University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 8/2022 (384)

ON THE INFERENCE ABOUT WILLINGNESS TO PAY DISTRIBUTION USING CONTINGENT VALUATION DATA

Mikołaj Czajkowski
Ewa Zawojska
Norman Meade
Ronaldo Seroa da Motta
Mike Welsh
Ramon Arigoni Ortiz

Warsaw 2022

WORKING PAPERS 8/2022 (384)

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

On the inference about willingness to pay distribution using contingent valuation data

Mikołaj Czajkowski[1], Ewa Zawojska[1], Norman Meade[2], Ronaldo Seroa da Motta[3], Mike Welsh[2], Ramon Arigoni Ortiz[2]

[1]*University of Warsaw, Faculty of Economic Sciences*
[2]*Independent consultant*
[3]*State University of Rio de Janeiro*
* *Corresponding authors: mc@uw.edu.pl, ewa.zawojska@uw.edu.pl*

Abstract: Although contingent valuation (CV) is one of the main sources of estimates of non-market values of environmental goods, little guidance exists regarding parametric approaches for modelling CV data, which would reliably estimate willingness-to-pay (WTP) values based on binary choice, payment card or open-ended preference elicitation data, among others. CV studies often rely on relatively simple approaches to modeling stated preference responses, without examining alternative modelling specifications. Lower-bound, non-parametric estimates seem to be preferred in legal cases, while studies that apply parametric approaches often select a specification among a limited set of commonly used distributions. To enhance the reliability of CV-based WTP estimates, we propose to adopt a more flexible approach to parametric modelling of a WTP distribution, by considering a wide range of parametric model specifications. We demonstrate the advantages of the proposed approach using databases from two large CV studies: the eutrophication reduction valuation for the Baltic Sea Action Plan and the Deepwater Horizon natural resource damage assessment. We find non-negligible differences in WTP value estimates across models with different assumed parametric distributions, and we observe the variation in the values to decrease when only better-fitting models are considered. This emphasizes the need for cautiously identifying the model best fitting to the data, instead of choosing a specification ad hoc without taking into account alternative parametric distributions. Focusing on the best-fitting parametric specifications, we provide alternative WTP value estimates for the two empirical cases studied.

Keywords: contingent valuation, parametric modelling, stated preferences, willingness to pay, welfare estimates

JEL codes: D61, H41, H43, Q51

Acknowledgements: The authors gratefully acknowledge the support of the National Science Centre of Poland (projects no. 2018/30/E/HS4/00388 and 2017/25/B/HS4/01076). Ewa Zawojska gratefully acknowledges the support of the Polish National Agency for Academic Exchange within the Bekker programme. The collection of the Baltic Sea Action Plan data was funded by four Finnish Ministries (the Ministry of Agriculture and Forestry, Ministry of Environment, Ministry of Transport and Communications, and Ministry of Finance), the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), the Danish Strategic Research Council, the German Federal Environment Agency (UBA, Fkz 3710 25 202), and the Swedish Environmental Protection Agency. We would like to thank the BalticSTERN Secretariat and the team consisting of researchers from the nine Baltic Sea countries for the joint work on the data collection and the willingness to share the data.

1. Introduction

Stated preference studies remain the main source of estimates of non-market values of environmental goods (Carson and Czajkowski 2014) and are essential for capturing the passive-use component of the values (Freeman, Herriges, and Kling 2014). Studies of that type elicit preferences with attribute-based approaches, such as discrete choice experiments (DCEs) involving a series of choice tasks with policy alternatives defined by attributes, or non-attribute approaches, such as a single binary choice or a payment card, among others. To ease further reference and comply with common nomenclature (e.g., Johnston et al. 2017), we henceforth refer to the non-attribute approaches as contingent valuation (CV).[1] While DCEs deliver information on values of individual attributes of goods and, hence, can help evaluate several possible policy outcomes and design policies balancing different quantifiable goals, CV is particularly useful to estimate values of goods inseparable into individual attributes (i.e., perceived as a whole) and provides a single estimate of a welfare change, such as a damage assessment in a litigation case. The methods are widely applied, as illustrated by about 4,700 studies using CV and nearly 6,200 studies employing DCEs, as indexed by Google Scholar in 2019 (Hanley and Czajkowski 2019). However, although there has been a considerable research interest in improving modelling techniques for DCE data (Mariel et al. 2020), there is little guidance regarding robust econometric approaches that would reliably estimate the population distribution of willingness to pay (WTP) based on CV data, including quantities especially relevant from a policy perspective such as mean and median WTP.

Applied CV studies typically rely on relatively simple approaches to modeling stated preference responses without considering various modeling alternatives, which may lead to biased results. In legal cases (e.g., damage assessment), conservative (lower-bound) non-parametric estimates seem to be preferred. For example, in one of the highest-profile studies of the past decade, focused on assessing the value of natural resources lost in the aftermath of the 2010 Deepwater Horizon oil spill, Bishop et al. (2017) use Lewbel (2000) and Watanabe (2010) estimator, which in their application[2] is equivalent to the ABERS (Ayer et al. 1955) and

---

[1] We acknowledge that stated preference literature is not uniform with respect to the nomenclature. In place of the differentiation described above, DCEs may be considered as a subset of CV when the latter is viewed as "a survey approach to place an economic value on a public good", "independent of any particular elicitation method" (Carson and Louviere 2011, p. 541-542).

[2] There was a finite set of bid levels and no monotonicity violations were observed for the increased probability of voting 'no' with increasing bid levels.

Turnbull (1976) methods. Provided its correct application, the Lewbel-Watanabe approach never results in model-driven value overestimation and, hence, may be easier to defend in a court. On the other hand, CV studies that apply parametric methods often rely on a limited set of parametric distributions, although there is no theoretical guidance on the choice of a parametric distribution for modelling WTP values, other than the aim of obtaining the best fit to data, theoretical consistency and mathematical tractability (e.g., Kerr 2000). The most commonly employed distributions are logistic and normal functions in a raw or logged form, and when zero WTP can be ruled out, the lognormal and Weibull distributions are fairly common (DWH Total Value Team 2015; Langford et al. 1998; Bateman et al. 1995). More flexible non-parametric, semi-parametric or parametric approaches, which may relax conventional rigid modelling structures, are not frequently used in CV studies. Yet, such more flexible methods can lead to a better fit of the distribution to observed data and, thus, to more valid and reliable estimation of mean and median WTP values.

This paper argues for a more flexible approach to parametric modelling of WTP values based on CV data. Parametric modelling offers a straightforward technique and can provide estimators with desirable statistical properties (e.g., high efficiency) (e.g., Creel and Loomis 1997)—these characteristics may make them attractive compared to non- and semi-parametric approaches. However, the key to obtaining unbiased welfare estimates with parametric modelling—and at the same time the major challenge—is to define the correct specification of the distribution. This need calls particularly for flexibility in the modelling by relaxing restrictive conventional assumptions and rigid structures to enhance the fit of the distribution to observed data (Cameron and Quiggin 1994; Araña and León 2005). Selection of a well-fitting specification is fundamental for deriving valid and reliable value estimates for policy making as evidence suggests that some statistics obtained with parametric modelling, such as the mean, may be sensitive to distributional assumptions (Carson, Wilks, and Imber 1994).

While economic theory provides hardly any motivation for making specific distributional assumptions for the correct parametric specification, stated preference literature appears to support an empirical approach to the specification selection by testing a range of distributional forms and conducting robustness checks (Kerr 2000; Creel and Loomis 1997). However, few empirical studies appear to undertake this path when determining the distribution for parametric modelling. In fact, there are not many studies investigating the importance of distributional assumptions. In the context of a fat-tails problem in CV data modelling, Kerr (2000) provides an initial investigation of parameterization effects and analyzes the performances of four common distributions (logistic, log-logistic, exponential and Weibull) and their modifications based, for example, on scaling (multiplying predicted probabilities by a constant factor) and shifting (adding some fixed amount to predicted probabilities). Another notable example is Werner (1999) who compares mixture models incorporating various distributional specifications: generalized gamma, Weibull, gamma and exponential, in order to improve accounting for zero WTP responses. Finally, there are some studies that employ new parametric modelling approaches that are potentially more flexible but typically also more difficult to estimate. For example, Layton and Moeltner (2005) implement a mixed model with a gamma distribution for WTP and a lognormal distribution for a scale parameter to model repeated dichotomous choice CV responses.

Grounded in the (arguably) surprising paucity of CV studies examining the sensitivity and fit of the wide range of familiar parametric distributions, our study investigates this approach focusing on the distribution selection guided by the goodness of fit to observed data. We examine the performance of numerous parametric specifications to modelling the WTP distribution based on CV responses. The comparison of numerous specifications based on different parametric distributions allows us to examine the extent to which ad hoc model selection can affect estimates of mean WTP when compared to WTP derived from distributions best fitting to observed data. Estimation results indicate a substantial variation in WTP estimates across models differing in (i) the assumed parametric distribution and (ii) the inclusion of a zero-inflation component (or lack thereof) to capture zero WTP values. These findings suggest that choosing a distribution ad hoc, without a proper consideration of various model specifications, may not result in the best model-driven WTP estimates. Furthermore, our empirical analysis indicates that the variation in estimates of WTP is notably smaller when only better-fitting models are considered, which can signal higher reliability of the estimates derived from the better-fitting models. In addition, for the empirical data considered here, commonly used distributions in parametric modelling of CV responses, such as logistic and normal distributions in a raw form, do not emerge to be among the distributions yielding the best fit to the data. This further reaffirms the importance of avoiding ad hoc model selection.

