensitivity.

**Table 6** Comparing improved scenarios to inferior scenarios with PSC threatened in both versions: method of convolutions confidence intervals of differences in means, $\text{WTP}_2(\text{HMS} = x, \text{SS} = y) - \text{WTP}_3(\text{HMS} = x, \text{SS} = y, \text{PSC} = T)$

| Two-species version | | Three-species version | | 95% Confidence interval | | 90% Confidence interval | |
|---|---|---|---|---|---|---|---|
| HMS ($x$) | SS ($y$) | HMS ($x$) | SS ($y$) | Lower bound | Upper bound | Lower bound | Upper bound |
| E | T | T | T | −$48.54 | −$6.55 | −$44.81 | −$10.10 |
| E | T | T | R | −$75.19 | −$29.38 | −$71.05 | −$33.26 |
| E | T | R | T | −$67.84 | −$25.43 | −$64.16 | −$29.12 |
| *E* | *T* | *E* | *R* | *−$22.25* | *$16.09* | *−$ 18.86* | *$12.62* |
| E | T | R | R | −$95.21 | −$47.73 | −$90.91 | −$51.84 |
| T | E | T | T | −$51.40 | −$9.43 | −$48.18 | −$13.53 |
| T | E | T | R | −$78.24 | −$32.45 | −$74.50 | −$36.71 |
| T | E | R | E | −$41.20 | −$0.19 | −$38.09 | −$4.21 |
| T | E | R | T | −$70.80 | −$28.36 | −$67.53 | −$32.51 |
| T | E | R | R | −$98.28 | −$50.86 | −$94.38 | −$55.29 |
| *T* | *T* | *T* | *R* | *−$35.14* | *$26.09* | *−$30.21* | *$20.69* |
| *T* | *T* | *R* | *T* | *−$28.15* | *$30.51* | *−$23.62* | *$25.27* |
| *T* | *T* | *R* | *R* | *−$55.06* | *$7.29* | *−$50.01* | *$1.88* |
| *R* | *E* | *R* | *T* | *−$33.21* | *$11.04* | *−$29.69* | *$7.21* |
| R | E | R | R | −$60.50 | −$11.43 | −$56.36 | −$15.65 |
| E | R | T | R | −$66.58 | −$11.53 | −$61.44 | −$15.67 |
| E | R | R | R | −$86.56 | −$30.07 | −$81.26 | −$34.37 |
| *R* | *T* | *R* | *R* | *−$42.70* | *$21.19* | *−$37.31* | *$15.84* |
| *T* | *R* | *R* | *R* | *−$20.15* | *$50.61* | *−$14.27* | *$43.50* |

*E* Endangered, *T* Threatened, *R* Recovered; *Italics* indicate comparisons that cannot reject scope insensitivity at the 5 and 10% levels

Unlike the Type 1 and Type 2 comparisons, Type 3 comparisons measure differences in welfare between WTP estimates that differ in both the number of species improved and the levels of improvement. As in the Type 1 comparisons, the PSC is assumed to be recovered in the WTP estimates from the three-species model, while the ESA status levels achieved by the HMS and SS are collectively better in the three-species version compared to the two-species version. For this set of 19 comparisons, each exhibits scope effects with 90% confidence intervals for the difference in means strictly in the negative range (see Table 7). Therefore, all Type 3 comparisons pass the external scope test and reject the scope insensitivity null hypothesis.

## 9 A Test for Proportional Scope Effects

The scope tests described above test the null hypothesis of scope insensitivity against the alternative hypothesis of scope sensitivity, but does not provide information about the extent or shape of the scope sensitivity, if found. A different, and stronger, scope test than the three types of external tests conducted above can be done to evaluate the presence of proportional scope effects with respect to the number of species improved; that is, whether total WTP for a scenario is proportional to the number of species whose status is being improved.

