Aggregating Social Benefits of Endangered Species Protection: The Case of the Cook Inlet Beluga Whale

Daniel K. Lew[♣♦]

Accepted version: June 2023

ABSTRACT

The Cook Inlet beluga whale (CIBW) is an endangered whale found in waters off the state of Alaska that is protected under the U.S. Endangered Species Act. The federal recovery plan estimates recovery costs will total $73 million (in 2013 dollars). In this study, we use data from a stated preference discrete choice experiment (CE) study to estimate the aggregate benefits of recovering the CIBW and generally for improving its conservation status and reducing extinction risk. Estimated CE models account and test for utility scale heterogeneity, attribute non-attendance, self-selection bias, and demographic effects. Aggregation methods that differ in adjustments made to the mean household welfare and/or to the number of population units are compared by assessing their impact on the resulting aggregate (population-level) welfare estimates. Alaska households are willing to pay between $34 million (95% CI of [$25 million, $44 million]) and $103 million ([$74 million, $131 million]) for full recovery of the CIBW, depending upon the model and aggregation assumptions. While some of the state-level recovery value estimates are below the total cost of combined federal and state recovery actions, accounting for welfare benefits beyond the state of Alaska justifies recovery actions by the benefit-cost criterion.

Key words: Attribute non-attendance, beluga whale, discrete choice experiment, endangered species, benefit-cost analysis, welfare aggregation

JEL Classification Codes: Q51, Q57

This article and its findings are those of the author and do not necessarily reflect the views of the National Marine Fisheries Service or the U.S. Department of Commerce. The author thanks Kristy Wallmo and an anonymous reviewer for useful comments that improved the paper. All remaining errors are the author's alone.

[♣]National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, WA 98115 USA
[♦]Corresponding author. E-mail: Dan.Lew@noaa.gov, Phone: (530) 554-1842, Mailing address: NOAA, National Marine Fisheries Service, Alaska Fisheries Science Center, Bldg 4., 7600 Sand Point Way NE, Seattle, WA 98115 USA

Introduction

The Cook Inlet beluga whale (CIBW) is a whale species found only in Cook Inlet, Alaska, USA. Its habitat is adjacent to Alaska's largest city, Anchorage, which has over a quarter million people and is the state's largest port as well as the center of the state's commerce. The CIBW was listed as an endangered species under the Endangered Species Act (ESA) in 2008 due to persistent population declines since the 1990s. Recent assessments suggest there are only about 331 animals remaining, down from about 1,300 in 1979 (Goetz et al. 2023; Muto et al. 2021). In 2016, the National Marine Fisheries Service (NMFS), the U.S. federal agency charged with managing and conserving marine species, adopted a recovery plan for the species (NMFS 2016).

The recovery plan identifies the anthropogenic and natural sources of current declines, outlines criteria for downlisting (improving from an endangered to a threatened status) and recovery (improving sufficiently to be delisted as a threatened or endangered species), and describes a set of recovery actions to take. The main sources of decline of ongoing concern include harmful noise from human activities, potential catastrophic events like oil spills and mass strandings, pathogens and disease, and habitat loss and degradation (NMFS 2016). Importantly, the recovery plan delineates a 50-year comprehensive federal recovery strategy of adaptive research, monitoring, management, and public education and outreach that is estimated to cumulatively cost $73 million (in 2013 dollars). The primary policy question that we seek to answer is whether the public benefits of recovering the CIBW outweigh the costs of recovery.

To answer this, we analyze stated preference (SP) discrete choice experiment (CE) data from a survey of Alaska households that collected information on attitudes and preferences for protecting the CIBW. We generate and compare aggregate Alaska-level welfare estimates for

CIBW recovery from alternative estimation models and by applying several aggregation approaches. Numerous empirical SP models exist in the literature to control for potential biases or response anomalies in CE data with the intent of generating more valid and reliable estimates of economic values (Johnston et al. 2017). Some modeling issues often accounted for in these empirical CE models include allowing for various forms of preference heterogeneity (Morey and Rossman 2003; Birol et al. 2006; Train 2009), controlling for utility scale heterogeneity (Lundhede et al. 2009; Fiebig et al. 2010; Olsen et al. 2011; Hess and Rose 2012) and sample selection bias (Cameron and DeShazo 2013; Bosworth, Cameron, and DeShazo 2015; Lewis et al. 2019), and accounting for choice heuristics like attribute non-attendance (ANA) (Hole 2011; Lew and Whitehead 2020). The sensitivity of individual or household-level welfare estimates to these modeling decisions are generally the focus of inquiries in the literature.

However, how these modeling decisions interact with or are offset by assumptions made in the aggregation process and the resulting effects on aggregate welfare are less explored. In particular, a key decision in aggregation is determining the "market extent"--i.e., the population whose values are non-zero (Smith 1993; Loomis 2000). Often, SP studies aggregating welfare do not assume the entire population should be included in the market extent. Instead, they adjust downward the population to which the sample-based mean or median welfare estimate is applied, reflecting an implicit assumption that some portion of the population has a zero willingness-to-pay (WTP) (Loomis 1987; Morrison 2000; Brouwer 2008). One common, and conservative, approach is to scale the population downward to reflect the proportion of the overall contacted sample that responded to the survey; that is, to assume that the survey response rate mirrors the proportion of the population that has a non-zero WTP. Morrison (2000) compared how several aggregation assumptions affected aggregate welfare estimates using a simple multinomial logit

CE model that allows for demographic interactions. Unsurprisingly, he found that large differences can arise.

While these aggregation assumptions about the relevant market extent have predictable first-order effects on aggregate welfare, it is unclear to what degree some modeling choices may contribute or offset these effects. For example, Lew and Whitehead (2020) identify several empirical approaches used in ANA studies for welfare measurement that implicitly affect the proportion of the sample assumed to contribute to the calculation of welfare. The extent to which accounting for ANA behavior impacts aggregate welfare has not been investigated.

Johnston et al. (2017) recommend that "calculation of welfare measures for policy guidance should recognize potential effects of sample selection, preference heterogeneity, and the extent of the market" (p. 368). To achieve this here, we compare aggregate welfare estimates of public values for CIBW recovery resulting from four CE estimation models embodying different assumptions about preference heterogeneity, utility scale differences, ANA, sample selection bias, and the role of demographics, and two market extent assumptions. In addition to the expected result that aggregation assumptions can lead to starkly different aggregate values, we show that modeling assumptions can have qualitatively large effects on aggregate welfare of a similar order of magnitude as different assumptions about the market extent do. In terms of the policy question about whether public benefits for species recovery outweigh the costs of recovery, we find that for the Alaska state population alone the benefits do not fully offset the costs in all cases. However, given the CIBW is a public resource of the United States, extending the public benefits beyond state borders, even under extremely conservative assumptions, enables the CIBW recovery program to pass the benefit-cost test.

In contrast to previous studies valuing CIBW recovery (Lew 2018, 2019), the emphasis here is on generating a population-level estimate of public WTP for protection of the species. Thus, it adds to the expanding knowledge of public values for threatened and endangered species protection (Martin-Lopez, Montes, and Benayas 2008; Richardson and Loomis 2009; Lew 2015). In so doing, the study also contributes to ongoing discussions of the benefit-cost ratios associated with recovery of endangered species in the United States. For example, Moore et al. (2022) review the public benefits for species protection and compare them to the costs of recovery, finding that even under very restrictive assumptions the benefit-cost ratios for 34 of 36 threatened and endangered species are greater than one.

The Survey and Data

Data were collected in a survey of the public's preferences and values for protecting the CIBW.[1] The survey was sponsored by NMFS and respondents were informed of this in all mailings and in the survey itself. Focus group input suggested that this enhanced the credibility of the survey and contributed to the belief that survey responses would be consequential.

The mail survey included information and questions about the CIBW, other threatened and endangered U.S. marine mammals, past and current population trends and management actions, and potential future actions to protect and recover it, as well as questions about the individual (e.g., socio-demographic questions). After being presented with a budget reminder and cheap talk script to reduce the potential for hypothetical bias (Cummings and Taylor 1999), respondents were presented with four CE questions. Each asks the respondent to choose their

[1] Details not provided here about the survey design, pretesting activities, and survey implementation are available in two other studies (Lew 2018, 2019), which each used a different subset of the data and explored other issues.

most preferred and least preferred option from among three alternatives (Alternatives A, B, and C) (see figure A1 in appendix).[2] The choice alternatives differ in the results of protection actions on the CIBW and in their costs. Alternative A in each CE question maintains the status quo, which results in the species remaining endangered with a 25% risk of extinction by 2112, and would not result in any added costs to the household. Alternatives B and C do more to protect the CIBW and lead to improvements in extinction risk and ESA listing status but result in an added cost to the household.

The payment vehicle used in the CE questions is a combination of added costs to the household from increases in costs of goods and services and federal taxes.[3] Three different payment schedules are presented in the study, differentiated by the number of years the household would pay the added costs (1, 5, and 10 years). For the 1-year payment survey version, the added cost for each alternative in the CE questions is presented as a lump sum payment. Respondents would not pay anything more beyond the single one-time payment. In the 5- and 10-year payment survey versions, the added cost is presented as an annual payment for the household that would occur for the stated number of years.

---

[2] Note that the rank choice (best-worst) format was employed in numerous CE studies in economics (e.g., Scarpa et al. 2011; Lancsar et al. 2013) around the time of this survey due in part to advantages the format has for estimating efficient model parameters. However, along with other SP question formats, it has since been used less after concerns were raised about incentive compatibility properties of SP question formats (Carson and Groves 2007; Vossler et al. 2012). In particular, Vossler et al. (2012) showed that a CE design exhibiting a particular set of properties ensure it is incentive compatible. Among those properties are that the choice design uses binary referenda between one alternative and the status quo, has independence of choice sets, has at most one alternative from the universe presented being implemented, and is viewed as consequential by respondents. The rank choice format used here does not have all of these features. Thus, it is one that cannot be assured to be fully incentive compatible, despite having features that support its consequentiality. As a result, there is the potential for strategic bias in the responses.