These findings cause us to argue for a flexible approach to parametric modelling of a WTP distribution based on CV data. In our study, flexibility, made possible thanks to advancements in computing, numerical optimization and the availability of parametric representations of empirical distributions, involves searching a wide range of parametric representation that fits data best. Improving estimation methods can generate more precise estimates of mean WTP, which in turn may result in more economically-efficient policy decisions. By suggesting a practical tool for enhancing the reliability of CV value estimates, the paper contributes to increasing accuracy of the estimates (Bishop and Boyle 2019).

We examine the performance of fitting various parametric specifications to CV data and demonstrate possible advantages of the proposed flexible approach. We do this using databases from two widely-known CV studies: the eutrophication reduction valuation for the Baltic Sea Action Plan and the Deepwater Horizon natural resource damage assessment. The former data comes from an investigation evaluating a program of reducing nutrient loadings to the Baltic Sea (Ahtiainen et al. 2014). That study focused on the social value of the Baltic Sea eutrophication reduction associated with the implementation of the Baltic Sea Action Plan (BSAP; HELCOM 2013). With surveys administered to nearly 10,000 respondents in nine Baltic Sea countries, it remains the most comprehensive and influential valuation study of eutrophication to date. The second data set comes from the study assessing environmental damages following the Deepwater Horizon oil spill (Bishop et al. 2017). The goal of the study was to provide a monetary value of the natural resource damage from the oil spill in support of a governmental lawsuit. 3,656 U.S. households were surveyed to evaluate the natural resource loss in the aftermath of the largest marine oil spill in U.S. history. While the Baltic Sea study used a payment card to elicit WTP values, the Deepwater Horizon study employed a single binary choice. This way, we present the application of the flexible approach to parametric modelling in CV research based on two common preference elicitation formats. Identifying the best-fitting parametric model allows us to provide alternative value estimates for each of the two data sets.

The remainder of the paper is structured as follows. Section 2 presents the econometrics of modelling CV data. Section 3 provides details about the two data sets subsequently used in the empirical part of the paper. Survey 4 discusses the estimation results, and Section 5 concludes.

2. Econometrics of modelling CV data

Inferring WTP values from preferences stated in surveys is grounded in economic theory (Freeman, Herriges, and Kling 2014). Original models of WTP are based on the random utility framework (McFadden 1974; Hanemann 1984) and infer the values indirectly (using the estimated utility function).[3] However, when the goal of a study is to provide a single estimate of WTP, a direct estimation of WTP may be preferred (Cameron 1988). Even though not all directly modeled WTP functions have straightforward representations as utility function differentials, they are more flexible and allow for avoiding complicated, and in some cases implausible, distributional consequences of indirect estimation of WTP based on preference functions.[4]

The most common CV preference elicitation formats include a single binary choice, a double-bounded binary choice, a payment card and an open-ended question. While the single binary choice question is the most straightforward to incentivize truthful preference disclosure (Carson and Groves 2007; Carson, Groves, and List 2014), the other formats reveal more information about respondents' underlying WTP and, hence, can be considered as alternatives in bias-variance trade-offs (Alberini 1995; Johnston et al. 2017). Recent studies by Vossler and Holladay (2018) and Vossler and Zawojska (2020) show that all these formats can deliver statistically indistinguishable WTP estimates when they are designed such that consistent economics incentives are provided, making the truthful preference statement the best response strategy for a respondent.

Respondents' answers to preference elicitation questions can be translated to bounds of their directly unobservable WTP. In the case of a single binary choice, a 'yes' response to a specific bid level indicates that the lower bound of the WTP is the bid amount, and the upper bound is unknown, potentially limited by the individual income. A 'no' response reveals that the respondent's WTP is between zero (the lower bound)[5] and the presented bid (the upper bound). If a double-bounded binary choice is used, 'yes' and 'no' responses are translated into bounds in a similar manner. For a payment card, a respondent's choice of the highest bid she would be willing to pay indicates that her maximum WTP lies in the range between the selected bid (the amount she is willing to pay; the lower bound) and the next bid (the amount she is not willing to pay; the upper bound). All these formats can therefore be considered interval-type data (Cameron and Huppert 1989). In the case of an open-ended question, the two bounds

[3] We refer the reader to excellent seminal textbooks such as Bateman et al. (2004), Champ, Boyle, and Brown (2017) and Haab and McConnell (2003) for a detailed presentation of welfare theory and the utility-maximization-based framework for non-market valuation.

[4] This is parallel to estimating DCE models in preference- or WTP-space (Scarpa, Thiene, and Train 2008).

[5] In some cases, a negative lower bound could be appropriate.

coincide and are equal to WTP stated by a respondent. The information on WTP bounds of individuals can be used to fit a parametric distribution describing the population WTP.

Assuming that the WTP distribution is of particular form (e.g., normal) with unknown parameters characterizing the distribution (e.g., a mean and a standard deviation), the probability of observing a particular answer is equal to the cumulative distribution function (CDF) of the assumed distribution evaluated at the upper bound (i.e., the probability that WTP is smaller than the upper bound) less the CDF of this distribution evaluated at the lower bound (i.e., the probability that WTP is smaller than the lower bound). The result of this subtraction is the probability that a respondent's WTP lies between the lower and the upper bound. The parameters of the selected parametric distribution can be found by maximizing the sum of these probabilities for the obtained answers of all respondents.

Formally, let us represent the probability that individual i's WTP lies between the observed lower bound $\left(b_{i,s}\right)$ and the upper bound $\left(b_{i,s+1}\right)$ as

$$P\left(b_{i,s} \leq WTP_i < b_{i,s+1}\right) = CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right), \tag{1}$$

where $CDF(\cdot)$ denotes a cumulative distribution function of the considered WTP distribution[6] and $\boldsymbol{\beta}_i$ is a vector of the distribution parameters (e.g., for a normal distribution, $\boldsymbol{\beta}_i$ consists of a mean and a standard deviation).[7]

The parameters of the distribution can be estimated using the maximum likelihood method. The probability specified in (1) expresses individual i's contribution to the likelihood function, while the log-likelihood function for a sample of N individuals can be formulated as:

$$\log L = \sum_{i=1}^{N} w_i \log\left[CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right)\right], \tag{2}$$

where w_i represent weights that account for the possible over- or underrepresentation of specific individuals in the sample relative to the target population.

---

[6] Actually, $CDF(x)$ is the probability that a random variable takes a value less than or equal to *x*. Although our representation in (1) slightly differs from this definition, it does not have any actual impact as the likelihood of a continuous random variable taking any specific value is zero.

[7] When the calculated difference in the CDF values is equal to zero, we operationalize the numerical estimation by using the value of the probability density function (PDF) evaluated at the lower bound instead of the difference in the CDF values.

The above equations are conditional on selecting a parametric distribution, for which CDFs are calculated. However, economic theory does not determine what parametric distribution is best for approximating the distribution of WTP in the population. Thus, the approach proposed in this paper relies on trying many parametric distributions to choose the one that fits the data best. Because the distributions vary with respect to the number of parameters and those with more parameters can lead to a better fit to the data than those with fewer parameters, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) appear as preferred measures for comparing the models, rather than using the value of the log-likelihood function, to account for the cost of additional parameters (and to penalize overfitting).

There is no theory guiding the choice of a parametric distribution—what parametric distribution fits the WTP distribution in the population is an empirical question. A flexibility in the approach to parametric modelling of CV data requires considering all typical parametric distributions; we included: normal, logistic, extreme value, generalized extreme value, t location-scale, uniform, Johnson SU, exponential, log-normal, log-logistic, Weibull, Rayleigh, gamma, Birnbaum-Saunders, generalized Pareto, inverse Gaussian, Nakagami, Rician, Johnson SB, Johnson SL, Poisson, and negative binomial. For the distributions with support in negative numbers, censoring negative values to zero can be used when simulating mean WTP in the sample.

In CV studies, it is usually found that there is a large share of respondents whose WTP is equal to zero, coupled with observing relatively few very small WTP amounts.[8] This can be captured in parametric modelling of the data by including a jump discontinuity in a probability density function (PDF) of any parametric distribution. There are a few econometric approaches to account for such a spike in the PDF, either by censoring the parametric distribution (Kriström 1997), truncating it or adopting a hurdle model (Cragg 1971) or using a zero-inflation/mixture model (Greene 2011).[9] We adopt the most flexible (zero-inflation) approach, in which respondents' WTP is modelled as a mixture of a Bernoulli distribution (a point mass at zero)

[8] This is particularly expected for payment card and open-ended elicitation formats (Carson and Czajkowski 2014). In discrete choice formats the issue may be indicated by the implied proportion of WTP that is zero (or less).

[9] The difference between the hurdle models and the zero-inflated models is that in zero-inflated models the zeros are modelled using a two-component mixture model. With a mixture model, the probability of the variable being zero is determined by both the main distribution and the mixture weight. For notable examples of CV studies accounting for zero responses see Werner (1999) and Strazzera et al. (2003).

and a given parametric distribution, allowing for an over-proportional share of zero responses (Gurmu and Trivedi 1996). As a result, the log-likelihood function becomes:

$$\log L = w_i \sum_{i=1}^{N} \left\{ (1-q_i) \cdot \log\left[CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right) \right] + q_i \cdot \log\left[CDF\left(0, \boldsymbol{\beta}_i\right) \right] \right\}, \tag{3}$$

where q_i is the probability that individual i has zero WTP.[10]

For specifications with complicated parametric distributions, the estimated parameters may not represent the mean and standard deviation, like in, for example, normal distribution. In this case, the needed characteristic of the WTP distribution (e.g., the mean) and its standard error can be simulated following the parametric bootstrapping method adapted from (1991); Krinsky and Robb (1986):

(i) To account for the uncertainty with which the estimates are known, use the parameter estimates and the inverted Hessian at convergence[11] to define a multivariate normal distribution[12] and use it to draw a large number (e.g., 10,000) of new sets of parameters.