**Table 7** Comparing improved scenarios with PSC recovered to inferior scenarios with PSC threatened: method of convolutions confidence intervals of differences in means, $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y) - \mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = \mathrm{R})$

| Two-species version | | Three-species version | | 95% Confidence interval | | 90% Confidence interval | |
|---|---|---|---|---|---|---|---|
| HMS ($x$) | SS ($y$) | HMS ($x$) | SS ($y$) | Lower bound | Upper bound | Lower bound | Upper bound |
| E | T | T | T | −$97.85 | −$51.53 | −$93.60 | −$55.15 |
| E | T | T | R | −$118.39 | −$69.64 | −$113.86 | −$73.43 |
| E | T | R | T | −$123.48 | −$73.77 | −$119.10 | −$77.84 |
| E | T | E | R | −$71.27 | −$28.72 | −$67.48 | −$32.37 |
| E | T | R | R | −$145.61 | −$92.49 | −$140.83 | −$96.87 |
| T | E | T | T | −$100.98 | −$54.54 | −$97.08 | −$58.64 |
| T | E | T | R | −$121.67 | −$72.71 | −$117.38 | −$76.95 |
| T | E | R | E | −$90.21 | −$44.79 | −$86.56 | −$48.95 |
| T | E | R | T | −$126.72 | −$76.88 | −$122.62 | −$81.37 |
| T | E | R | R | −$148.93 | −$95.68 | −$144.44 | −$100.39 |
| T | T | T | R | −$77.98 | −$14.21 | −$72.60 | −$19.67 |
| T | T | R | T | −$83.11 | −$18.90 | −$77.94 | −$24.55 |
| T | T | R | R | −$104.96 | −$38.02 | −$99.43 | −$43.69 |
| R | E | R | T | −$88.77 | −$37.39 | −$84.47 | −$41.76 |
| R | E | R | R | −$110.86 | −$56.25 | −$106.20 | −$60.76 |
| E | R | T | R | −$109.29 | −$51.70 | −$103.86 | −$55.91 |
| E | R | R | R | −$136.30 | −$74.98 | −$130.62 | −$79.62 |
| R | T | R | R | −$92.37 | −$24.03 | −$86.56 | −$29.65 |
| T | R | R | R | −$69.68 | $5.20 | −$63.38 | −$2.03 |

*E* Endangered, *T* Threatened, *R* Recovered

Similar tests are often applied in health risk contingent valuation studies to evaluate scope sensitivity of WTP to changes in risk probabilities (e.g., Hammitt and Graham 1999; Hammitt 2000).

The test involves a deeper assessment of the Type 1 comparisons since these comparisons hold improvements to SS and HMS constant between the two-species and three-species survey versions, with the difference being WTP estimates from the three-species version assume a recovered PSC. Thus, the only difference between the mean WTP estimates is the additional improvement in the three-species version of the PSC, such that $\mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y)$ and $\mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = \mathrm{R})$. For this test, we calculate the distribution of the difference between the *average per-species WTP values* for each survey version, $\mathrm{AvgWTP}_2 = \mathrm{WTP}_2(\mathrm{HMS} = x, \mathrm{SS} = y)/s_2$ and $\mathrm{AvgWTP}_3 = \mathrm{WTP}_3(\mathrm{HMS} = x, \mathrm{SS} = y, \mathrm{PSC} = \mathrm{R})/s_3$, where $s_2$ is the number of species with improved ESA status in the two-species version (takes on values of 1 or 2), and $s_3$ is the number of species with improved ESA status in the three-species version (takes on values of 2 or 3). Using the method of convolutions method, we calculated the distribution of the difference, $\mathrm{AvgWTP}_2 - \mathrm{AvgWTP}_3$ to test the null hypothesis that $\mathrm{AvgWTP}_2 = \mathrm{AvgWTP}_3$, which is equivalent to a null hypothesis of WTP being proportional to the number of species being improved. For this strong external scope test, acceptance of the null hypothesis means that WTP appears to be proportional to the number of species (in going from 1 or 2 species improved to 2 or 3 species improved, respectively).

Table 8 presents the results of this method of convolutions-based test. Recall that in Type 1 comparisons (Table 5), 2 of the 9 comparisons could not reject the null hypothesis of scope insensitivity, indicating no statistical difference in WTP between two-species and three-species versions. These same two cases, corresponding to scenarios where SS and HMS are both recovered and HMS is threatened while SS is recovered, reject the null hypothesis of proportional WTP (equality of AvgWTP values). One other case, where HMS is recovered, while SS remains endangered, rejects the null hypothesis at the 5% level in this scope test, suggesting that WTP is not proportional to the number of species. However, the remaining 5 comparisons indicate that the 95% (and 90%) confidence bounds include zero, and thus the null hypothesis of proportional WTP cannot be rejected. These results, together with the Type 1 tests, suggest that for these five comparisons scope effects exist and, furthermore, are linearly related to the number of species being improved.