[3] Focus group and cognitive interview testing suggested people generally assumed the federal tax was certain and binding and would make up most, if not all, of the "added cost" to their household, which reduces but does not eliminate concerns about this payment vehicle not being fully "fixed and unmalleable" (Johnston et al. 2017).

For each payment schedule treatment, there are 16 survey versions that differ in the attribute levels (population statuses, risk of extinction, and costs) seen by respondents. Attribute levels for the conservation status under Alternatives B and C are--in decreasing order of endangerment--"endangered," "threatened," or "recovered".[4] There are ten extinction risk levels that could be achieved under Alternatives B and C ranging from "less than 1%" to 22%, which is less than the 25% extinction risk associated with Alternative A, the status quo program. The experimental design also included non-zero added cost amounts for Alternatives B and C ranging from \$10 to \$350. The final experimental design was selected based on a D-efficiency criterion.[5] The resulting 48 survey versions (16 versions × 3 payment treatments) were distributed randomly and in equal numbers in the mail-out sample.

The survey was fielded in 2013 to 4,200 randomly drawn Alaska households.[6] The survey was implemented using a mixed mode approach (Dillman et al. 2014) with five contacts: an advance letter, the initial mailing with a \$5 monetary incentive, a postcard reminder, a follow-up telephone call, and a second full mailing. In total, 1,747 completed surveys were returned. The overall response rate, excluding undeliverables, was 44.4%.[7] The data used for the analysis

---

[4] "Recovery" is defined as reducing the extinction risk to a near-zero level and removing the species from the ESA list of threatened and endangered species.

[5] A two-step process was used to select the experimental design used in the study. First, a program developed in GAUSS selected the 20 most D-efficient experimental designs based on a main effects utility specification from among 100,000 randomly drawn candidate experimental design combinations that excluded fully dominating choices. Then a series of Monte Carlo simulations were done to evaluate how well each of the 20 designs performed when analyzed with a main-effects rank-ordered multinomial logit model using pseudo-data constructed from several sets of assumed true utility parameters. The designs were evaluated in terms of their ability to estimate the true utility parameters, and the best performing one was selected for use.

[6] Rural households were oversampled (50% of the mail-out sample), so the estimation models presented here are sample-weighted. The sample was drawn from the Marketing Systems Group's address-based database of U.S. households, which has nearly 100% coverage of households in Alaska since it is based on the U.S. Postal Services' Computerized Delivery Sequence File. The survey implementation was carried out by Ipsos.

[7] Response rates varied from 33% to 55% over the 48 survey versions. The response rate excluding undeliverables also excludes those who were deceased or moved out of Alaska (273 total undeliverables).

consist of 1,316 respondents.[8] This excludes individuals who did not answer any of the choice questions or indicated the same choice for "best" and "worst" option (134) and respondents who strongly rejected the payment vehicle identified through their responses to a series of Likert questions that followed the CE questions (297).

Table 1 presents summary statistics of several characteristics of the estimation sample. The sample is statistically similar to the State of Alaska's population in terms of gender, proportion of full-time workers and Alaska Natives/American Indians, and household income. However, relative to the population, the sample is older, more educated, and consists of a higher proportion of white respondents and homeowners.

Analyzing Discrete Choice Experiment Responses

A variety of mixed logit models that account for preference heterogeneity and the ordered and panel nature of the data are estimated. Extensions of a base panel rank-ordered model are developed to evaluate and control for several response anomalies and to allow for aggregation adjustments. In particular, we estimate models that allow for differences in utility scale due to varying levels of confidence or certainty in responses (Lundhede et al. 2009), for choice heuristic behavior as reflected in attribute non-attendance behavior (Lew and Whitehead 2020), and for the effects of sample selectivity (Bosworth, Cameron, and DeShazo 2015; Lewis et al. 2019) and demographics on choice behavior (Morey and Rossman 2003; Birol et al. 2006). Given the different payment schedule treatments (one, five, and ten years), an endogenous discounting model (Bond et al. 2009; Lew 2018; Vasquez-Lavin et al. 2021) is employed that allows for estimation of an implied discount rate.

[8] 719 from urban areas and 597 from rural areas.

The random utility maximization (RUM) models assume each of the J = 3 choice alternatives (indexed by j) in each CE question has a utility U_j that consists of a deterministic (V_j) and stochastic (ε_j) component. V_j is linear in the present value of the annual cost (PVC_j) and non-cost attributes ($\boldsymbol{x}_j$) of choice alternative j. Thus, for a given individual, the utility of the jth choice alternative is

$$U_j = V_j + \varepsilon_j = \boldsymbol{\beta}\Box\boldsymbol{x}_j - \gamma\Box\text{PVC}_j + \varepsilon_j \quad , \tag{1}$$

where $\boldsymbol{\beta}$ and γ are parameters to be estimated, with γ > 0 to conform with theory and ensure finite moments of the WTP (Carson and Czajkowski 2019). Assuming equal-sized annual costs of size C_j over T time periods, the PVC_j can be represented by the product of a cumulative discount factor (ψ) and C_j; i.e., $\text{PVC}_j = C_j\cdot\psi$. Under exponential discounting, ψ is defined as a function of the discount rate (r):

$$\begin{aligned}\psi \quad &= \sum_{t=0}^{T-1}\left(\frac{1}{1+r}\right)^t \\ &= \left(1+\frac{1}{r}\right)\cdot\left[1-\frac{1}{(1+r)^T}\right] \quad .\end{aligned} \tag{2}$$

Thus, equation 1 becomes

$$U_j = \boldsymbol{\beta}\Box\boldsymbol{x}_j - \gamma\Box C_j\Box\psi + \varepsilon_j \quad . \tag{3}$$

Assuming ε_j follows a Gumbel distribution, the probability an individual chooses the jth alternative as best takes the familiar multinomial logit form:

$$\text{P}_j = \Pr[j \text{ is best}] = \frac{exp(\mu\cdot V_j)}{\sum_{k=1}^{J} exp(\mu\cdot V_k)} \quad , \tag{4}$$

where μ is the scale factor that is inversely related to the variance, which is $\pi^2/6\mu^2$ for the Gumbel distribution (Swait and Louviere 1993).

The scale factor can be allowed to vary over individuals either stochastically (Fiebig et al. 2010; Hess and Rose 2012) or systematically (Lundhede et al. 2009; Davis, Burton, and Kragt

2019). In this study, we allow the scale factor to differ across individual responses to a stated certainty question that directly follows the CE questions and assesses how confident respondents are in their CE responses (Lundhede et al. 2009).[9] Responses to the certainty question are measured on a 5-point scale, from 1 = not at all confident to 5 = extremely confident. The sample mean was 3.5. In the model, we specify the scale parameter as a function of two dummy variables, one indicating a lower confidence response between 1 and 3 (CONF13) and the other indicating a higher confidence response of 4 (CONF4) on the 5-point scale; i.e., $\mu = \exp(\upsilon_{13}\cdot \text{CONF13} + \upsilon_{4}\cdot \text{CONF4})$, where υ_{13} and υ_{4} are parameters to be estimated. Since this specification implies the baseline confidence level is extremely confident, which one would expect to have the lowest utility variance, our expectation is that υ_{13} and υ_{4} are negatively signed, which implies a smaller scale parameter and thus larger variance.

Respondents are asked to choose both the best (most preferred) alternative and the worst (least preferred) alternative from among the J (=3) alternatives. Thus, the responses provide a full rank-ordered preference ordering. Denote the three choice alternatives as j, k, and l (where j, k, and l □ {A, B, C}). If j was selected as the most preferred alternative and l was selected as worst, the implied preference order of j as best, k as second best, and l as worst (denoted $j > k > l$) can be represented as a joint probability equal to the product of the unconditional probability of j being best and the conditional probability of k being second best if j is best:

$$\Pr[j > k > l] = \pi_{jkl} = \mathrm{P}_j \times \mathrm{P}_{k|j} \,, \tag{5}$$

[9] The specific question wording is provided in Appendix Figure A2.

where $P_{k|j}$ = Pr[k is second best|j is best] = $\frac{exp(\mu \cdot V_k)}{\sum_{m \neq j} exp(\mu \cdot V_m)}$, j,m,k,l = {A, B, C}. This is the rank-ordered conditional logit model (Chapman and Staelin 1982).

The joint probability of observing the sequence of choices an individual makes is the product of individual choice probabilities under the assumption that the Gumbel errors are independent across the repeated choices. Letting r_q equal the rank ordering in the qth choice question, the joint probability of observing the Q = 4 rank orderings is represented as

$$\Pr[r_1, r_2, r_3, r_4] = \pi_{r_1} \cdot \pi_{r_2} \cdot \pi_{r_3} \cdot \pi_{r_4}, \tag{6}$$

where π_{r_q} (for q = 1, 2, 3, 4) is defined by Equation 5.

To account for unobserved preference heterogeneity, the random parameters logit, or mixed logit (MXL), model is used. The parameters $\boldsymbol{\beta}$ in Equation 3 (i.e., the non-cost parameters) are allowed to be individual-specific and assumed to follow a multivariate normal distribution (with mean vector $\bar{\beta}$ and variance-covariance matrix Σ). In addition, the cost parameter γ is assumed to be randomly distributed over a symmetric positively-constrained triangular distribution (Hensher and Greene 2003), which is defined by a single hyper-parameter, υ (=ln(γ)).[10]

In the MXL model formulation, the individual's joint probability of observing the rank orderings {r_1, r_2, r_3, r_4} over the Q = 4 choice questions is evaluated over the assumed random

[10] Allowing the cost parameter (γ in equation 3) to be normally distributed led to model results that suggest a significant percentage of parameter values in the distribution are in the negative orthant, which is counter to theory. We imposed consistency with demand theory by restricting the cost parameter to be positive, which meant estimating models that specify the cost parameter to be fixed, lognormally distributed, or triangularly distributed. Since the lognormally-distributed cost parameter models failed to converge, we restrict ourselves to the triangularly distributed cost parameter models (and fixed cost parameter model results are available upon request).

parameter joint distribution $f(\Omega)$, where Ω (=$\{\bar{\beta}, \Sigma, \upsilon\}$) are the hyper-parameters defining the marginal distributions of the random parameters:

$$\Pr[r_1, r_2, r_3, r_4] = \int \prod_{q=1}^{Q} \pi_{r_q}(\beta, \gamma) \cdot f(\Omega) d\Omega, \tag{7}$$

for all j, k, and l □ {A, B, C}.