(ii) For each set of parameters estimated in step (i), draw 10,000 empirical WTP values. This follows the assumed parametric distribution of WTP (possibly with a zero-inflation component).

(iii) Observing variation in the needed characteristic (e.g., a mean) of the WTP distributions simulated in step (ii), driven by each set of parameters generated in step (i), allows for estimating the uncertainty associated with the characteristic.

Overall, the proposed approach enables us to fit a large range of parametric distributions to CV data and to choose the one with desired properties. In addition, evaluating multiple parametric distributions provides an insight into the uncertainty arising from the model selection. In what follows, we employ two empirical data sets to demonstrate the advantages of the proposed approach and examine the possibility to increase the reliability of WTP estimates by deriving them from the best-fitting distributions.

---

[10] $q(\cdot)$ is usually assumed to be a CDF of a normal or logistic distribution, leading to a probit or logit model, respectively, to account for the zero inflation. In what follow, we assume a normal distribution.

[11] This is to approximate the asymptotic variance covariance matrix.

[12] Maximum likelihood estimates are asymptotically normal.

3. Data sets

This section describes two empirical data sets that we use to illustrate the proposed flexible approach to fitting a parametric distribution to CV data. Identifying best-fitting models for the data sets help us provide alternative value estimates for the two considered cases.

3.1. Baltic Sea eutrophication reduction

The Baltic Sea is a flagship case for studying major environmental perturbations and the efficacy of various management responses. Its main environmental problems include eutrophication along with its consequences, such as deterioration of the water transparency, increased toxic algal blooms, the expansion of oxygen-minimum zones and changes in fish stock (HELCOM 2014). From the natural science perspective, the Baltic Sea is one of the most intensely studied areas in the world. The high-density data, many long-term data series and coordinated macro-regional research agenda provide crucial inputs for science-based management (Reusch et al. 2018). However, the science-based management is not possible without robust inputs from social sciences, particularly without valid estimates of economic benefits and costs of various policy actions.

Evaluation of benefits of improving the Baltic Sea environment in monetary terms has been the focus of several studies to date. Most notably, Ahtiainen et al. (2014) conducted a CV study and estimated the value of alleviating eutrophication in the Baltic Sea at 3.6 billion EUR annually. The economic value of the reductions in eutrophication has earlier been measured in the Stockholm archipelago of Sweden (Söderqvist and Scharin 2000), and in Lithuania, Poland and Sweden (Markowska and Żylicz 1999). Tuhkanen et al. (2016) estimate the value of benefits for improvements of water quality in the Estonian waters, while Pakalniete et al. (2017) provide WTP estimates for the improved coastal water quality for recreation in Latvia.

The study by Ahtiainen et al. (2014) is unique in providing the value of eutrophication reduction for all nine Baltic Sea countries. This data was collected in 2011 with a survey that aimed at estimating people’s WTP for reducing eutrophication in the open-sea areas of the Baltic Sea by 2050 as envisaged by the implementation of the Baltic Sea Action Plan. In total, 10,564 respondents from all countries around the Baltic Sea were surveyed.

To assess the Baltic Sea eutrophication reduction in monetary terms, Ahtiainen et al. (2014) employed the CV method and asked respondents to evaluate the program of meeting the nutrient load reduction targets defined in the HELCOM’s Baltic Sea Action Plan (HELCOM

2013). The questionnaire compared the improved environmental quality resulting from the program implementation against the eutrophication level as predicted for 2050 if no new investments in nutrient abatement measures were made. The effects of the eutrophication in the improvement program scenario and in the no-improvement baseline were defined on a basis of state-of-the-art marine models for the Baltic Sea (Ahlvik et al. 2014; Kiirikki et al. 2006; Kiirikki et al. 2001; Maar et al. 2011) and professional expertise by marine ecologists.

The concept of eutrophication was introduced in the survey by linking it to five ecosystem effects: water clarity, blue-green algal blooms, a condition of underwater meadows, a composition of fish species and oxygen content in deep-sea bottoms. Each effect was described on a five-step colored water quality scale, in which colors depicted different levels of the effect intensity and were labelled from “best possible water quality” to “worst possible water quality”. After explaining to respondents the meaning and construction of the scale, the improvement scenarios were shown in a form of color-coded maps that illustrated eutrophication levels in all sub-basins of the Baltic Sea in 2050. The visual representation of the scenario was supported with a verbal description.[13]

After being introduced to the nutrient loadings reduction program, respondents were asked whether in principle they would be willing to pay anything for eutrophication reduction in the Baltic Sea. This type of question might be referred to as an “in-the-market” question, because it showed whether a respondent was interested at all in having the good provided. These responses are included in the modelling in the share of people with zero WTP, who add to the jump density of the WTP distribution at zero in the zero-inflated specifications (following equation (3)).

In the preference elicitation question, respondents were asked to indicate the maximum amount they would be willing to pay for the improvement using a payment card. The exact wording of the question (upon translation into English) was: “What is the most you would be willing to pay every year to reduce eutrophication in the Baltic Sea as shown in the maps [presented in Figure A1 in Appendix A]?”. The payment mechanism described in the survey was a tax which each individual and each firm in the Baltic Sea countries would need to pay annually upon the implementation of the environmental improvement program.[14] The

[13] The color scale with precise descriptions as presented to respondents is available in Appendix A. The appendix also presents the maps used for the program illustration.

[14] Pre-testing showed that mentioning firms was important to respondents for the reason of fairness. This formulation, however, could have incentivized respondents to understate their WTP for the program if they

description of the payment mechanism highlighted that the tax would be used for reducing the Baltic Sea eutrophication. Pre-testing showed that the tax vehicle was perceived both credible and acceptable by the interviewed populations.

The payment cards were designed in an analogical way in every country, based on responses from pilot studies. Each card included 18 positive bids,[15] a zero bid. Specific bid amounts differed between the countries and were adjusted to national currencies. The last bid was defined as "over" the last but one bid, which allowed for avoiding truncation of the bid distribution at the upper bound, since this has been evidenced to possibly affect value estimates (Roach, Boyle, and Welsh 2002; Rowe, Schulze, and Breffle 1996).

The survey informed the respondents that by answering the questionnaire, they could affect the environmental policy projects related to controlling eutrophication levels in the Baltic Sea area. Precisely, the survey said that "[respondents'] answers will help governments around the Baltic Sea to develop appropriate water quality improvement programs".

A particular emphasis was placed on designing the questionnaire to be equally relevant and accurate in each Baltic Sea coastal country. Pre-testing included five expert reviews, three focus groups, sixteen cognitive interviews in different countries and pilot surveys in all nine littoral countries. This helped develop an identical survey instrument for every country, which was translated into national languages.

Data collection involved different modes in different countries: Computer-Assisted Web Interviews (CAWI) were administered in Denmark, Estonia, Finland, Germany and Sweden; Computer-Assisted Personal Interviews (CAPI) were conducted in Latvia, Lithuania and Russia; and both CAWI and CAPI were employed in Poland. The choice of a mode in every country was guided mainly by considerations of the costs of the survey administration and of a share of people in a given country with access to the internet. Further details of the survey design and implementation are available in Ahtiainen et al. (2014).

3.2. Deepwater Horizon damage assessment

The 2010 Deepwater Horizon oil spill in the Gulf of Mexico is considered the largest marine oil spill in US waters. To help estimate the monetary value of the natural resources lost in the

---

believed that they would need to pay twice for the improvement—through the firms they worked in and individually.

[15] The only exception is the questionnaire used in Russia, which included 14 positive bids.

aftermath of the oil spill in support of a lawsuit, Bishop et al. (2017) conducted a comprehensive, nationally representative CV study, generating a broad and unique data set.

We use the Deepwater Horizon damage assessment data set to illustrate our proposed procedure of identifying the best-fitting model and deriving WTP estimates on its basis. The data set includes observations for 3,656 households that completed the CV survey. The data was collected through face-to-face interviews in the period from October 2013 to July 2014. The surveyed adult individuals were randomly selected from the population of English-speaking households in the contiguous U.S.

Following an introduction describing the state of the Mexico's Gulf before the Deepwater Horizon oil spill, the cause of the spill and natural resource damages to the Gulf's environment, respondents were asked to evaluate a proposed program of preventing a similar accident in the future. The questionnaire explained that in order to prevent adverse environmental effects of a next oil spill, a second pipe would need to be drilled for every well at the same time the first pipe is constructed. This would enable closing a well quickly in a case of a leak—instead of waiting three months for a relief pipe to be drilled, the closure could be completed in two days. The proposed prevention program assumed that the government would pay to install a second pipe in each of the four-hundred new wells that would be drilled during the next fifteen years in the Mexico's Gulf. Thus, this program could be seen as insurance against the effects of another oil spill. After explaining the program, respondents were asked a single binary-choice CV question: "Do you vote for or against the prevention program, which will cost you and your family living with you the one-time tax of $T?", where T was replaced with a randomly assigned tax amount from the set of $15, $65, $135, $265 and $435. The questionnaire informed that new tax revenue would be needed to implement the proposed program. Each respondent faced a single tax amount. The tax amounts were selected based on optimal design techniques, which included, in particular, a minimum-bias criterion aimed at minimizing the difference between true WTP and the Lewbel-Watanabe estimates.

To examine the sensitivity of WTP responses to the scope of the natural resource damage, two versions of the questionnaire were designed and randomly assigned across the interviewed sample. The questionnaire versions differed with respect to the set of environmental injuries considered—one version presented a smaller set of injuries and another version discussed a larger set of injuries. Specifically, the smaller set described thousands of miles of oiled marshes and beaches, millions of dead birds and one hundred million of lost recreation trips. In addition to these injuries, the larger set included injuries to bottlenose dolphins, deep-water corals, fish, sea turtles, snails and worms.