Thus, the test of proportional WTP confirms that the two cases that could not reject scope insensitivity in the "weak" scope test (Type 1 comparison) also reject proportional WTP, which is not at all surprising. What is somewhat surprising, however, is that the proportionality test results suggest that several comparisons (five) not only exhibit scope effects, but proportional scope effects. The surprising part of this is that the test is looking at an average WTP per species improved without consideration for whether the improvement is from endangered to recovered, threatened to recovered, or endangered to recovered. This seems curious given results in Tables 2 and 3 show the parameter and WTP estimates and their corresponding significance levels that individuals appear to have distinct preferences for improvements to SS and HMS to a threatened status from endangered, as well as to recovered from endangered, and for an improvement to recovered from threatened for PSC. Note that for the two comparisons that may be viewed as "apples to apples" comparisons, one assuming an improvement in one status level across all species (i.e., SS and HMS improve to threatened in both versions, and PSC improves to recovered in the three-species version) and the other assuming all species are recovered, proportional WTP cannot be rejected.

## 10 Discussion

A total of 46 external scope test comparisons that check for the presence of scope insensitivity were conducted to evaluate the sensitivity of public preferences to the number of species protected and the amount of protection afforded each species.[14] In a subset of the scope test comparisons where scope insensitivity was refuted, a stronger test of scope effects, one testing for proportionality of WTP with respect to the number of species, was conducted and found scope effects were proportional to the number of species in most cases. Though this follow-up scope test was applicable for only a small subset of cases, the results suggest there may be a proportional relationship along this dimension, but clearly further research is needed beyond this single application.

Although the majority of the external scope test comparisons rejected scope insensitivity (thus indicating sensitivity to scope), including those involving simultaneous changes to the number of species valued and amount of improvements to each species, scope insensitivity was found in several comparisons of welfare estimates that involve one type of change in

[14] All of the scope tests conducted here are based on the distributions of differences in welfare measures, and are thus sensitive to the precision with which the welfare measures are measured.

**Table 8** Confidence bounds for difference in average WTP per species from two-species version (AvgWTP2) and three-species version (AvgWTP3) with identical improvements to SS and HMS, but with PSC recovered in three-species version

| Two-species version | | Three-species version | | | | 95% CI | | 90% CI | |
|---|---|---|---|---|---|---|---|---|---|
| HMS ($x$) | SS ($y$) | HMS ($x$) | SS ($y$) | AvgWTP2 | AvgWTP3 | Lower bound | Upper bound | Lower bound | Upper bound |
| *R* | *R* | *R* | *R* | *$ 75.95* | *$56.29* | *$1.42* | *$37.63* | *$4.47* | *$34.48* |
| T | E | T | E | $46.49 | $45.30 | −$15.04 | $19.83 | −$12.47 | $15.91 |
| *R* | *E* | *R* | *E* | *$85.66* | *$57.25* | *$9.49* | *$47.68* | *$12.53* | *$44.20* |
| E | T | E | T | $50.45 | $40.50 | −$7.84 | $27.51 | −$4.53 | $23.90 |
| E | R | E | R | $64.50 | $50.10 | −$10.23 | $36.36 | −$5.56 | $32.91 |
| T | T | T | T | $48.45 | $41.46 | −$6.50 | $21.74 | −$4.28 | $19.15 |
| R | T | R | T | $55.38 | $49.59 | −$8.82 | $20.76 | −$6.40 | $18.29 |
| *T* | *R* | *T* | *R* | *$ 68.10* | *$47.81* | *$4.17* | *$37.16* | *$6.84* | *$33.68* |

*E* Endangered, *T* Threatened, *R* Recovered; *Italics* indicate comparisons that *reject* proportional scope effects at the 5 and 10% levels scope, either in the number of species or the amount of protection for each species, but not when both dimensions change. Moreover, within these results scope insensitivity seems to occur for specific types of comparisons. Specifically, it appears that scope insensitivity may be found when comparing WTP for different amounts of improvement to the same species in the upper range of status improvements (i.e., improvements to recovered versus to threatened), holding everything else the same. That is, along the ESA improvement dimension, preferences appear to show scope insensitivity. This result was seen in 6 of the 9 cases where scope insensitivity could not be rejected and seems to be consistent with the diminishing marginal value argument discussed by Rollins and Lyke (1998). However, the argument for diminishing marginal utility does not fully explain the other three cases of scope insensitivity. Rather, the two Type 1 comparisons not displaying scope sensitivity appear to occur for the largest status improvements, but at the same time PSC improves in the three-species version, possibly suggesting there is some interaction effects in this range of improvements that is driving the result. The remaining case (the 14th Type 2 comparison) showing scope insensitivity compares two welfare measures that both lead to a recovered HMS but differ in whether or not SS remains endangered (two-species version) or improves to threatened (three-species version), suggesting that preferences may not be sensitive to scope for a small improvement to SS when HMS is recovered, which may also be related to interaction effects between species improvements or be an artifact of the precision with which welfare measures are calculated.[15]