To control for attribute non-attendance, an inferred ANA approach is employed (Lew and Whitehead 2020). Specifically, we adopt the endogenous attribute attendance model of Hole (2011) to allow for ANA behavior. Let A_s be the set of attributes (out of K total attributes) to which an individual pays attention in the CE questions. The probability of observing an attribute κ (κ □ A_s) being paid attention to is specified as a logit probability:

$$\Pr[\kappa] = \exp(\theta_\kappa)/[1+\exp(\theta_\kappa)], \tag{8}$$

where θ_κ is an attribute-specific constant. The probability of observing A_s being paid attention to is the product of the binomial logit probabilities in equation 8 across all attributes:

$$\Pr[A_s] = H_{A_s} = \prod_{\kappa \in A_s} \left(\frac{exp(\theta\ )}{1+exp(\theta\ )} \right) \times \prod_{\kappa \notin A_s} \left(\frac{1}{1+exp(\theta\ )} \right). \tag{9}$$

This assumes independence of the individual attribute attendance probabilities. Note that there are $S = 2^K$ possible combinations of attributes being attended to, implying 2^K possible patterns of attribute attendance. The ANA version of the MXL model can be cast in a latent class specification:

$$\Pr[r_1, r_2, r_3, r_4] = \int \sum_1^{2^K} \{H_{A_s} \prod_{q=1}^{Q} \pi_{r_q}(\beta^{A_s})\} f(\Omega) d\Omega, \tag{10}$$

where β^{As} is the vector of all parameters (β and γ) with zeros imposed for attributes that are not included in A_s, so that the attributes do not contribute to utility (they are zeroed out since they are ignored by the respondent).

These panel rank-ordered MXL models are estimated using maximum simulated likelihood. The likelihood function for the MXL model is approximated through simulation by taking R = 2,000 Sobol draws from a multivariate normal distribution and symmetric positive-constrained triangular distribution and evaluating the conditional joint choice probabilities at each draw. The multivariate normal distribution allows for correlation between the random parameters, $\boldsymbol{\beta}$. The symmetric triangular distributed cost parameter is assumed to be independent of the other random parameters. The triangular distribution is described by a single parameter, υ, representing the spread (and mean and mode equal to $\upsilon/2$) with a lower bound of 0 (see appendix for details). The simulated joint probability is the mean over the R draws.[11]

Model Specifications

The four estimation models presented in this analysis are variants of the panel rank-ordered MXL model described above. They differ in whether or not they account for sample selection bias, ANA, utility scale differences due to differences in response certainty, and demographic interaction effects. All of the models assume utility of the non-status quo alternatives (Alternatives B and C) is a function of the logged reduced risk of extinction (ln(RR+1)), the ESA listing status as captured through two dummy variables --one for a threatened status (THR) and one for a recovered status (REC)--and the present value of cost for the alternative (PVC). The logged extinction risk allows for utility to increase with extinction risk reductions at a decreasing rate.[12] The separate ESA listing status dummies allow for marginal utility for status improvements beyond reductions in extinction risk. The status quo alternative is represented in the model as an alternative specific constant (SQ).

[11] The simulated maximum likelihood procedure was programmed in GAUSS version 22. Halton and Sobol quasi-random draws of varying sizes (500 to 2,000) were evaluated and model parameter stability supported use of 2,000 Sobol draws.
[12] Other model specifications that are quadratic in RR had similar model fit and statistical significance but allowed for negative marginal utility for large extinction risk reductions. Here we impose that the utility associated with increases in RR are non-negative.

In the first model, the MXL-Base model, there are no considerations for sample selection bias, attribute non-attendance, utility scale differences, or demographic interaction effects. The second model, the MXL-Adj model, controls for two potential issues of concern: ANA and utility scale differences. Specifically, the MXL-Adj model applies the endogenous attribute attendance model to account for ANA behavior and controls for utility scale differences due to differences in respondent's response certainty, as described in the previous section. Attribute non-attendance behavior is assumed to be possible in relation to the reduced extinction risk (ln(RR+1)) and ESA conservation status (THR and REC) attributes.[13] The remaining two models build on the MXL-Adj model by controlling for sample selection bias (MXL-Sbias) or by controlling for potential demographic effects on utility (MXL-Demo) in addition to accounting for ANA behavior and utility scale differences.

The MXL-Sbias model follows the approach employed by Cameron and DeShazo (2013), Bosworth, Cameron, and DeShazo (2015), and Lewis et al. (2019) to control for sample selection bias. The approach involves using individual-level predictions of the deviation from the mean survey response propensity to control for deviations in the utility parameters attributable to systematic sample selection effects. This is implemented in three steps: First, we model an individual's propensity to respond to the survey as a function of observable characteristics. To facilitate this, auxiliary data about all households in the mail-out sample were procured by the survey vendor, Ipsos. The auxiliary data included variables primarily related to the type and location of a household's mailing address and demographics from Census data. Specifically, the response propensity logit model used as covariates dummy variables for the presence of a post office box address, whether the household lived in a high-rise building, if the household lived in

[13]Models that also included cost attribute non-attendance were estimated but not used since likelihood ratio tests rejected the cost attribute being ignored.

an area adjacent to the Cook Inlet, if it was a single-person household, and whether the household was located in a low (< $30K/year) or high (>$150K/year) income area. Second, the estimated response propensity model (appendix table A1) is then used to generate individual-level predicted response propensities. Third, a response propensity deviation (RPDEV) variable, which is the individual-level predicted response propensity subtracted from the mean response propensity, is interacted with the SQ and REC variables in the CE model (see Table 1).[14]

The fourth and final model we present is the MXL-Demo model, which allows for the interaction of demographic variables with the ln(RR+1), THR, and REC variables. Although specifications with other demographic variables were tried, only LOWEDUC (a dummy variable for whether individual has less than a high school diploma or equivalent) was consistently statistically significant and was therefore the one used here. This model allows for welfare estimates to be adjusted to reflect differences in demographics between the sample and population.

Welfare and Aggregation Assumptions

The four models described above were selected to be consistent with aggregation approaches that embody alternative assumptions about (1) the similarity in preferences between sample respondents and non-respondents; (2) differences between the sample and the population (POP); and (3) the proportion of the total population of interest (PP) to which the sample-based mean WTP is applied (see Figure 1). When respondents and non-respondents have different underlying preferences and auxiliary data is available to assess differences between them, sample selection models can be used to deal with non-random non-response. The sample-based utility

[14] RPDEV interactions with other utility variables were consistently not statistically significant.

models can incorporate features that allow for adjusting the sample mean WTP to account for the underlying sample selection bias, as well as potential response biases or behaviors (ANA and utility scale effects) as described above for MXL-Sbias. In this case, the sample mean WTP calculated from a sample selection bias-corrected model must be applied to the full population, POP (such that PP=1.0). If instead respondents and non-respondents are assumed to have the same preferences (i.e., an absence of selection bias), then the sample mean WTP is assumed representative of some or all of the population of interest. Thus, sample mean WTP estimates derived from MXL-Base and MXL-Adj can be applied to the portion of the population believed to be in the market extent. Utility scale differences and attribute non-attendance can be controlled for (MXL-Adj) to account for potential response biases prior to aggregation. Additionally, if preferences are believed to systematically differ with demographics that differ between the sample and population, then sample mean WTP can be adjusted by using population-level demographics in a model that accounts for these effects (MXL-Demo). When this demographic-adjustment approach is undertaken, the assumption is that the sample mean can be applied to the entire population (PP = 1).

For the initial aggregation, we define POP to be all households in the state of Alaska, USA, for the year corresponding to when the survey was conducted. The 2013 American Community Survey 5-year estimate of the number of households in Alaska is 251,899. The most recent estimate is not much larger (253,346)[15], suggesting that the growth in households in Alaska over the last several years is fairly level (less than a 0.6% change). When assuming only a proportion of the full population should be in the market extent in welfare aggregation, we define PP as the percentage of the mail-out sample (minus undeliverables) that was used to

[15] See https://www.census.gov/quickfacts/AK.

estimate the models; PP= 1316/2927 = 33.5%. This assumption is more conservative than some studies that use the unadjusted survey response rate (e.g., Morrison 2000). It assumes that the same proportion of respondents who did not respond to the choice questions (item respondents) and those who rejected the payment vehicle, and thus were not used in model estimation, exists in the population and should not have the estimated welfare applied to them.

For the MXL-Base model described above, the expected household willingness-to-pay, E[WTP], resulting from a change is

$$E[WTP] = \int (1/\gamma)\cdot(V^1 - V^0)\, f(\Omega) d\,\Omega, \tag{11}$$

where γ 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 improved state of the world. Given V^0 and V^1 are functions of random parameters, E[WTP] is calculated over the distribution of parameters in parallel fashion to the simulation-based estimation procedure (with 2,000 Sobol parameter vector draws). Note that since the utility is a function of the present value of costs, E[WTP] represents the present value of a household's willingness to pay.

For all other models, ANA must be accounted for in the welfare calculation. To this end, we propose an expected weighted average WTP:

$$E[WTP] = \int \Sigma_s\, [\Pr[A_s] \times (1/\gamma)\cdot(V^1(\beta^{As}) - V^0(\beta^{As}))]\, f(\Omega) d\Omega, \tag{12}$$

where $\Pr[A_s]$ is the probability of observing the sth ($s = 1,\ldots,2^K$) combination of attributes being paid attention to as defined by Equation 9, and $V_0(\beta^{As})$ and $V_1(\beta^{As})$ are the corresponding conditional indirect utility specifications with parameter vectors that reflect the associated attribute non-attendance pattern. This welfare measure explicitly takes into consideration the probability of each specific pattern of attribute non-attendance behavior (including full attendance). For the MXL-Sbias model, the effect of the response propensity deviation is zeroed out in calculating welfare. For the MXL-Demo model, E[WTP] is calculated assuming population-level demographics from Census data (LOWEDUC = 0.3845). 95% confidence intervals are constructed using the simulation approach of Krinsky and Robb (1986) using 1,000 iterations.