The survey design process took more than three years and involved many rounds of cognitive interviews, focus groups and pilot studies. Numerous steps were undertaken to assure, among other things, that the questionnaire conveyed needed information for respondents to understand the questions and that the proposed prevention program sounded plausible and effective. For these purposes, for example, before the CV question, respondents were provided with descriptions of the pre-spill condition of the affected resources, the effects of the spill on the resources and the time needed for them to return to the earlier state. To inform respondents that their responses might have actual consequences, prior to the interview each sampled household received a letter on the official U.S. Department of Commerce letterhead discussing the importance of the survey for policy-making.

4. Results

The Baltic Sea and Deepwater Horizon data sets are used to fit parametric distributions of WTP. For the data collected for the evaluation of the Baltic Sea eutrophication reduction program, we consider each country separately and, in addition, we treat separately the CAPI and CAWI data in Poland, which was the only country where the two data collection modes were employed.[16] This results in ten cases for which we aim at identifying the best-fitting WTP distribution.[17] For the Deepwater Horizon damage assessment data, we treat separately the elicited preferences towards avoiding the smaller set and the larger set of injuries. This gives two cases for which we search for the best-fitting distribution of WTP. The detailed results are provided in Appendices B and C for the Baltic Sea and Deepwater Horizon data, respectively, while Tables 1 and 2 summarize the results for the best-fitting parametric distributions.

[16] In the estimation, the data for each country in the Baltic Sea data set is weighted to match the general population of a given country with respect to the shares of females, unemployed individuals and individuals with a university degree. In addition, the weighting for Poland includes the following socio-demographic characteristics: income (four categories), occupational status (four categories), attained education (four categories), household size (six categories), number of children (four categories) and age (four categories).

[17] The questionnaires used national currencies, which differed across the surveyed countries. For the purpose of the comparison, we convert non-EUR currencies, which were in Denmark, Latvia, Lithuania, Poland, Russia and Sweden, into EUR using the PPP corrected exchange rates for 2011 as provided by OECD.Stat (retrieved on June 12, 2017, from https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE4).

For each of the twelve cases, we apply the flexible approach detailed in Section 2 to find the best-fitting parametric distribution. This means that for every case, we consider twenty-two distributions, which, combined with including or not the zero-inflation component, leads to forty-four models examined for each case. The results included in Appendices B and C describe the models that successfully converge and deliver non-extreme parameter estimates.[18] The models listed in the Appendices are ordered according to their fit to the data measured with the Bayesian information criterion (BIC). When models yield the same values of the BIC, they are ordered by the Akaike information criterion (AIC), and in a case when the two criteria are equal, the order is determined based on the log-likelihood values. The BIC is selected as the leading data fit indicator because it implies a more stringent penalty for increasing the number of parameters than the AIC. However, our conclusions are largely unaffected by the choice of the fit measure. For every model in Appendices B and C, we provide the values of the information criteria, the log-likelihood value, the number of the model parameters and the implied estimate of mean WTP together with its standard error.

Table 1 provides a summary of the results for the Baltic Sea data set. We report therein the results for the best-fitting distribution for every country that we obtain with our flexible procedure. The original study by Ahtiainen et al. (2014) estimates WTP values using the "spike model", which is equivalent to our zero-inflated log-normal distribution. In Table 1, we also present replicated results of that model,[19] along with a comparison to the non-parametric Lewbel-Watanabe estimates. While the zero-inflated log-normal distribution selected by Ahtiainen et al. (2014) performs reasonably well (for every country, it is among the first four best-fitting distributions), in all but two cases there are other models that fit the data better. We find that the best fitting distributions for this data set include Birnbaum-Saunders, exponential, generalized Pareto and inverse Gaussian (all with the zero-inflation component).

[18] The estimation is conducted using a custom code developed in Matlab available at https://github.com/czaj/DistFit under CC BY 4.0 license.

[19] We replicate the models using the weights as described in footnote 16. The original study does not report weighting of the observations.

Table 1. Mean annual WTP per person (in 2011 PPP-corrected EUR) for the marine eutrophication reduction in every Baltic Sea country

	Replicated results of Ahtiainen et al. (2014) "spike model" (log-normal distribution with zero inflation)			Results of our flexible approach: The best-fitting distribution for each country			Our results for the lower-bound non-parametric Lewbel-Watanabe estimator
	Log-likelihood	BIC/n	WTP	Log-likelihood	BIC/n	WTP	WTP
Denmark	-2117.76	4.70	38.35*** (3.12)	-2109.09	4.68	37.42*** (2.49)	32.18*** (5.31)
Estonia	-1089.23	4.84	27.94*** (3.03)	-1086.77	4.83	28.41*** (3.16)	23.78*** (4.82)
Finland	-3597.37	4.59	43.40*** (3.04)	-3597.37	4.59	43.40*** (3.04)	36.10*** (7.75)
Germany	-3100.70	4.45	26.16*** (1.85)	-3090.28	4.43	25.94*** (1.82)	22.35*** (4.43)
Latvia	-1331.91	3.98	5.92*** (0.73)	-1315.17	3.92	5.31*** (0.38)	4.49*** (1.33)
Lithuania	-1244.90	4.29	9.03*** (0.86)	-1232.00	4.24	9.22*** (0.89)	8.30*** (3.56)
Poland (CAPI)	-1617.90	3.48	7.72*** (0.98)	-1585.74	3.42	8.06*** (3.48)	6.44*** (2.48)
Poland (CAWI)	-2385.95	5.11	17.98*** (1.29)	-2385.31	5.10	18.92*** (1.05)	16.07*** (1.90)
Russia	-1990.21	2.94	8.54*** (1.55)	-1961.16	2.90	8.66*** (0.89)	6.61*** (2.90)
Sweden	-2304.29	5.00	84.10*** (5.65)	-2304.29	5.00	84.10*** (5.65)	71.72*** (15.84)

Notes: Standard errors are given in brackets. *** denotes WTP statistically different from zero at a 1% significance level. The "spike model" and the best-fitting distributions have three parameters, with the exceptions of the distributions for Latvia and Poland (CAWI) with two parameters and for Poland (CAPI) with four parameters. The BIC is divided by the number of observations (n).

A summary of the results for the Deepwater Horizon data set is provided in Table 2. The table presents the lower-bound non-parametric Lewbel-Watanabe estimates as reported in the original study by Bishop et al. (2017) and the results of our estimation from the best-fitting parametric model and from the Lewbel-Watanabe method. The parametric distribution that appears to fit the data best is the Birnbaum-Saunders model without the zero-inflation component.

Naturally, we warn against generalizing these findings to other CV data sets, as the choice of the best-fitting parametric distribution is strongly data- and preference-elicitation-format-specific.[20] Instead, by presenting the results from these two data sets, we encourage

[20] We find that some of the distributions that are often used with CV data (e.g., Weibull distribution) do not work well with the payment card scales employed in this study—the very large span of bid amounts results here in extreme probability values and numerical errors when evaluating the CDF of these distributions.

considering a wide range of parametric distributions when modelling CV responses, which can indicate the model specification closely matching observed data.

Table 2. Mean WTP per household (in 2013/2014 USD) for preventing future oil-spill-caused injuries to natural resources in the Gulf of Mexico

	Replicated results of Bishop et al. (2017) lower-bound non-parametric Lewbel-Watanabe estimator	Results of our flexible approach: The best-fitting distribution for each set of injuries		
	WTP	Log-likelihood	BIC/n	WTP
Smaller set of injuries	132.36*** (5.38)	-1166.82	1.28	367.52*** (60.93)
Larger set of injuries	152.25*** (5.65)	-1197.21	1.32	528.41*** (86.11)

Notes: Standard errors are given in brackets. *** denotes WTP statistically different from zero at a 1% significance level. The best-fitting distributions have two parameters. The BIC is divided by the number of observations (n).

Turning to the WTP estimates, the values of the Baltic Sea eutrophication reduction derived from the best-fitting distributions do not differ significantly from the estimates obtained from the model specification used in the original study by Ahtiainen et al. (2014).[21] The mean annual WTP values per person range from 5.31 EUR in Latvia to 84.10 EUR in Sweden. The concurrence across the estimates from the original-study specification and our examination is expected, as the log-normal model with the zero-inflation component applied by Ahtiainen et al. (2014) is among the best-fitting models for the wide range of distributions included in our flexible approach. However, differences between the WTP estimates across model specifications appear to substantially increase as a larger spectrum of parametric distributions is considered and, hence, distributions with a worse fit are taken into account (see Appendix B). We return to this issue in a further part of this section.

For the Deepwater Horizon data set, the lower-bound, non-parametric Lewbel-Watanabe estimates of WTP are 132 USD for preventing the smaller set of injuries and about 152 USD for preventing the larger set of injuries. While it is understandable that the conservative, lower-bound estimates may be preferred for applications in legal procedures, it is likely that they do not fully capture the true mean WTP. The mean WTP values derived from

[21] While Ahtiainen et al. (2014) model all data for Poland pooled, we analyze separately the two samples differing by the data collection mode. Observing a significant difference in the value estimates for Poland between CAPI and CAWI respondents (see Table 1), we discuss and examine the mode-related differences in the estimates in more detail in Appendix D.

our best-fitting model are significantly larger.[22] They reach levels of about 368 USD and 528 USD for the smaller and larger, respectively, sets of injuries. Similarly as for the Baltic Sea data, extending the range of parametric models to also consider those with a lower fit to the data seems to increase the variation across the WTP estimates (see Appendix C).