Of course, a number of other possible explanations have been set forth in the literature for the failure of scope tests to find sensitivity to scope that may be argued for here. Czajkowski and Hanley (2009), for example, suggest that scope insensitivity may be rooted in omitted variables in the choice experiments playing a more significant role than the included attributes. They explored the use of explicit (descriptive) labeling of choice alternatives in the choice experiment (which results in alternative specific constants in the utility specification) to control for non-attribute effects that may confound scope effects associated with attributes. In the Czajkowski and Hanley (2009) work, labeling of the choice alternatives appears to provide additional information that respondents may have filled in for themselves in its absence.

This is representative of a larger issue discussed by Fischhoff et al. (1993) and others related to whether respondents are actually answering the same choice question the researchers intend. Although considerable effort was spent pretesting the survey instruments in focus groups, cognitive interviews, and peer review, it is possible that unintended response effects may be present in the data. For example, even though the choice question options are cast in certain terms such that it is implied the results will occur in 50 years with 100% certainty, an alternative explanation for insensitivity to larger improvements (e.g., improving a species' status from endangered to recovered) is that respondents are placing a small probability of success on large improvements (Fischhoff and Furby 1988).

Another potential explanation for declining marginal utility for improvements to species is that respondents view large improvements in species' status as leading to additional costs to society beyond those that are paid out-of-pocket by their household. For example, an improvement from an endangered status to recovered for the SS over the 50 year time period may be viewed by respondents as requiring extreme measures, such as stopping all coastal development and beach recreation in South Florida, which would severely impact businesses, tourism, and individuals living there. This suggests respondents "fill in the blanks" left by

[15] As one reviewer pointed out, these individual scope tests are independent. Formal scope tests that jointly test numerous hypotheses of scope insensitivity are beyond the scope of the current paper, but are a useful future direction.

the fact that only program *results* are being described, not the specific means to achieve those results.[16] Alternatively, one can make an argument that people's preferences are insensitive to the protection of the PSC in addition to the other two species because it is already at a threatened ESA status, and thus is doing better relative to the other species at present. Consequently, it is possible that individuals may feel it less important than the other species to protect when confronted with such a choice.

Since we use choice experiments, some issues that arise with scope tests using sequential and nested contingent valuation questions (Carson et al. 1998; Bateman et al. 2004) are avoided. This can simultaneously be viewed as a benefit and a difficulty with the approach. Respondents do respond to a sequence of choice questions in each survey, but the attribute levels for each option are determined from a statistical-based experimental design and are not selected to control for path dependency of the goods being valued. Respondents are asked to compare a "new set" of options in each question, which is intended to encourage respondents to base their response to each question independently of how much they said they were willing to pay in previous questions. Although ordering effects may still exist in SPCE surveys, assessing this is beyond the scope of the present analysis and is left for future research.

In a recent paper, Heberlein et al. (2005) argued that conventional scope tests, ones focusing on comparing differences in the distributions of measures of central tendencies of willingness to pay, may not be appropriate as a means of evaluating scope insensitivity since they ignore important individual-specific characteristics that may drive behavior. Their study involved conducting individual-level scope tests and investigating the role of each individual's attitudes toward the good being valued and behavior that may lead to scope test failures. The analysis in this paper follows the conventional approach (sample-level, not individual-level, analysis), but future analysis along the lines of Heberlein et al. (2005) may provide further insights into the patterns of WTP estimates presented here and may be worth pursuing in the context of non-CV stated preference applications generally.

In summary, this paper has presented the results from a series of scope tests to evaluate scope insensitivity along two dimensions of the preferences for threatened and endangered species using data from SP choice experiments. The scope tests evaluate the sensitivity of preferences across two samples of US residents over a broad range of improvements to two or three threatened and endangered marine species. Not surprisingly, over the tests performed, not all indicate scope sensitivity as diminishing marginal values appear to play an important role along at least one dimension of scope changes, the improvements to each species. When both dimensions simultaneously change in the valuation, WTP appears to display scope sensitivity, suggesting the changes are sufficiently large along that area of the utility surface to discern measurable differences. However, further research is needed to evaluate scope effects resulting from valuing different numbers of species, as the results here were evaluated over a small range of species (2 or 3) and had two cases displaying scope insensitivity; though, for the cases shown to be sensitive to scope, proportionality of WTP to the number of species was usually evidenced. Despite this, these results overall appear to support well-behaved preferences for these threatened and endangered species in this case, though further research is needed to assess whether this holds for other species and for larger numbers of species.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

[16] Indeed, in explaining their choices to the SPCE questions some focus group participants did discuss the likely mechanisms that would need to be put in place to bring about the changes being valued.