Estimation Results

Table 2 presents the estimated means of the distributions of the non-cost random parameters, along with other estimated model parameters and model fit statistics. However, it excludes the Choleski matrix elements associated with the multivariate normally-distributed random parameters (reported in Appendix Table A-2). The signs and statistical significance of the estimated mean parameters for all four models conform to our expectations. Logged reduction in extinction risk (ln(RR+1)), improving the ESA status from endangered to threatened (THR), and improving the status to recovered (REC) are all positively related to utility, all else equal, and are statistically significant at conventional levels. In the MXL-Demo model, demographic interactions with the low educational attainment dummy variable (LOWEDUC) were statistically significant in two cases: when interacted with extinction risk reductions (ln(RR+1)) and with the THR dummy variable. In both cases, the interaction effect was negative and statistically different from zero (at least at the 10% level of significance) and reduced the net positive utility effect from risk reductions and achieving a threatened ESA status, suggesting that all else equal individuals with lower education have a lower utility for CIBW improvements than others. In the MXL-Sbias model, the parameters on the RPDEV interaction with SQ and REC were both large, negative, and statistically significant. This suggests a sample selection bias in the marginal utility of recovering the CIBW, with recovery having on average a stronger positive effect on utility among respondents than the broader population. It also suggests a lesser propensity among respondents to choose the status quo alternative relative to non-respondents.

In all models, the symmetric triangularly-distributed cost parameter, which is constrained to conform to theoretical expectations, is statistically significant, and the implied discount rate individuals apply to future payments is large and statistically significant. The implied exponential discount rates across the four models were generally large, ranging from a low of 428% ($\exp(1.46) \cong 4.28$) for the MXL-Sbias model to a high of 648% ($\exp(1.87) \cong 6.48$) for the MXL-Demo model. These implied discount rate estimates suggest individuals heavily discount future payments. The mean SQ parameter in all models is negative and statistically significant, indicating a preference by respondents for choosing alternatives that enhance protection for the CIBW rather than the status quo. The likelihood ratio index (LRI), a pseudo-R^2 value, is at least 0.372 for the four models indicating the overall models are statistically significant.

Appendix Table A-2 presents the estimated Choleski matrix parameters for the variance-covariance matrix of the normally-distributed random parameters. For all models, the diagonal Choleski matrix parameters are all statistically significant and larger in magnitude relative to the corresponding mean parameters, thus indicating the presence of preference heterogeneity. However, the off-diagonal Choleski matrix parameters are all positive, with most statistically different from zero. As Mariel and Artabe (2020) have shown, one can only interpret negative off-diagonal parameters because of scale issues. Thus, the results suggest the presence of nonsystematic preference heterogeneity but is inconclusive about whether there is, and the direction of, correlation across random parameters.

Controlling for ANA and utility scale effects from response uncertainty improves model fit, as suggested by the higher LRI (0.375) and lower Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) for the MXL-Adj model relative to the MXL-Base model (as well as direct likelihood ratio [LR] tests that were done). Conforming to our prior expectations, the CONF13 and CONF4 parameters in the MXL-Adj, MXL-Sbias, and MXL-Demo models were negative and statistically different from zero, indicating larger utility variances for respondents who are less confident in their responses to the CE questions. Attribute non-attendance was present in these models, but only in relation to extinction risk reduction. For extinction risk, the probability of attendance was 54.3% (95% CI of [47.6%, 60,3%]) in the MXL-Adj model, 56.7% ([50.0%, 63.5%]) in the MXL-Sbias model, and 83.3% ([75.6%, 89.3%]) in the MXL-Demo model. The statistical insignificance of the estimated ANA parameters related to ESA status in these models (95% CIs of [~0.1%, 100%]) indicates the inability of these models to identify ANA behavior in relation to ESA status improvements.[16] For each of the MXL-Adj, MXL-Sbias, and MXL-Demo, LR tests reject the null hypotheses of no ANA ($p < 0.1$) and equal utility variance across certainty responses ($p < 0.001$), both separately and jointly. LR tests also reject the separate null hypotheses of no sample selection bias ($p < 0.001$) and no demographic interactions ($p < 0.001$).

Welfare Results

[16] Fixed parameter conditional logit specifications of these models indicated statistical significance of ANA parameters for both ESA status and extinction risk reductions, but the lack of significance in the random parameters specifications highlights the potential for ANA to be confounded with preference heterogeneity.

Table 3 reports estimates of the household-level mean present value of WTP associated with full recovery (REC = 1, THR = 0, and RR = 25%) and the present value of marginal WTP associated with reducing extinction risk by one percentage point and improving the ESA status to threatened and to recovered. All WTP estimates are statistically different from zero except those for the marginal value of ESA status changes, regardless of the estimation model. The estimated mean household present value of WTP for full recovery ranged from a low of $221 (95% CI of [$117, $322]) for the MXL-Demo model to a high of $409 ([$295, $520]) for the MXL-Adj model. The full recovery value estimate for the MXL-Base model is very similar to the MXL-Demo model estimate ($228 with a 95% CI of [$112, $339]), while the MXL-Sbias model's estimate is more similar to the MXL-Demo model's ($395 [$280, $511]). A similar pattern emerges for the marginal WTP of a one percentage point extinction risk reduction, with household-level value estimates ranging from $49 [$29, $68] for the MXL-Demo model to $88 [$65, $113] for the MXL-Adj model. The estimated marginal value of a status change to a threatened level is statistically different from zero for the MXL-Base model ($35 [$14, $56]) and the MXL-Sbias model ($26 [$0.25, $50]), but not for the other two models at the 5% level. None of the marginal values for a recovered status change are statistically different from zero at the 5% level.

Using a confidence interval-based test procedure (see appendix), we assess whether there are statistical differences between WTP estimates from the four models. We find that full recovery household-level values for the MXL-Base and MXL-Demo are statistically the same, and those from the MXL-Adj and MXL-Sbias models are statistically the same (at the 5% level). However, we can reject the equality of full recovery values between the MXL-Base (or MXL-Demo) and MXL-Adj (or MXL-Sbias). A parallel set of outcomes occur in tests involving the marginal values of extinction risk reductions. In contrast, for the marginal values associated with ESA status changes, tests for a given marginal change (in THR or REC) suggest that there are no statistical differences between any estimates of the four models at the 5% level.

Table 4 presents the aggregate (population-level) WTP estimates for full recovery for each of the four models assuming the mean household-level WTP is applied to the full population (second column; PP=1.0) or to a proportion of the population (third column; PP = 0.335). The full population aggregate WTP estimates ranged from $55.7 million [$29.5M, $81.1M] for the MXL-Demo model to $103.0 million [$74.4M, $99.4M] for the MXL-Adj model. The aggregate WTP estimates applied to a proportion of the population is reported for the models that are consistent with such an approach (the MXL-Base and MXL-Adj models). For these models, the aggregate WTP is lower of course—$19.2 million [$9.5M, $28.6M] for the MXL-Base model and $34.5 million [$24.9M, $43.9M] for the MXL-Adj model.

Net Benefits of CIBW Recovery

We evaluate whether the comprehensive recovery efforts outlined in the 2016 NMFS CIBW recovery plan pass the benefit-cost test (net benefits > 0) by comparing the aggregate WTP estimates in Table 4 with the aggregate cost of recovery programs of $73 million. For these benefit-cost tests, we do not use the MXL-Base model aggregate WTP values since the other models perform better and are preferred. This leaves four aggregate WTP values to be evaluated. Two of the aggregate WTP values, for the MXL-Adj and MXL-Sbias models that assume PP=1, exceed the $73 million total recovery costs, suggesting the net social benefits of the programs are positive.[17] However, the other two aggregate WTP values, for the MXL-Adj

[17] The 95% and 90% CIs for these aggregate WTP value estimates, respectively, do not contain the aggregate recovery cost amount, suggesting the costs are statistically lower than benefits in these cases.

model assuming PP=0.335 and for the MXL-Demo model, are lower than the aggregate costs, although the total recovery cost amount is within the 95% CI of the MXL-Demo aggregate value. Thus, if one were to limit the evaluation to the state level, there is evidence in these two cases that the net benefits of CIBW recovery are not positive.

Recall, however, that the aggregate WTP values in Table 4 represent the total public benefits to Alaska households but that the CIBW is a species protected in the interest of the U.S. public. Thus, comparing the Table 4 aggregate benefits to the recovery costs is a very conservative benefit-cost test that limits the market extent to the population that is nearest the CIBW (i.e., those in the State of Alaska), and ignores values for CIBW recovery held by those living in other U.S. states.

Expanding WTP estimates beyond the sampled population can be viewed as a type of benefits transfer. The differing socio-demographics between the larger U.S. population and Alaska population may influence differences in WTP, not to mention attitudinal differences toward the CIBW and endangered species protection. Moreover, research into the spatial distribution of non-consumptive SP values like those in this study suggests the likelihood of spatially disperse but clustered areas of high and low values across the market extent (Johnston et al. 2015; Czajkowski et al. 2017), which underscores the dangers of applying a distance decay rule to any extrapolated values (Glenk et al. 2020). Nevertheless, past research on public values for endangered species protection derived from U.S. national samples (Wallmo and Lew 2012, 2016) indicates the market extent likely extends beyond the Alaska state border and should be accounted for when evaluating net benefits of state and federal recovery efforts.

Our objective is not to accurately estimate national level benefits of CIBW recovery, which would require a rigorous benefits transfer approach (Johnston et al. 2021). Instead, our aim is to assess whether the CIBW recovery efforts have net positive economic benefits. To conservatively account for U.S. public values for CIBW recovery, we employ a simple approach that extends estimated household-level WTP to the entire United States. To do this, we adopt a strong, yet conservative, assumption about what proportion of each non-Alaska U.S. state's households has a zero WTP for CIBW recovery (99.0%) and what proportion has a WTP for CIBW recovery equal to the mean household WTP value that we estimated for Alaska households (1.0%). Under the assumption that 1% of households in each state outside Alaska have a WTP equal to that estimated in a given model, Figure 2 displays the cumulative present value of aggregate WTP as one includes welfare benefits from states moving farther away from Alaska. Two cumulative WTP lines are shown: one using the MXL-Adj model's household WTP estimate assuming PP=0.335 within Alaska and the other using the MXL-Demo model's household WTP with PP=1.0. The other two aggregate WTP values already lead to net benefits being positive, and so are excluded from this illustration. Each of the cumulative WTP lines in Figure 2 begins at the corresponding aggregate WTP levels for Alaska from Table 4, then increases as WTP from nearby states (Washington state, then Oregon, and so on) are added.