Considering a wide range of parametric distributions can provide insights into potential WTP differences and uncertainty associated with the model selection. To examine the possible extent of the differences resulting from a not-best-fitting model selection, we calculate separately for each country and each set of injuries (i.e., separately for each of the twelve cases to which we apply our flexible modelling approach) the following ratios:

$$\frac{\text{standard deviation of mean WTP estimates from } x \text{ best-fitting distributions for case } i}{\text{mean WTP from the best-fitting model for case } i} \cdot 100\%,$$

where x are selected integers from 2 to the total number of correctly converged models for case i, and cases i include: Denmark, Estonia, Finland, Germany, Latvia, Lithuania, Poland (CAPI), Poland (CAWI), Russia and Sweden for the Baltic Sea data set, and a smaller and a larger set of injuries for the Deepwater Horizon data set. Henceforth, we refer to this ratio as *relative variation*, as it expresses the variation across a selected set of WTP estimates relatively to the WTP obtained from the best-fitting model for a given case. The relative variation for the non-parametric Lewbel-Watanabe estimator is calculated in a similar manner as a ratio of the standard deviation between the Lewbel-Watanabe WTP estimate and the WTP estimate from the best-fitting model divided by the WTP estimate from the best-fitting model.

The calculated ratios of relative variation are presented graphically in Figures 2 and 3 for the Baltic Sea data set, and in Figures 4 and 5 for the Deepwater Horizon data set. Figures 2 and 4 present the relative variation separately for each case (i.e., a country or a set of injuries), while Figures 3 and 5 average the ratios over the respective data set.

[22] We believe that the difference between non-parametric and parametric estimates in the Baltic Sea Action Plan and Deep Water Horizon data arises due to the differences in elicitation formats. While the former used a payment card, effectively bounding each observation from below and above, the binary choice question used in the latter study provides a bound only from below or above.

Figure 2. Relative variation of WTP estimates resulting from $x = \{2, 3, ..., 10, \text{all}\}$ best-fitting distributions for the Baltic Sea data set

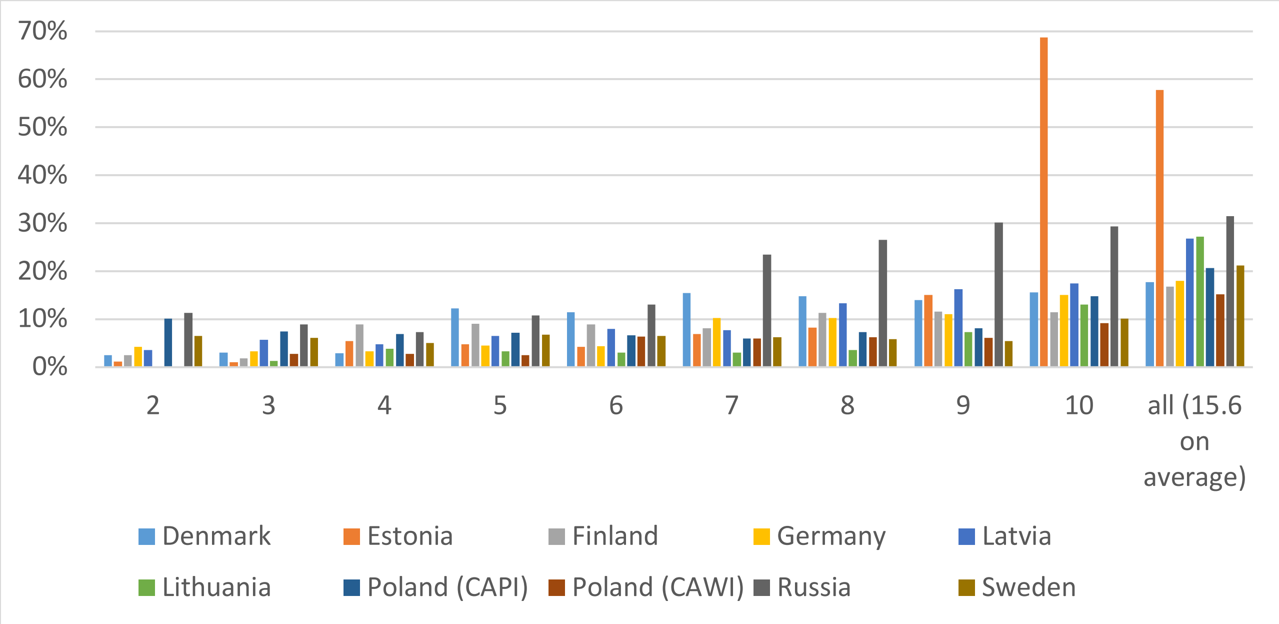
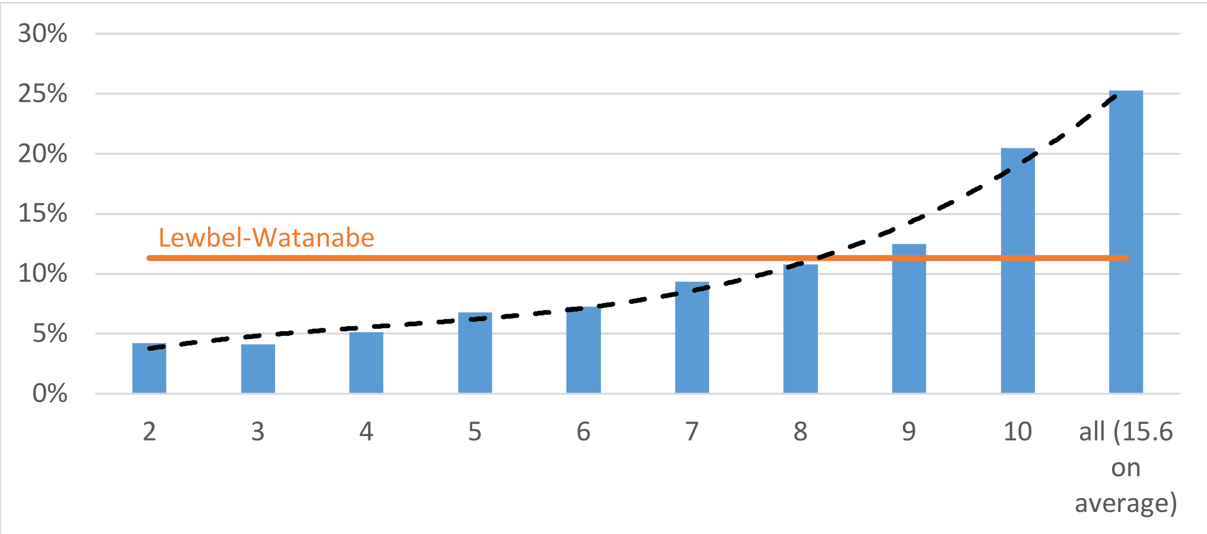


Figure 3. Average relative variation of WTP estimates resulting from $x = \{2, 3, ..., 10, \text{all}\}$ best-fitting distributions for the Baltic Sea data set



Notes: The horizontal orange line represents the relative variation for the non-parametric Lewbel-Watanabe estimator averaged over all cases (i.e., countries) in the data set. The dashed black line results from a fitted polynomial to depict a trend in the relative variation when an increasing number of distributions is considered.

Figure 4. Relative variation of WTP estimates resulting from $x = \{2,3,...,15,\text{all}\}$ best-fitting distributions for the Deepwater Horizon data set