## References


Adamowicz W, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. Am J Agric Econ 80:64–75

Adamowicz W, Louviere J, Williams M (1994) Combining revealed and stated preference methods for valuing environmental amenities. Am J Agric Econ 26:271–292

Alpizar F, Carlsson F, Martinsson P (2003) Using choice experiments for non-market valuation. Econ Issues 8(1):83–110

Arrow K, Solow R, Portney P, Leamer E, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. Fed Regist 58(10):4601–4614

Bateman I, Brouwer R (2006) Consistency and construction in stated WTP for health risk reductions: a novel scope-sensitivity test. Res Energy Econ 28:199–214

Bateman I, Carson R, Day B, Hanemann W, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E, Pearce D, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar, Northampton

Bateman I, Cole M, Cooper P, Georgiou S, Hadley D, Poe G (2004) On visible choice sets and scope sensitivity. J Environ Econ Manage 47:71–93

Bateman I, Cooper P, Georgiou S, Navrud S, Poe G, Ready R, Riera P, Ryan M, Vossler C (2005) Economic valuation of policies for managing acidity in remote mountain lakes: examining validity through scope sensitivity testing. Aquat Sci 67:274–291

Berrens R, Bohara A, Silva C, Brookshire D, McKee M (2000) Contingent values for new Mexico instream flows: with tests of scope, group-size reminder and temporal reliability. J Environ Manage 58:73–90

Berrens R, Ganderton P, Silva C (1996) Valuing the protection of minimum instream flows in new Mexico. J Agric Res Econ 21(2):294–309

Carson R (1997) Contingent valuation surveys and tests of insensitivity to scope. In: Kopp R, Pommerhene W, Schwartz N (eds) Determining the value of non-marketed goods: economic, psychological, and policy relevant aspects of contingent valuation methods. Kluwer, Boston

Carson R, Groves T (2007) Incentive and information properties of preference questions. Environ Res Econ 37:181–210

Carson R, Mitchell R (1993) The issue of scope in contingent valuation studies. Am J Agric Econ 75(5): 1263–1267

Carson R, Mitchell R (1995) Sequencing and nesting in contingent valuation surveys. J Environ Econ Manage 28:155–173

Carson R, Flores N, Hanemann W (1998) Sequencing and valuing public goods. J Environ Econ Manage 36:314–323

Cummings R, Taylor L (1999) Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. Am Econ Rev 89(3):649–665

Czajkowski M, Hanley N (2009) Using labels to investigate scope effects in stated preference methods. Env Res Econ, Online First

Desvousges W, Johnson F, Dunford R, Boyle K, Hudson S, Wilson K (1993) Measuring natural resource damages with contingent valuation: tests of validity and reliability. In: Hausman J (ed) Contingent valuation: a critical assessment. North Holland Publishers, Amsterdam

Diamond P, Hausman J, Leonard G, Denning M (1993) Does contingent valuation measure preferences? experimental evidence. In: Hausman J (ed) Contingent valuation: a critical assessment. North Holland Publishers, Amsterdam

Dillman D (2000) Mail and internet surveys: the tailored design method. John Wiley, New York

Fischhoff B, Furby L (1988) Measuring values: a conceptual framework for interpreting transactions with special reference to contingent valuation of visibility. J Risk Uncertain 1(2):147–184

Fischhoff B, Quadrel M, Kamlet M, Loewenstein G, Dawes R, Fischbeck P, Klepper S, Leland J, Stroh P (1993) Embedding effects: stimulus representation and response mode. J Risk Uncertain 6:211–234

Giraud K, Loomis J, Johnson R (1999) Internal and external scope in willingness-to-pay estimates for threatened and endangered wildlife. J Environ Manage 56:221–229

Hammitt J (2000) Evaluating contingent valuation of environmental health risks: the proportionality test. Assoc Environ Res Econ Newslett 20(1):14–19

Hammitt J, Graham J (1999) Willingness to pay for health protection: inadequate sensitivity to scope? J Risk Uncertain 8:33–62