In both cases, the cumulative WTP lines cross the $73 million total recovery cost level after the eight closest (by geographic centroid) states have been accounted for, then continue increasing as states farther away from Alaska are added. This simple approach adopts a very conservative assumption about WTP in non-Alaska U.S. states, yet clearly shows that accounting for public benefits outside of Alaska leads to all models passing the benefit-cost test.

Discussion and Concluding Remarks

Mean household values for CIBW recovery were found to be between $221 [$117, $322] and $409 [$295, $520] and can be compared to other value estimates for recovering other threatened and endangered whale species in the country. Wallmo and Lew (2016) estimated U.S. household WTP for the recovery of the then-endangered humpback whale and endangered Southern Resident killer whale. For these species, the estimated mean annual household WTP values were $65 (2013 dollars, 95% CI of [$61-$69]) and $90 ([$84, $95]), respectively. Additionally, Wallmo and Lew (2012) estimated similar mean annual U.S. household WTP values for recovery of both the North Atlantic ($77 [$72, $81]) and North Pacific right whales ($75 [$70, $79]), both of which are also endangered.

To compare these with our results, we convert the present value of WTP estimates for CIBW recovery to equivalent annual payments assuming a 7% discount rate, a rate commonly used in federal BCA analyses (OMB Circular A-94[18]), and a 10-year planning horizon to match the one implicit in the CE questions used in the other studies. The equivalent annual mean household WTP for CIBW recovery ranges from $30 [$16, $43] and $54 [$39, $69], which are statistically lower than those for both right whales and the Southern Resident killer whale. In the humpback whale case, we cannot reject the null hypothesis of the recovery value being the same.

Regardless, the finding of significantly lower recovery values for the CIBW than three other U.S. endangered whale species is noteworthy. This difference could reflect differences in the general preferences of the study population in this study (Alaska households) versus those of the broader U.S. population for endangered species recovery used in these other studies, but they may also reflect differences in CE modeling decisions used in these studies. At the same time, there is evidence that recovery values differ across species (Wallmo and Lew 2012), suggesting

[18] Available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/a94/a094.pdf

CIBW recovery values may just be lower. A more formal exploration of this is left for future research.

One of the objectives of this study was to explore whether CE modeling decisions impact aggregate WTP estimates at a level similar to how decisions about the market extent do. For this, we began with a panel rank-ordered mixed logit model as a base model, then estimated three additional models that all account for ANA behavior and utility scale differences reflecting respondent certainty differences. Two of those models also, separately, embody assumptions about whether there are utility differences between the sample and population (sample selection bias) and utility differences arising from demographics (systematic preference heterogeneity). Accounting for only ANA and utility scale differences (MXL-Adj model) led to almost a doubling of the household and aggregate WTP for full recovery relative to those corresponding to the baseline model (MXL-Base). This difference is similar in magnitude to the effect of assuming only a portion of the population has a non-zero WTP, as is assumed in many aggregation studies (Morrison 2000; Birol et al. 2006; Brouwer 2008). When additionally accounting for educational attainment differences between the sample and population in the demographic-interactions model (MXL-Demo), estimated WTP values were similar to those of the base model. The model that instead adjusts for sample selection bias in addition to ANA and utility scale differences (MXL-Sbias) yielded similar welfare estimates to the ANA and scale effects model (MXL-Adj). These findings suggest that modeling decisions in this case do have fairly large impacts on aggregate welfare.

A potential limitation of using the demographic-interactions model in aggregation of household WTP can occur when there is a mismatch, or at least imperfect information, on the population level for the corresponding demographic variable(s) used in the model. For example, the final MXL-Demo model interacted attributes with an education dummy that reflected the educational attainment for the respondent, assumed to be the household's decision-maker. As there was no Census-level information on the distribution of educational attainment for household decision-makers, we used educational attainment for adults (18 and older) in the population. This discrepancy points to a potential source for bias, but also highlights the need to further explore the relationship between individual and household WTP (Bateman and Munro 2009; Lindjhem and Navrud 2009) and the implications of using demographic-adjusted WTP to generate aggregate welfare. Moreover, it points more generally to the importance of modeling decisions on aggregation.

Another objective of this work was to evaluate whether public benefits outweigh the costs of recovery for the CIBW. For this, we examined the aggregate WTP for Alaska, the state in which the CIBW exists, and compared it to the estimated $73 million price tag for combined federal and state recovery efforts. At the state-level, we found that some of the estimated models and aggregation assumptions led to net benefits being positive. However, given the CIBW is a species protected by the federal government, a national-level market extent is appropriate. Under very conservative assumptions about recovery values in other U.S. states, we found that expanding the market extent to include additional U.S. states led to positive net benefits for CIBW recovery across all the models and aggregation assumptions. Since there are still uncertainties about the drivers of the CIBW decline (Muto et al. 2021), the full cost of recovery may change as more is learned. This study provides evidence of substantial social benefits of CIBW recovery that easily exceed the current estimate of recovery costs, which underscores evidence of the importance of endangered species protection to the U.S. public and is consistent with the finding of net positive benefits at the national level for dozens of other ESA-protected species by Moore et al. (2022). However, we lack the ability to generate a more accurate estimate of the national-level WTP for CIBW recovery due to our survey being limited to Alaska households and therefore cannot say how much larger the budget for recovery efforts can get before the public benefits are surpassed. Additionally, future work to understand the spatial distribution within Alaska and across the U.S. of both the preferences for CIBW recovery and who bears the costs of recovery actions (e.g., restrictions on activities in or near critical habitat) can aid in assessing distributional effects of species recovery.

References

Bateman, I. J., & Munro, A., 2009. Household versus individual valuation: What's the difference? *Environmental and Resource Economics*, *43*, 119-135.

Birol, E., Karousakis, K., & Koundouri, P., 2006. Using a choice experiment to account for preference heterogeneity in wetland attributes: The case of Cheimaditida wetland in Greece. *Ecological Economics*, *60*(1), 145-156.

Bond, C.A., K.G. Cullen, and D.M. Larson, 2009. Joint estimation of discount rates and willingness to pay for public goods. *Ecological Economics*, 68: 2751–2759.

Bosworth, R., Cameron, T. A., & DeShazo, J. R., 2015. Willingness to pay for public health policies to treat illnesses. *Journal of Health Economics*, *39*, 74-88.

Brouwer, R., 2008. The potential role of stated preference methods in the Water Framework Directive to assess disproportionate costs. *Journal of Environmental Planning and Management*, *51*(5), 597-614.

Cameron, T.A., and J.R. DeShazo, 2013. Demand for Health Risk Reductions. *Journal of Environmental Economics and Management*, *65*: 87-109.

Carson, R. T., & Czajkowski, M., 2019. A new baseline model for estimating willingness to pay from discrete choice models. *Journal of Environmental Economics and Management 95*: 57-61.

Carson, R.T., & Groves, T. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*(1), 181-210.

Chapman, R. G., & Staelin, R., 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, *19*(3), 288-301.

Cummings, R. G., & Taylor, L. O., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review 89*(3): 649-665.

Czajkowski, M., W. Budziński, D. Campbell, M. Giergiczny, and N. Hanley, 2017. Spatial heterogeneity of willingness to pay for forest management. *Environmental and Resource Economics* 68(3): 705-727.

Davis, K. J., Burton, M., & Kragt, M. E., 2019. Scale heterogeneity and its implications for discrete choice analysis. *Land Economics*, *95*(3), 353-368.

Dillman, D. A., Smyth, J. D., & Christian, L. M., 2014. *Internet, phone, mail, and mixed-mode surveys: The Tailored Design Method*. John Wiley & Sons.

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N., 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, *29*(3), 393-421.

Glenk, K., Johnston, R. J., Meyerhoff, J., & Sagebiel, J., 2020. Spatial dimensions of stated preference valuation in environmental and resource economics: methods, trends and challenges. *Environmental and Resource Economics*, *75*(2), 215-242.

Goetz, K. T., K. E. W. Shelden, C. L. Sims, J. M. Waite, and P. R. Wade. 2023. Abundance of belugas (Delphinapterus leucas) in Cook Inlet, Alaska, June 2021 and June 2022. *AFSC Processed Rep*. 2023-03, 47 p. Alaska Fish. Sci. Cent., NOAA, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle WA 98115. https://media.fisheries.noaa.gov/2023-06/PR2023-03.pdf

Hensher, D. A., & Greene, W. H., 2003. The mixed logit model: the state of practice. *Transportation*, *30*(2), 133-176.

Hess, S., & Rose, J. M., 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, *39*(6), 1225-1239.

Hole, A. R., 2011. A discrete choice model with endogenous attribute attendance. *Economics Letters*, *110*(3), 203-205.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., & Vossler, C. A., 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists 4*(2): 319-405.

Johnston, R. J., Boyle, K. J., Loureiro, M. L., Navrud, S., & Rolfe, J., 2021. Guidance to enhance the validity and credibility of environmental benefit transfers. *Environmental and Resource Economics*, *79*(3), 575-624.

Johnston, R. J., Jarvis, D., Wallmo, K., & Lew, D. K., 2015. Multiscale spatial pattern in nonuse willingness to pay: applications to threatened and endangered marine species. *Land Economics*, *91*(4), 739-761.

Krinsky, I., & Robb, A. L., 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, *68*(4), 715-719.

Lew, D. K., 2015. Willingness to pay for threatened and endangered marine species: a review of the literature and prospects for policy use. *Frontiers in Marine Science*, *2*, 96.