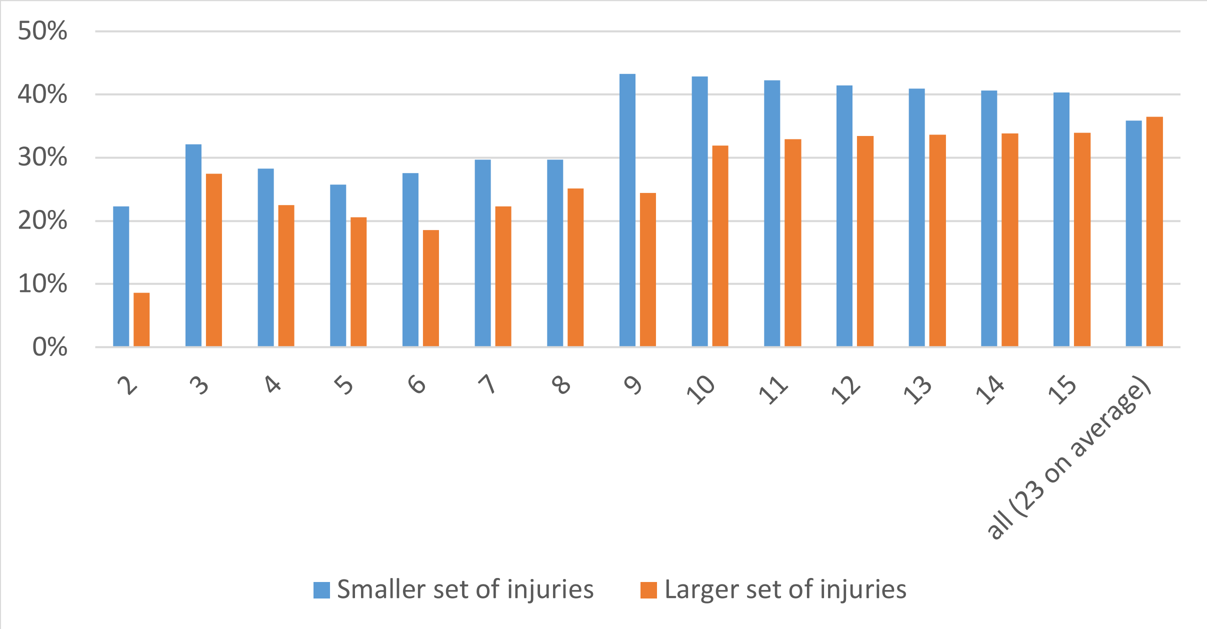
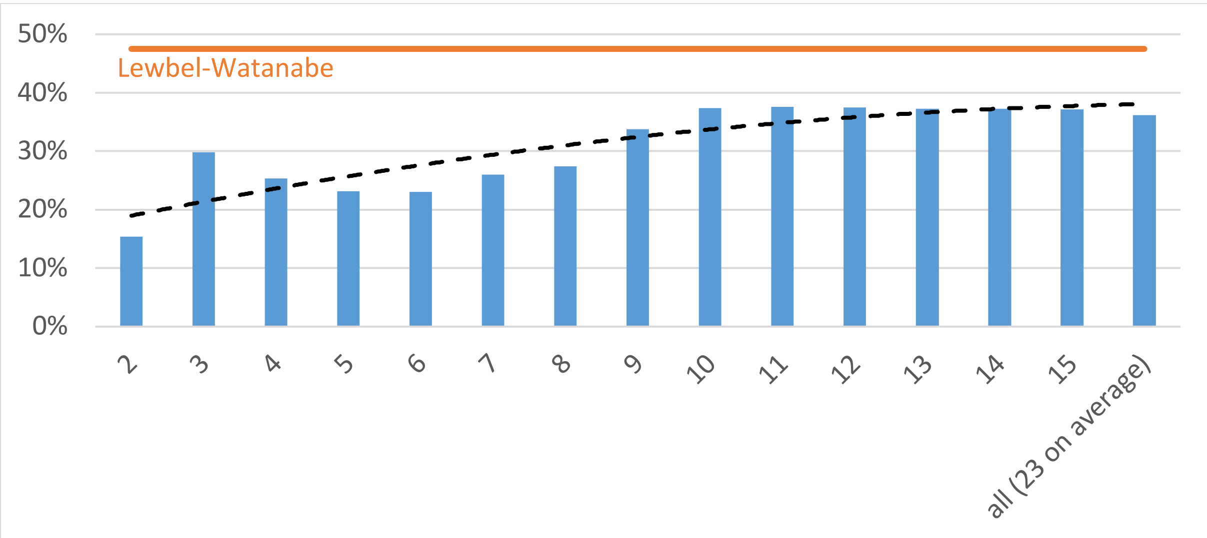


Figure 5. Average relative variation of WTP estimates resulting from $x = \{2,3,...,15,\text{all}\}$ best-fitting distributions for the Deepwater Horizon data set



Notes: The horizontal orange line represents the relative variation for the non-parametric Lewbel-Watanabe estimator averaged over all cases (i.e., sets of injuries) in the data set. The dashed black line results from a fitted polynomial to depict a trend in the relative variation when an increasing number of distributions is considered.

The major finding emerging from Figures 2 to 5 is that, on average, there is a positive trend in the relative variation as the set of considered WTP estimates is expanding. This trend is particularly visible in Figures 3 and 5, where the relative variation ratios averaged over the data sets are illustrated. This relationship appears especially pronounced for the Baltic Sea data set. For the Deepwater Horizon data, we also observe a similar, increasing trend, but the changes in average relative variation when moving from considering x to $x+1$ best-fitting distributions are not always monotonic. Figures 2 and 4 depict the changes in the relative variation in more detail. While we do not observe the relationship between the number of considered WTP estimates and the relative variation to be monotonic, the overall trend is increasing. In brief, these findings suggest that the variation in WTP values is smaller when only better-fitting models are considered. Hence, the selection of models with a good fit to the data appears to be crucial for the reliability of parametric WTP estimates.

5. Conclusions

Stated preference studies deliver important inputs to policy analyses, by providing information on the economic benefits and costs of considered policies in various domains, ranging from environment and health to transportation and culture. Given their widespread use for policy and legal purposes (e.g., Hanley and Czajkowski 2019), it is crucial that the survey-based value estimates provide reliable welfare measures. A factor largely contributing to this reliability is appropriate modelling of stated preference data. Although a considerable share of the literature pays attention to the modelling of attribute-based (i.e., DCE) stated preference data, there is a (arguably) surprising paucity of attention to enhancing modelling approaches for non-attribute (commonly referred as contingent valuation, CV) stated preference data, particularly those relying on parametric modelling. Motivated by this imbalance, in this paper, we approach the question of improving empirical modelling of CV data and propose a practical approach for how the modelling choices by researchers could be guided by observed data. Specifically, we focus on the process of identifying the parametric distribution that matches best the empirical data collected with a CV survey.

The bottom line emerging from our investigation is simple: the reliability of CV-based WTP estimates can be enhanced by considering many parametric distributions in modelling the CV data to select the one that fits the data best. Based on two flagship databases for CV studies, we observe non-negligible differences in value estimates across models differing in assumed parametric distributions. These findings indicate that choosing a model specification ad hoc may reduce the model fit to the data and lower the precision of value estimates. For the two considered databases, we observe that the variation in value estimates is smaller when only better-fitting models are considered. This further emphasizes the need for caution when choosing a parametric specification for modelling CV data and for identifying the best-fitting model in order to derive reliable value estimates.

Improving estimation methods help deliver more precise and, thus, more reliable value estimates, which in turn can generate more economically efficient policy decisions. We believe there is a substantial need for advancing CV data modelling approaches, especially as they have experienced little development over the last decades, which stays in a large contrast to advances to modelling valuation data from other sources, such as discrete choice experiments (Johnston et al. 2017). The importance of this need gets even more pronounced when considering that a single binary choice—a CV preference elicitation format—constitutes a gold standard in stated preference elicitation approaches. The use of this format has been recommended since the NOAA panel report (Arrow et al. 1993), as it remains the most straightforward approach for incentivizing truthful preference disclosure by eliminating incentives for strategic responses. We believe that the findings from this paper combined with the proposed practical approach for improving the reliability of value estimates derived from CV data provide an important step towards advancing CV modelling procedures.

References

Ahlvik, L., Ekholm, P., Hyytiäinen, K., & Pitkänen, H. (2014). An economic–ecological model to evaluate impacts of nutrient abatement in the Baltic Sea. *Environmental Modelling & Software, 55*(0), 164-175. doi:http://dx.doi.org/10.1016/j.envsoft.2014.01.027

Ahtiainen, H., Artell, J., Czajkowski, M., Hasler, B., Hasselström, L., Huhtala, A., . . . Semeniene, D. (2014). Benefits of meeting nutrient reduction targets for the Baltic Sea – a contingent valuation study in the nine coastal states. *Journal of Environmental Economics and Policy, 3*(3), 1-28. doi:10.1080/21606544.2014.901923

Alberini, A. (1995). Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models. *Journal of Environmental Economics and Management, 29*(2), 169-180.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. *Federal Register, 58*, 4601-4614.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An Empirical Distribution Function for Sampling with Incomplete Information. *Ann. Math. Statist., 26*(4), 641-647. doi:10.1214/aoms/1177728423

Balderas Torres, A., MacMillan, D. C., Skutsch, M., & Lovett, J. C. (2015). 'Yes-in-my-backyard': Spatial differences in the valuation of forest services and local co-benefits for carbon markets in México. *Ecological Economics, 109*(0), 130-141. doi:http://dx.doi.org/10.1016/j.ecolecon.2014.11.008

Bateman, I. J., Carson, R. T., Day, B., Hanemann, M. W., Hanley, N., Hett, T., . . . Swanson, J. (2004). *Economic Valuation with Stated Preference Techniques: A Manual*. Northampton, MA: Edward Elgar.

Bell, J., Huber, J., & Viscusi, W. K. (2011). Survey Mode Effects on Valuation of Environmental Goods. *International Journal of Environmental Research and Public Health, 8*(4), 1222. Retrieved from http://www.mdpi.com/1660-4601/8/4/1222

Bishop, R. C., & Boyle, K. J. (2019). Reliability and Validity in Nonmarket Valuation. *Environmental and Resource Economics, 72*(2), 559-582. doi:10.1007/s10640-017-0215-7

Bishop, R. C., Boyle, K. J., Carson, R. T., Chapman, D., Hanemann, W. M., Kanninen, B., . . . Scherer, N. (2017). Putting a value on injuries to natural assets: The BP oil spill. *Science, 356*(6335), 253-254. doi:10.1126/science.aam8124

Cameron, T. A. (1988). A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression. *Journal of Economics and Management, 15*(3), 355-379.

Cameron, T. A., & Huppert, D. D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management, 17*(3), 230-246.

Canavari, M., Nocella, G., & Scarpa, R. (2005). Stated Willingness-to-Pay for Organic Fruit and Pesticide Ban. *Journal of Food Products Marketing, 11*(3), 107-134. doi:10.1300/J038v11n03_07

Cardamone, A., Eboli, L., & Mazzulla, G. (2014). Drivers' road accident risk perception. A comparison between face-to-face interview and web-based survey. *Advances in Transportation Studies*(33), 59-72.

Carson, R. T., & Czajkowski, M. (2014). The Discrete Choice Experiment Approach to Environmental Contingent Valuation. In S. Hess & A. Daly (Eds.), *Handbook of choice modelling*. Northampton, MA: Edward Elgar.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics, 37*(1), 181-210. Retrieved from http://dx.doi.org/10.1007/s10640-007-9124-5

Carson, R. T., Groves, T., & List, J. A. (2014). Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. *Journal of the Association of Environmental and Resource Economists, 1*(1/2), 171-207. Retrieved from http://www.jstor.org/stable/10.1086/676450

Champ, P. A., Boyle, K. J., & Brown, T. C. (2017). *A Primer on Nonmarket Valuation*. Amsterdam: Springer.