Hanley N, Wright R, Adamowicz W (1998) Using choice experiments to value the environment: design issues, current experience, and future prospects. Environ Res Econ 11(3–4):413–428

Heberlein T, Wilson M, Bishop R, Schaeffer N (2005) Rethinking the scope test as a criterion for validity in contingent valuation. J Environ Econ Manage 50:1–22

Herriges J, Kling C (1999) Nonlinear income effects in random utility models. Rev Econ Stat 81(1):62–72
Huber J, Zwerina K (1996) The importance of utility balance in efficient choice designs. J Mark Res 33(3): 307–317
Kahneman D (1986) Comments. In: Cummings R, Brookshire D, Schuze W (eds) Valuing environmental goods: an assessment of the contingent valuation method. Rowman and Allanheld, Totowa
Kahneman D, Knetsch J (1992) Valuing public goods: the purchase of moral satisfaction. J Environ Econ Manage 22:57–70
Layton D (2000) Random coefficient models for stated preference surveys. J Environ Econ Manage 40:21–36
Layton D, Brown G, Plummer M (2001) Valuing multiple programs to improve fish populations. Unpublished report to the Washington state department of ecology
Layton D, Levine R (2005) Bayesian approaches to modeling stated preference data. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics. Springer, Dordrecht
Leiter A, Pruckner G (2009) Proportionality of willingness to pay to small changes in risk: the impact of attitudinal factors in scope tests. Env Res Econ 42:169–186
Lew D, Layton D, Rowe R (2010) Valuing enhancements to endangered species protection under alternative baseline futures: the case of the steller sea lion. Mar Res Econ 25:133–154
Loomis J, Ekstrand E (1997) Economic benefits of critical habitat for the mexican spotted owl: a scope test using a multiple-bounded contingent valuation survey. J Agric Res Econ 22(2):356–366
Loomis J, Larson D (1994) Total economic values of increasing gray whale populations: results from a contingent valuation survey of visitors and households. Mar Res Econ 9:275–286
Louviere J, Hensher D, Swait J (2000) Stated choice methods—analysis and application. Cambridge University Press, Cambridge
Morey E, Rowe R, Watson M (1993) A repeated nested-logit model of Atlantic Salmon fishing. Am J Agric Econ 75:578–592
Olar M, Adamowicz W, Boxall P, West G (2007) Estimation of the economic benefits of Marine mammal recovery in the St. Lawrence Estuary. Report to the policy and economics branch, fisheries and oceans Canada, regional branch Quebec
Poe G, Giraud K, Loomis J (2005) Computational methods for measuring the difference of empirical distributions. Am J Agric Econ 87(2):353–365
Poe G, Severance-Lossin E, Welsh M (1994) Measuring the difference (X-Y) of simulated distributions: a convolutions approach. Am J Agric Econ 76(4):904–915
Reaves D, Kramer R, Holmes T (1999) Does question format matter? Valuing an endangered species. Environ Res Econ 14:365–383
Richardson L, Loomis J (2009) The total economic value of threatened, endangered and rare species: an updated meta-analysis. Ecol Econ 68:1535–1548
Rollins K, Lyke A (1998) The case for diminishing marginal existence values. J Environ Econ Manage 36: 324–344
Rudd M (2009) National values for regional aquatic species at risk in canada. Endanger Species Res 6:239–249
Samples K, Hollyer J (1990) Contingent valuation of wildlife resources in the presence of substitutes and complements. In: Johnson R, Johnson G (eds) Economic valuation of natural resources: issues, theory, and applications. West View Press, Boulder pp 177–193
Schkade D, Payne J (1994) How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. J Environ Econ Manage 26:88–109
Siikamaki J, Layton D (2007) Discrete choice survey experiments: a comparison using flexible methods. J Environ Econ Manage 53:122–139
Smith V (1992) Arbitrary values, good causes, and premature verdicts. J Environ Econ Manage 22:71–89
Smith V, Osborne L (1996) Do contingent valuation estimates pass a 'Scope' test? A meta-analysis. J Environ Econ Manage 31:287–301
Train K (1998) Recreation demand models with taste differences over people. Land Econ 74(2):230–239
Train K (2003) Discrete Choice methods with simulation. Cambridge University Press, Cambridge
Veisten K, Hoen H, Navrud S, Strand J (2004) Scope insensitivity in contingent valuation of complex environmental amenities. J Environ Manage 73:314–331