Lew, D. K., 2018. Discounting future payments in stated preference choice experiments. *Resource and Energy Economics*, *54*, 150-164.

Lew, D. K., 2019. Place of residence and cost attribute non-attendance in a stated preference choice experiment involving a marine endangered species. *Marine Resource Economics*, *34*(3), 225-245.

Lew, D. K., & Whitehead, J. C., 2020. Attribute non-attendance as an information processing strategy in stated preference choice experiments: Origins, current practices, and future directions. *Marine Resource Economics*, *35*(3), 285-317.

Lewis, D. J., Dundas, S. J., Kling, D. M., Lew, D. K., & Hacker, S. D., 2019. The non-market benefits of early and partial gains in managing threatened salmon. *PloS One*, *14*(8), e0220260.

Lindhjem, H., & Navrud, S., 2009. Asking for individual or household willingness to pay for environmental goods? Implication for aggregate welfare measures. *Environmental and Resource Economics*, *43*, 11-29.

Loomis, J.B., 1987. Expanding contingent valuation sample estimates to aggregate benefit estimates: Current practices and proposed solutions. *Land Economics* 63(4): 396-402.

Loomis, J.B., 2000. Vertically summing public goods demand curves: An empirical comparison of economic versus political jurisdictions. *Land Economics* 76(2): 312-321.

Lundhede, T. H., Olsen, S. B., Jacobsen, J. B., & Thorsen, B. J., 2009. Handling respondent uncertainty in choice experiments: Evaluating recoding approaches against explicit modelling of uncertainty. *Journal of Choice Modelling*, *2*(2), 118-147.

Mariel, P., & Artabe, A., 2020. Interpreting correlated random parameters in choice experiments. *Journal of Environmental Economics and Management*, *103*, 102363.

Martin-Lopez, B., Montes, C., & Benayas, J., 2008. Economic valuation of biodiversity conservation: the meaning of numbers. *Conservation Biology*, *22*(3), 624-635.

Moore, F. C., Stokes, A., Conte, M. N., & Dong, X., 2022. Noah's ark in a warming world: Climate change, biodiversity loss, and public adaptation costs in the United States. *Journal of the Association of Environmental and Resource Economists*, *9*(5), 981-1015.

Morey, E., & Greer Rossmann, K., 2003. Using stated-preference questions to investigate variations in willingness to pay for preserving marble monuments: Classic heterogeneity, random parameters, and mixture models. *Journal of Cultural Economics*, *27*(3), 215-229.

Morrison, M., 2000. Aggregation biases in stated preference studies. *Australian Economic Papers* 39(2): 215-230.

Muto, M. M., V. T. Helker, B. J. Delean, N. C. Young, J. C. Freed, R. P. Angliss, N. A. Friday, P. L. Boveng, J. M. Breiwick, B. M. Brost, M. F. Cameron, P. J. Clapham, J. L. Crance, S. P. Dahle, M. E. Dahlheim, B. S. Fadely, M. C. Ferguson, L. W. Fritz, K. T. Goetz, R. C. Hobbs, Y. V. Ivashchenko, A. S. Kennedy, J. M. London, S. A. Mizroch, R. R. Ream, E. L. Richmond, K. E. W. Shelden, K. L. Sweeney, R. G. Towell, P. R. Wade, J. M. Waite, and A. N. Zerbini, 2021. Alaska marine mammal stock assessments, 2020. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-421, 398 p.

National Marine Fisheries Service (NMFS), 2016. Recovery Plan for the Cook Inlet Beluga Whale (*Delphinapterus leucas*). National Marine Fisheries Service, Alaska Region, Protected Resources Division, Juneau, AK.

Olsen, S. B., Lundhede, T. H., Jacobsen, J. B., & Thorsen, B. J., 2011. Tough and easy choices: testing the influence of utility difference on stated certainty-in-choice in choice experiments. *Environmental and Resource Economics*, *49*(4), 491-510.

Richardson, L., & Loomis, J., 2009. The total economic value of threatened, endangered and rare species: an updated meta-analysis. *Ecological Economics*, *68*(5), 1535-1548.

Smith, V.K., 1993. "=Welfare effects, omitted variables, and the extent of the market. *Land Economics* 69(2): 121-131.

Swait, J., & J. Louviere, 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, *30*(3), 305-314.

Train, K. E., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Vásquez-Lavín, F., Carrasco, M., Barrientos, M., Gelcich, S., & Oliva, R. D. P., 2021. Estimating discount rates for environmental goods: Are People's responses inadequate to frequency of payments? *Journal of Environmental Economics and Management*, *107*, 102446.

Wallmo, K., & Lew, D. K., 2012. Public willingness to pay for recovering and downlisting threatened and endangered marine species. *Conservation Biology*, *26*(5), 830-839.

Wallmo, K., & Lew, D. K., 2016. A comparison of regional and national values for recovering threatened and endangered marine species in the United States. *Journal of Environmental Management*, *179*, 38-46.

Table 1. Characteristics of the sample and population

Variable	Description	Sample (n=1,316)	Alaska**
Gender	% male	51.7	52.2
Full-time workers	% of sample working full-time	47.7	46.4
Education – HS or less (LOWEDUC)	% with high school diploma or equivalent or less	18.6*	38.4
Education – grad school+	% with at least some graduate or professional school	25.2*	8.4
Ethnicity – white (WHITE)	% white/caucasian	81.2*	66.9
Ethnicity – Alaska Native/Am. Indian	% Alaska Native or American Indian	16.0	14.1
Home ownership	% owning own home/residence	79.7*	63.8
Age	Median age (years)	53.2*	35-44***
Income	Median household income ($)	70,000	70,760
Confidence (CONF)	Mean score (on 5-point scale) with 1 = not at all confident and 5 = extremely confident in choice experiment responses	3.5	n/a
Response propensity deviation (RPDEV)	Mean of deviation of predicted individual response propensity from mean propensity	-0.0083	n/a

*Sample estimate is statistically different from the population level at the 5% level.

**2009-2013 American Community Survey 5-year estimate. See https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t.

***Median age category

Table 2. Model results for MXL-Base, MXL-Adj, MXL-Sbias, and MXL-Demo models

	MXL-Base		**MXL-Adj**		**MXL-Sbias**		**MXL-Demo**	
Parameter	Estimate	Std error	Estimate	Std error	Estimate	Std error	Estimate	Std Error
<u>Mean parameter</u>								
Status quo (SQ)	-1.6303	0.2952***	-3.3005	0.3692***	-3.4625	0.3943***	-2.6925	0.6010***
SQ×RPDEV						3.1368**		
ln(RR+1)	0.8814	0.1810***	4.1079	0.7194***	3.8949	0.6842***	2.524	0.4214***
×LOWEDUC								0.5401***
THR	0.3703	0.1246***	0.5095	0.1998**	0.5511	0.2124***	0.856	0.3122**
×LOWEDUC								0.5744**
REC	0.7255	0.2354***	1.1721	0.4195***	1.1771	0.4406***	1.233	0.5532*
×LOWEDUC								0.9455
×RPDEV						3.5400**		
COST (υ=ln(γ))	-3.4155	0.0940***	-2.9976	0.1088***	-2.9733	0.1098***	-2.8122	0.1172***
DISC (ln(r))	1.7636	0.6974**	1.4937	0.4283***	1.4562	0.4116***	1.8685	0.6058***
ANA_RR			0.169	0.135	0.2709	0.1433*	1.6278	0.2882***
ANA_STAT			7.8925	328.4203	3.6594	4.7143	11.593	18100.82
CONF13 (υ_{13})			-0.5522	0.0978***	-0.58	0.1006***	-0.7996	0.1075***
CONF4 (υ_4)			-0.2686	0.1039***	-0.2883	0.1071***	-0.4722	0.1117***
Mean log-likelihood value	-4.50205	-7.3831						
Likelihood ratio index (LRI)	0.3718							
AIC	11881.4							
BIC	11964.3	-4.47925		-4.4742		-4.4650		

Note: *, **, and *** denote statistically different from zero at the 10, 5, and 1% levels, respectively. Estimation sample based on responses from 1,316 respondents (four CE questions each). Estimated Choleski matrix parameters reported in Appendix Table A2.

		0.3750		0.3757		0.3770
		11829.4		11820.0		11797.8
		11933.0		11934.0		11917.0
	-8.0118					
	-2.058					

Table 3. Household-level welfare estimates ($/household) for full recovery and marginal values

	Full recovery			Marginal value of extinction risk reduction			Marginal value of THR			Marginal value of REC		
Model	Lower bound	Mean	Upper bound	Lower bound	Mean	Upper bound	Lower bound	Mean	Upper bound	Lower bound	Mean	Upper bound
MXL-Base	112.14	227.98	338.89	36.25	55.66	76.44	14.17	35.10	56.28	-97.05	-33.37	24.76
MXL-Adj	295.24	408.70	520.21	64.85	87.92	112.64	0.00	15.07	46.75	-51.54	-2.96	38.51
MXL-Sbias	280.15	394.56	510.94	62.30	84.09	106.85	0.25	26.20	49.55	-52.51	0.23	48.33
MXL-Demo	117.01	221.24	322.03	29.24	49.39	68.14	0.00	10.43	39.02	-70.19	-12.01	25.09

Note: Lower and upper bounds are for 95% confidence intervals of the mean WTP calculated using the Krinsky-Robb simulation method with 1,000 iterations. All values are in 2013 dollars.

Table 4. Aggregate welfare estimates (in millions of 2013 dollars) for full recovery under different aggregation assumptions (MXL-all model only)

Model/Aggregation assumption	Full population (POP)	POP × estimation sample response rate
MXL-Base	57.4 (28.2, 83.4)	19.2 (9.5, 28.6)
MXL-Adj	103.0 (74.4, 131.0)	34.5 (24.9, 43.9)
MXL-Sbias	99.4 (70.6, 128.7)	N/A
MXL-Demo	55.7 (29.5, 81.1)	N/A

Note: Lower and upper bounds are for 95% confidence intervals of the aggregate WTP calculated using the Krinsky-Robb simulation method with 1,000 iterations.