Covey, J., Robinson, A., Jones-Lee, M., & Loomes, G. (2010). Responsibility, scale and the valuation of rail safety. *Journal of Risk and Uncertainty, 40*(1), 85-108. doi:10.1007/s11166-009-9082-0

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4 ed.): Wiley.

Freeman, A. M., Herriges, J. A., & Kling, C. L. (2014). *The measurement of environmental and resource values: Theory and methods*: Routledge.

Fricker, R. D., & Schonlau, M. (2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods, 14*(4), 347-367. doi:10.1177/152582202237725

Goethals, F., Leclercq-Vandelannoitte, A., & Tütüncü, Y. (2012). French consumers' perceptions of the unattended delivery model for e-grocery retailing. *Journal of Retailing and Consumer Services, 19*(1), 133-139. doi:https://doi.org/10.1016/j.jretconser.2011.11.002

Goldenbeld, C., & de Craen, S. (2013). The comparison of road safety survey answers between web-panel and face-to-face; Dutch results of SARTRE-4 survey. *Journal of Safety Research, 46*(Supplement C), 13-20. doi:https://doi.org/10.1016/j.jsr.2013.03.004

Greene, W. H. (2011). *Econometric Analysis* (7 ed.). Upper Saddle River, NJ: Prentice Hall.

Gurmu, S., & Trivedi, P. K. (1996). Excess Zeros in Count Models for Recreational Trips. *Journal of Business & Economic Statistics, 14*(4), 469-477. doi:10.2307/1392255

Haab, T., & McConnell, K. (2003). *The Econometrics Of Non-Market Valuation*. Northampton, MA: Edward Elgar.

Hamzaoui-Essoussi, L., & Zahaf, M. (2012). Canadian Organic Food Consumers' Profile and Their Willingness to Pay Premium Prices. *Journal of International Food & Agribusiness Marketing, 24*(1), 1-21. doi:10.1080/08974438.2011.621834

Hanemann, W. M. (1984). Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of of Agricultural Economics, 71*, 1057-1061.

Hanley, N., & Czajkowski, M. (2019). The Role of Stated Preference Valuation Methods in Understanding Choices and Informing Policy. *Review of Environmental Economics and Policy, 13*(2), 248-266. doi:10.1093/reep/rez005

HELCOM. (2013). *Taking Further Action to Implement the Baltic Sea Action Plan - Reaching Good Environmental Status for a healthy Baltic Sea*. Retrieved from Copenhagen, Denmark:

HELCOM. (2014). Eutrophication status of the Baltic Sea 2007-2011 - A concise thematic assessment. *Baltic Sea Environment Proceedings, 143*.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International statistical review, 78*(1), 3-20. doi:10.1111/j.1751-5823.2010.00102.x

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., . . . Vossler, C. A. (2017). Contemporary Guidance for Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists, 4*(2), 319-405.

Kiirikki, M., Inkala, A., Kuosa, H., Pitkänen, H., Kuusisto, M., & Sarkkula, J. (2001). Evaluating the effects of nutrient load reductions on the biomass of toxic nitrogenfixing cyanobacteria in the Gulf of Finland, Baltic Sea. *Boreal Environment Research, 6*, 1-16.

Kiirikki, M., Lehtoranta, J., Inkala, A., Pitkänen, H., Hietanen, S., Hall, P. O. J., . . . Sarkkula, J. (2006). A simple sediment process description suitable for 3D-ecosystem modelling — Development and testing in the Gulf of Finland. *Journal of Marine Systems, 61*(1–2), 55-66. doi:http://dx.doi.org/10.1016/j.jmarsys.2006.02.008

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics, 68*(4), 715-719. Retrieved from http://www.jstor.org/stable/1924536

Krinsky, I., & Robb, A. L. (1991). Three methods for calculating the statistical properties of elasticities: A comparison. *Empirical Economics, 16*(2), 199-209. Retrieved from http://dx.doi.org/10.1007/BF01193491

Kriström, B. (1997). Spike Models in Contingent Valuation. *American Journal of Agricultural Economics, 79*(3), 1013-1023.

Lee, C.-K., Kim, T.-K., & Mjelde, J. W. (2016). Comparison of preservation values between Internet and interview survey modes: the case of Dokdo, South Korea. *Journal of Environmental Planning and Management, 59*(1), 22-43. doi:10.1080/09640568.2014.980900

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics, 97*(1), 145-177. doi:https://doi.org/10.1016/S0304-4076(00)00015-4

Lindhjem, H., & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics, 70*(9), 1628-1637. doi:http://dx.doi.org/10.1016/j.ecolecon.2011.04.002

Maar, M., Møller, E. F., Larsen, J., Madsen, K. S., Wan, Z., She, J., . . . Neumann, T. (2011). Ecosystem modelling across a salinity gradient from the North Sea to the Baltic Sea. *Ecological Modelling, 222*(10), 1696-1711. doi:http://dx.doi.org/10.1016/j.ecolmodel.2011.03.006

Maier, E., Wilken, R., & Dost, F. (2015). The double benefits of consumer certainty: combining risk and range effects. *Marketing Letters, 26*(4), 473-488. doi:10.1007/s11002-014-9282-5

Manski, C. F. (2017). Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment. *Theory and Decision, 83*(2), 155-173. doi:10.1007/s11238-017-9592-1

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., . . . Thiene, M. (2020). *Environmental Valuation with Discrete Choice Experiments. Guidance on Design, Implementation and Data Analysis*: Springer.

Markowska, A., & Żylicz, T. (1999). Costing an International Public Good: The Case of the Baltic Sea. *Ecological Economics, 30*, 301-316.

Marta-Pedroso, C., Freitas, H., & Domingos, T. (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics, 62*(3), 388-398. doi:https://doi.org/10.1016/j.ecolecon.2007.02.005

McFadden, D. (1974). Conditional Logit Analysis of Qualititative Choice Behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York, NY: Academic Press.

Menegaki, A. N., Olsen, S. B., & Tsagarakis, K. P. (2016). Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *Journal of Choice Modelling, 18*(Supplement C), 18-50. doi:https://doi.org/10.1016/j.jocm.2016.04.005

Mitchell, R. C., & Carson, R. T. (1989). *Using Surveys to Value Public Goods: the Contingent Valuation Method*. Washington, DC: Resources for the Future.

Mjelde, J. W., Kim, T.-K., & Lee, C.-K. (2016). Comparison of Internet and interview survey modes when estimating willingness to pay using choice experiments. *Applied Economics Letters, 23*(1), 74-77. doi:10.1080/13504851.2015.1051648

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary Choice Health State Valuation and Mode of Administration: Head-to-Head Comparison of Online and CAPI. *Value in Health, 16*(1), 104-113. doi:https://doi.org/10.1016/j.jval.2012.09.001

Nielsen, J. S. (2011). Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. *Resource and Energy Economics, 33*(1), 119-129. doi:https://doi.org/10.1016/j.reseneeco.2010.01.006

Pakalniete, K., Aigars, J., Czajkowski, M., Strake, S., Zawojska, E., & Hanley, N. (2017). Understanding the distribution of economic benefits from improving coastal and marine ecosystems. *Science of The Total Environment, 584-585*, 29-40.

Ready, R., Fisher, A., Guignet, D., Stedman, R., & Wang, J. (2006). A pilot test of a new stated preference valuation method: Continuous attribute-based stated choice. *Ecological Economics, 59*(3), 247-255. doi:https://doi.org/10.1016/j.ecolecon.2005.11.004

Reichl, J., Schmidthaler, M., & Schneider, F. (2013). The value of supply security: The costs of power outages to Austrian households, firms and the public sector. *Energy Economics, 36*(Supplement C), 256-261. doi:https://doi.org/10.1016/j.eneco.2012.08.044

Reusch, T. B. H., Dierking, J., Andersson, H. C., Bonsdorff, E., Carstensen, J., Casini, M., . . . Zandersen, M. (2018). The Baltic Sea as a time machine for the future coastal ocean. *Science Advances, 4*(5). doi:10.1126/sciadv.aar8195

Roach, B., Boyle, K. J., & Welsh, M. P. (2002). Testing Bid Design Effects in Multiple Bounded Contingent Valuation. *Land Economics, 78*(1), 121-131.

Rowe, R. D., Schulze, W. D., & Breffle, W. S. (1996). A Test for Payment Card Biases. *Journal of Environmental Economics and Management, 31*(2), 178-185.