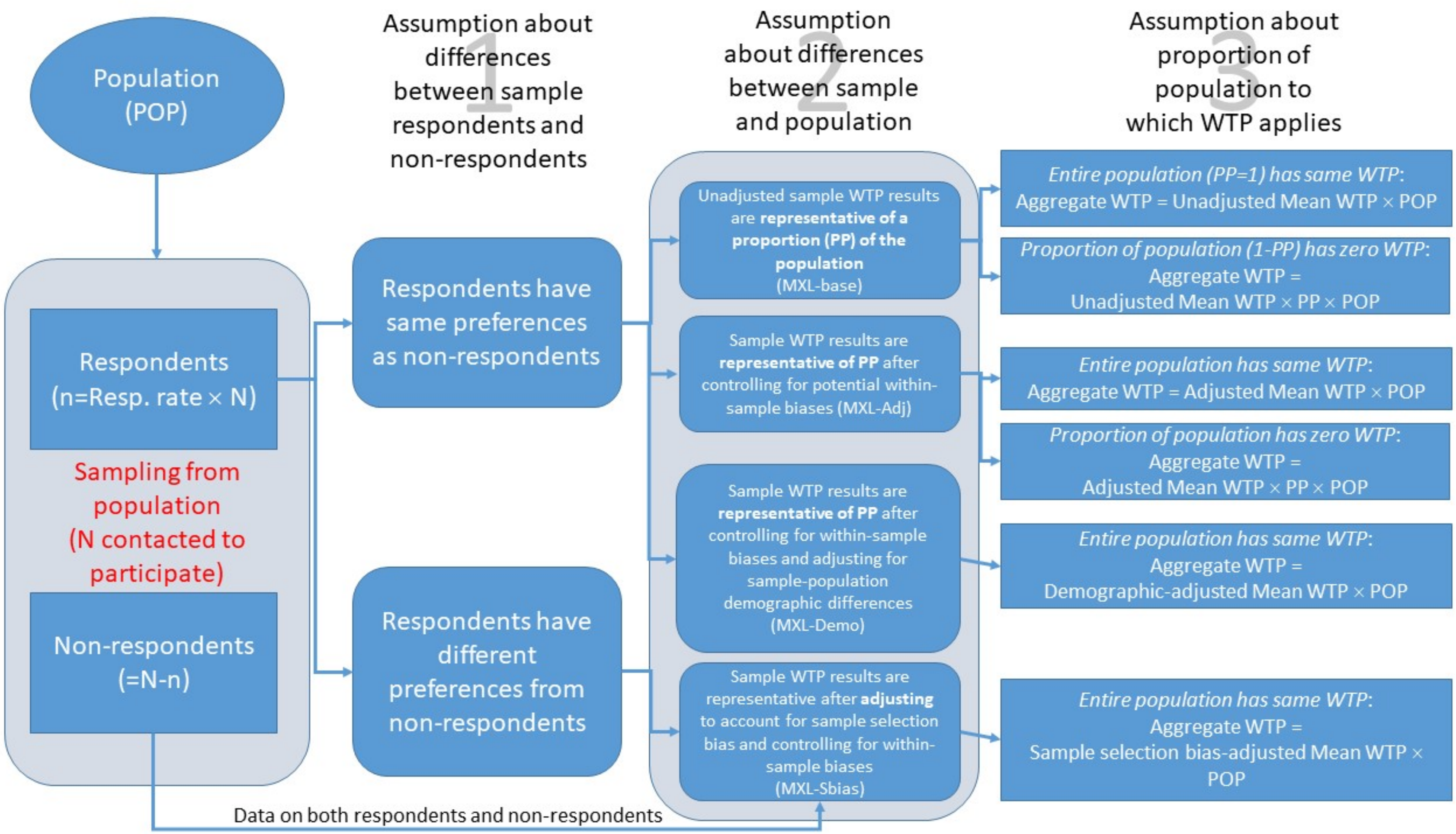

Figure 1. Flowchart of assumptions about the sample and population that contribute to the construction of aggregate willingness-to-pay (WTP) estimates. POP = population size, N = contacted sample, n = respondent sample, N-n = non-respondents, PP = population proportion for which sample-based WTP is representative.

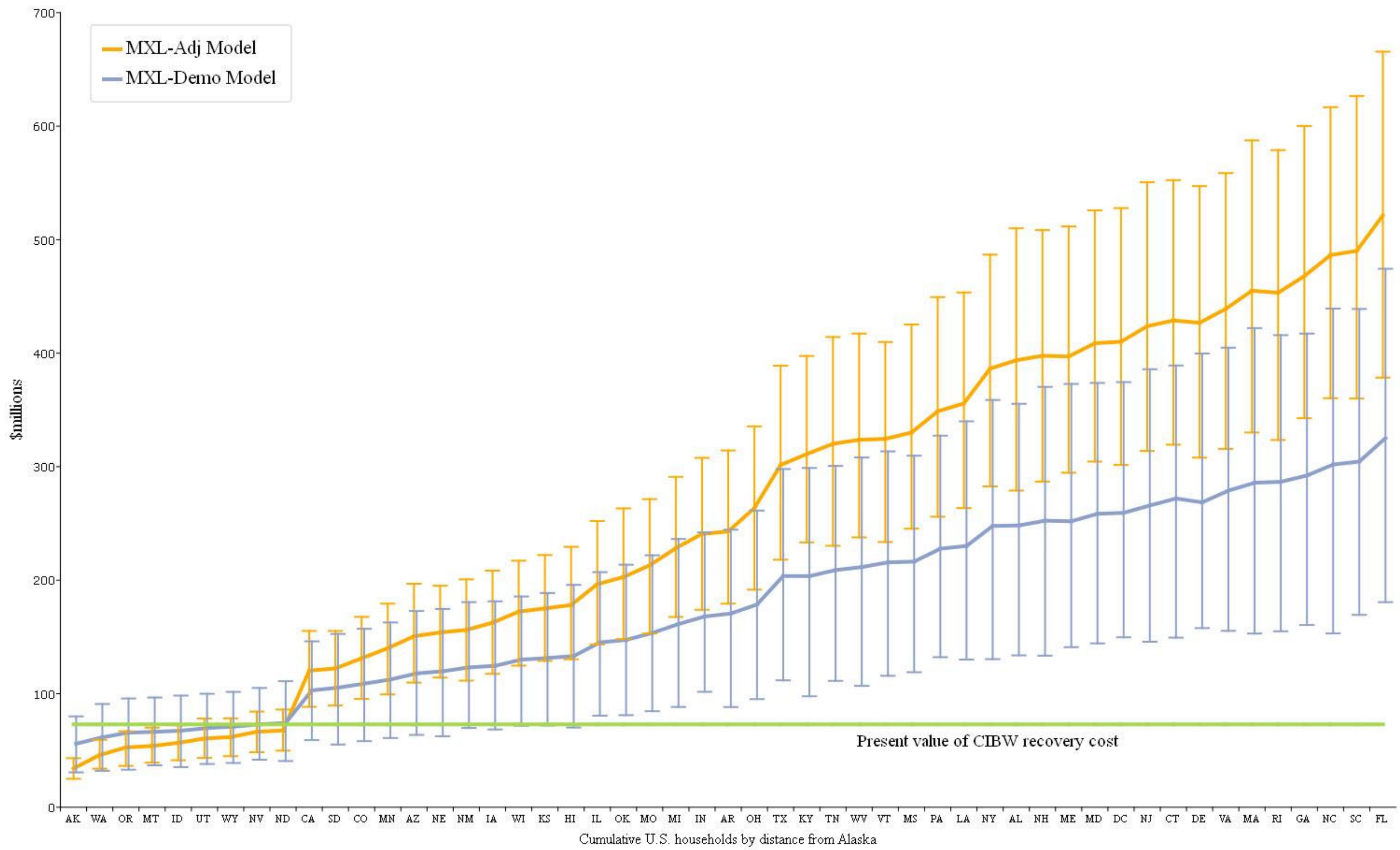

Figure 2. Cumulative U.S. aggregate present value of WTP for recovery (and 95% confidence intervals) calculated under the conservative assumption that non-Alaska states have 1% of households with a WTP equal to the mean WTP for Alaska, with all other households having a zero WTP. Aggregate values are presented using the MXL-Demo Model (blue line) and the MXL-Adj Model (yellow line). The green line indicates the total present value of recovery costs.

APPENDIX

Symmetric Triangular Distribution

Our utility specifications allow for symmetric positively-constrained triangular distributed cost parameters. The symmetric triangular distribution is described in Hensher and Greene (2003). The more general triangular distribution (asymmetric triangular distribution) is presented in Dekker (2016). Figure A.1 shows the density function, which increases from 0 to c, then decreases thereafter until b. The density is zero for values at or below zero, as well as at and above b. The center of the distribution (c) equals b/2 and is both the mean and mode. The standard deviation is $\frac{b}{2\sqrt{6}}$. The distribution is defined by a single hyperparameter, b (or alternatively c).

For the maximum simulated likelihood procedure, symmetric positively-constrained triangular distributed cost parameters are drawn as follows: First, we draw *R* standard uniformly-distributed Sobol quasi-random draws, $\mathbf{u} = u^1, u^2,\ldots,u^R$. Second, for the rth draw the random triangular distributed variate (υ^r) is calculated as $\upsilon^r = \sqrt{u^r \cdot \frac{b^2}{2}}$ when $u^r < 0.5$ and $\upsilon^r = b - \sqrt{(1-u^r) \cdot \frac{b^2}{2}}$ when $u^r \geq 0.5$. For estimation, we estimate $\upsilon = \ln(b)$, such that $b = \exp(\upsilon)$.

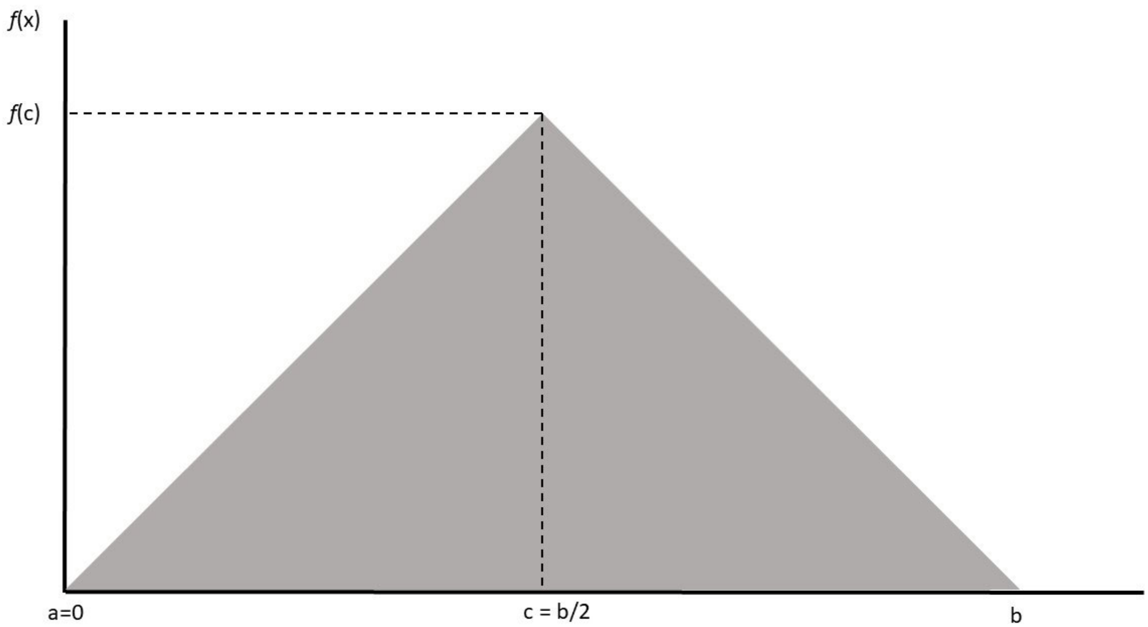

Figure A1. Symmetric positively-constrained triangular distribution.

An Approximately Normal Confidence Interval (CI)-based Test

We use a simple confidence interval (CI)-based test for testing for statistical differences between the means of two empirical willingness-to-pay distributions calculated using the Krinsky-Robb simulation approach (Krinsky and Robb 1986), as an alternative to the computationally-intensive method of convolutions-based approach (Poe et al. 2005) or to the Delta method. Poe, Severance-Lossin, and Welsh (1994) showed that assessing whether there is a statistical difference between means $\underline{X}$ and $\underline{Y}$ by determining if the (1- α)% confidence intervals for them do not overlap leads to an overly conservative test at a level of significance less than α. In fact, they show that under the null hypothesis of no difference and assuming equal variances, one can show that the non-overlapping confidence intervals assumes a critical value of ($2 \times Z_{\alpha/2}$/sqrt(2)), where $Z_{\alpha/2}$ is the Z value evaluated at α/2. We extend this intuition to assess the statistical difference by deriving the distance between means that is consistent with the level of α desired for the test. Recall that for a normally distributed variable X a (1-α)% CI for sample mean $\underline{X}$ with a population mean of μ, standard deviation of σ_x and sample size n is

$$\underline{X} \pm Z_{\alpha/2} \cdot \sigma_x/\sqrt{n}. \qquad \text{(A.1)}$$

Note that we are comparing mean welfare estimates from different models or different welfare scenarios, so the sample size (*n* = number of K-R mean WTP values) is the same but the standard deviations will differ. The (1-α)% CI for the difference between two sample means, $\underline{X}$ and $\underline{Y}$, when the sample size is the same but the standard deviations differ is

$$(\underline{X} - \underline{Y}) \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}{n}}\ . \qquad \text{(A.2)}$$

Testing for the statistical difference in means implies the test statistic:

$$Z = \frac{\underline{X} - \underline{Y}}{\sqrt{\frac{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}{n}}} = \frac{\underline{X} - \underline{Y}}{\sqrt{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}} \qquad Z \sim N(0,1)\ . \qquad \text{(A.3)}$$

The critical distance between means associated with an alpha of α is $D = Z_{\alpha/2} \cdot \sqrt{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}$. Normalizing this reveals the critical normalized distance, $Z = \frac{D}{\sqrt{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}}$, which under the null hypothesis is distributed standard normal. Thus, if we want a test that has a level of significance of α = 0.05, then we want to have the critical normalized distance Z = 1.96. To ensure this, we can find a value for D that results in a $Z = Z_{\alpha/2}$. The general solution to this distance (D_0) is $D_0 = Z_{\alpha/2} \times \sqrt{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}$. Given this, the test is implemented in two steps:

1. For the desired alpha level, calculate $D_0 = Z_{\alpha/2} \cdot \sqrt{\sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2}$. This is the critical distance between the means that is consistent with the alpha level. If the means are further apart than D_0 then we reject the null hypothesis of equal means. If they are closer, we cannot reject the null hypothesis.

2. We then compare the distance between the lowest bound of either (1-α)% CI bound for $\underline{X}$ and $\underline{Y}$ and the highest bound of the (1-α)% CI's (i.e., the max range of the union of the two CIs). Define this distance as D_{max}.
   a. If $D_{max} > D_0 + Z_{\alpha/2} \cdot (\sigma_{\underline{x}} + \sigma_{\underline{y}})$, then this is equivalent to the critical distance between means being exceeded and we reject the null hypothesis.
   b. If $D_{max} \leq D_0 + Z_{\alpha/2} \cdot (\sigma_{\underline{x}} + \sigma_{\underline{y}})$, then we cannot reject the null hypothesis.

This approach assumes normality of the empirical mean WTP distributions, so it is very similar to the delta method but instead of starting with the estimated function (in this case the utility function), we start with the empirical distribution of the parameter (here WTP) of interest.

References

Dekker, T., 2016. Asymmetric triangular mixing densities for mixed logit models. *Journal of Choice Modelling*, *21*, 48-55.

Hensher, D. A., & Greene, W. H., 2003. The mixed logit model: the state of practice. *Transportation*, *30*(2), 133-176.

Krinsky, I., & Robb, A. L., 1986. On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 715-719.

Poe, G. L., Severance-Lossin, E. K., & Welsh, M. P., 1994. Measuring the difference (X—Y) of simulated distributions: A convolutions approach. *American Journal of Agricultural Economics*, *76*(4), 904-915.

Poe, G. L., Giraud, K. L., & Loomis, J. B., 2005. Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, *87*(2), 353-365.

Table A1. Response propensity (sample selection) model

Parameter	Estimate	Std err
Constant	-0.0391	0.0362
PO Box address	-0.4582	0.0775***
High rise residence	-0.4078	0.1104***
Adjacent to Cook Inlet	0.1394	0.0712**
One adult in household	-0.2101	0.0701**
Low income (<$30K)	0.0399	0.0779
High income (≥$150K)	0.1122	0.0756
Mean Log-Likelihood	-0.6763	
LRI	0.0243	
AIC	5307.93	
BIC	5351.84	
N	3914	

Notes: *, **, and *** denote statistically different from zero at the 10, 5, and 1% levels, respectively. All non-constant parameters are associated with dummy variables. Data are compiled from Census data by Ipsos. The logit model estimates the probability of responding to the survey. Individual-level predicted values are calculated and used to form a response propensity deviation that enters the sample selection bias-corrected model (MXL-Sbias) as a covariate interacted with utility parameters.

Table A2. Estimated Choleski matrix parameters from MXL-Base, MXL-Adj, MXL-Sbias, and MXL-Demo models

	MXL-Base		**MXL-Adj**		**MXL-Sbias**		**MXL-Demo**	
Parameter	Estimate	Std error	Estimate	Std error	Estimate	Std error	Estimate	Std Error
SD(SQ)	5.2114	0.3887***	4.2339	0.4484***	4.2111	0.4602***	9.500	1.1233***
Chol(SQ,RR)	2.6214	0.2492***	2.3797	0.5675***	1.2629	0.5628**	5.908	0.6718***
SD(ln(RR+1))	2.4938	0.1210***	8.9061	1.2092***	8.5629	1.1210***	4.440	0.5222***
Chol(SQ,THR)	0.0505	0.2604	0.9	0.3505**	1.1245	0.3946***	0.513	0.4747
Chol(RR,THR)	0.0832	0.1746	1.2913	0.9158	1.4377	0.8966	0.210	0.4412
SDTHR	1.9297	0.2087***	2.9439	0.3885***	3.08	0.4085***	-3.331	0.6262***
Chol(SQ,REC)	1.244	0.3884***	2.3777	0.6372***	2.619	0.6956***	3.059	1.0605***
Chol(RR,REC)	0.4714	0.3144	4.1943	1.7120**	4.6283	1.6710***	0.892	0.8712
Chol(THR,REC)	1.3322	0.4119***	2.1434	0.5343***	2.0737	0.5849***	-2.159	0.9979**
SDREC	1.701	0.2930***	1.8809	0.4299***	1.8663	0.4589***	-2.647	0.6943***

Q12 **Here is the current program with two alternatives. Which alternative do you most prefer and which alternative do you least prefer?** *Please indicate your responses below the table.*

	Alternative A Current program	Alternative B	Alternative C
Population status in 50 years…..... (endangered now)	**Endangered**	**Threatened**	**Threatened**
Risk of extinction by the year 2112…………………………….. (25% now)	**25%**	**15%**	**10%**
Added cost to your household (<u>one-time payment</u>)…………....	**$0**	**$40**	**$50**
	Alternative A	**Alternative B**	**Alternative C**
Which alternative do you prefer the most? *"X" only one box*	☐	☐	☐
Which alternative do you prefer the least? *"X" only one box*	☐	☐	☐

Figure A2. Example of choice experiment question (Pay1 survey version)

Q17 **These questions were asked to obtain public input for decision makers to consider along with information from scientists and planners. People feel differently about how confident they are with their selection of alternatives and the costs they would have to pay.**

How confident are you that your answers in Q12 through Q15 accurately reflect how you feel about the alternatives for protecting Cook Inlet beluga whales? *(Please "X" only one box).*

Not at all confident	Slightly confident	Somewhat confident	Very confident	Extremely confident
☐	☐	☐	☐	☐

Figure A3. Response certainty question.