Sandorf, E. D., Aanesen, M., & Navrud, S. (2016). Valuing unfamiliar and complex environmental goods: A comparison of valuation workshops and internet panel surveys with videos. *Ecological Economics, 129*(Supplement C), 50-61. doi:https://doi.org/10.1016/j.ecolecon.2016.06.008

Scarpa, R., Thiene, M., & Train, K. (2008). Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps. *American Journal of Agricultural Economics, 90*(4), 994-1010. doi:10.1111/j.1467-8276.2008.01155.x

Ščasný, M., & Alberini, A. (2012). Valuation of Mortality Risk Attributable to Climate Change: Investigating the Effect of Survey Administration Modes on a VSL. *International Journal of Environmental Research and Public Health, 9*(12), 4760. Retrieved from http://www.mdpi.com/1660-4601/9/12/4760

Söderqvist, T., & Scharin, H. (2000). *The regional willingness to pay for a reduced eutrophication in the Stockholm archipelago*. Retrieved from Beijer Discussion Paper Series No. 128:

Stephenson, L. B., & Crête, J. (2011). Studying Political Behavior: A Comparison of Internet and Telephone Surveys. *International Journal of Public Opinion Research, 23*(1), 24-55. doi:10.1093/ijpor/edq025

Tuhkanen, H., Piirsalu, E., Nõmmann, T., Karlõševa, A., Nõmmann, S., Czajkowski, M., & Hanley, N. (2016). Valuing the benefits of improved marine environmental quality under multiple stressors. *Science of The Total Environment, 551-552*, 367-375. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048969716302224

Turnbull, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society. Series B (Methodological), 38*(3), 290-295. Retrieved from http://www.jstor.org/stable/2984980

van der Heide, C. M., van den Bergh, J. C. J. M., van Ierland, E. C., & Nunes, P. A. L. D. (2008). Economic valuation of habitat defragmentation: A study of the Veluwe, the Netherlands. *Ecological Economics, 67*(2), 205-216. doi:https://doi.org/10.1016/j.ecolecon.2008.04.012

Vossler, C. A., & Holladay, J. S. (2018). Alternative value elicitation formats in contingent valuation: Mechanism design and convergent validity. *Journal of Public Economics, 165*, 133-145. doi:https://doi.org/10.1016/j.jpubeco.2018.07.004

Watanabe, M. (2010). Nonparametric Estimation of Mean Willingness to Pay from Discrete Response Valuation Data. *American Journal of Agricultural Economics, 92*(4), 1114-1135. doi:10.1093/ajae/aaq034

Appendix A. Description of eutrophication-related water quality levels used in the survey

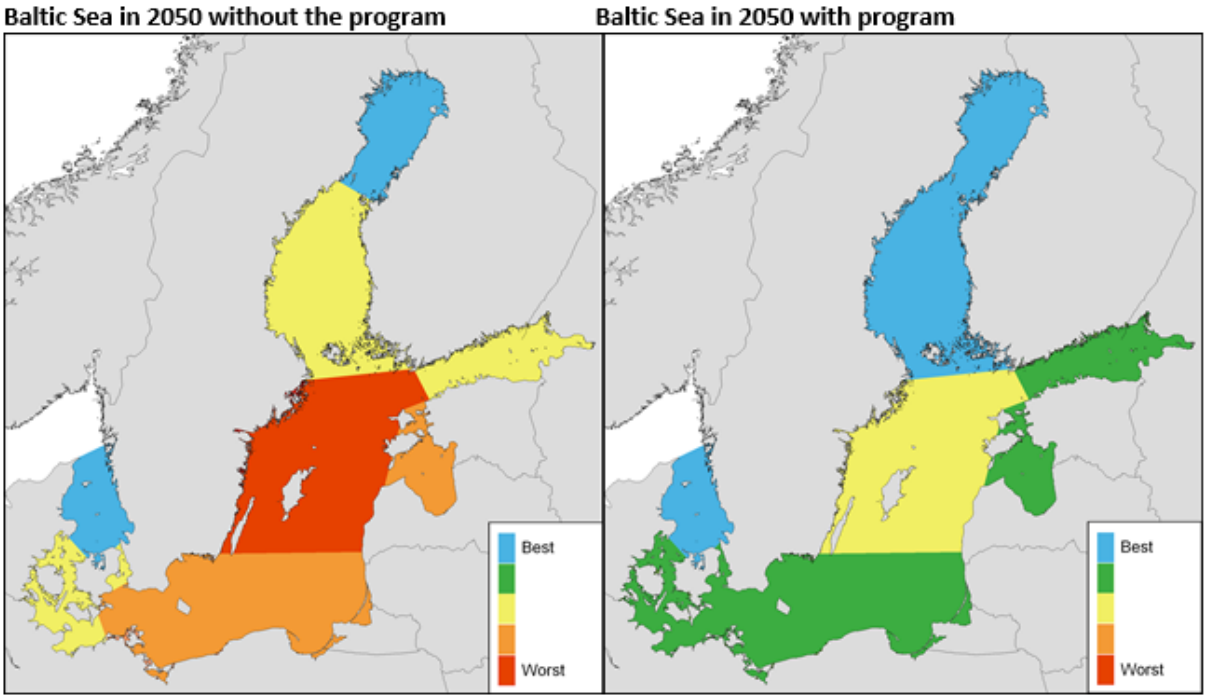
Figure A1. Colored water quality scale describing intensity levels of the considered ecosystem effects

The effects of eutrophication on water quality in open sea areas

Marine scientists have prepared a colour scale to show how serious eutrophication is in the different parts of open Baltic Sea. Before answering to the following questions, we would like you to familiarise with the colour scale below.

Water quality	Description of the effects of eutrophication					**Water quality**
	Water clarity	*Blue-green algal blooms*	*Underwater meadows*	*Fish species*	*Deep sea bottoms*	
Best possible water quality	Clear	Seldom	Excellent condition Good for fish spawning and feeding	Cod, herring and perch common	No oxygen deficiency Bottom animals common	**Best possible water quality**
	Mainly clear	Sometimes	Patchy vegetation Good for fish spawning and feeding	Cod, herring and perch common	Oxygen deficiency in large areas Bottom animals common	
	Slightly turbid	In most summers	Cover a small area Less good for fish spawning	Fewer cod, but herring and perch common More roach, carp and bream	Oxygen shortages often in large areas Some bottom animals rare	
	Turbid	Every summer	Cover a small area Bad for fish spawning	Fewer cod, herring and perch More roach, carp and bream	Oxygen shortages often in large areas Some bottom animal groups have disappeared	
Worst possible water quality	Very turbid	On large areas every summer	Almost gone Not suitable for fish spawning	Almost no cod, fewer herring and perch Lots of roach, carp and bream	Oxygen shortages always in large areas No bottom animals in many areas	**Worst possible water quality**

Figure A2. A map illustrating changes in the open-sea water eutrophication as a result of implementing the nutrient loadings reduction program



Notes: The colors follow a detailed scale defined in the survey, as presented in Figure A1.

Appendix B. Comparison of different parametric distributions fitted to the payment card data for each Baltic Sea country

Denmark							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						32.18	5.31
Birnbaum-Saunders	yes	-2109.09	3	4.67	4.68	37.42	2.49
Inverse Gaussian	yes	-2116.43	3	4.68	4.70	36.12	2.86
Log-normal	yes	-2117.76	3	4.69	4.70	38.35	3.12
Exponential	yes	-2135.42	2	4.72	4.73	36.17	2.09
Log-logistic	yes	-2132.74	3	4.72	4.74	47.06	9.95
Generalized Pareto	yes	-2129.78	4	4.72	4.74	35.79	2.56
Generalized extreme value	yes	-2135.71	4	4.73	4.75	49.62	21.50
Negative binomial	yes	-2152.70	3	4.76	4.78	35.74	2.25
Negative binomial	no	-2276.25	2	5.03	5.05	37.40	3.64
t location-scale	yes	-2299.93	4	5.09	5.11	29.58	2.16
Logistic	yes	-2347.13	3	5.19	5.21	34.12	1.65
Normal	yes	-2431.45	3	5.38	5.40	41.69	2.06
Rayleigh	yes	-2479.44	2	5.48	5.49	47.21	1.45
Rician	yes	-2479.44	3	5.49	5.50	44.35	1.29
Extreme value	yes	-2688.91	3	5.95	5.96	47.02	2.80
Exponential	no	-2830.65	1	6.26	6.26	36.69	1.65
Logistic	no	-3472.35	2	7.68	7.69	34.00	1.70
Normal	no	-3630.06	2	8.03	8.04	45.44	2.10
Extreme value	no	-4092.27	2	9.05	9.06	55.17	3.54

Notes: *n* (the number of observations) = 905. "No. of param." stays for the number of parameters. WTP is in 2011 EUR.

Estonia							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						23.78	4.82
Inverse Gaussian	yes	-1086.77	3	4.80	4.83	28.41	3.16
Log-normal	yes	-1089.23	3	4.81	4.84	27.94	3.03
Birnbaum-Saunders	yes	-1090.95	3	4.82	4.85	27.82	2.49
Log-logistic	yes	-1092.14	3	4.82	4.85	31.08	10.98
Generalized Pareto	yes	-1095.53	4	4.84	4.88	28.11	3.92
Exponential	yes	-1115.65	2	4.92	4.94	28.48	1.99
Logistic	yes	-1266.13	3	5.59	5.62	24.33	1.56
Normal	yes	-1326.73	3	5.86	5.89	32.11	2.14
Rayleigh	yes	-1390.74	2	6.14	6.15	39.60	1.48
Uniform	yes	-1389.64	3	6.13	6.16	90.18	2.22
Exponential	no	-1396.85	1	6.16	6.17	26.91	1.55
Extreme value	yes	-1446.76	3	6.39	6.41	36.97	2.67
Logistic	no	-1742.77	2	7.69	7.70	23.49	1.45
Normal	no	-1862.60	2	8.21	8.23	35.61	2.16
Extreme value	no	-2081.92	2	9.18	9.20	46.19	3.00

Notes: *n* (the number of observations) = 454. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Finland							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						36.10	7.75
Log-normal	yes	-3597.37	3	4.58	4.59	43.40	3.04
Birnbaum-Saunders	yes	-3599.56	3	4.59	4.60	41.83	2.42
Inverse Gaussian	yes	-3608.78	3	4.60	4.61	42.38	2.96
Log-logistic	yes	-3614.57	3	4.61	4.62	50.21	19.85
Generalized extreme value	yes	-3623.19	4	4.62	4.63	48.89	10.09
Generalized Pareto	yes	-3627.39	4	4.62	4.64	41.32	2.50
Exponential	yes	-3644.93	2	4.64	4.65	43.87	2.17
t location-scale	yes	-3863.29	4	4.92	4.94	34.10	2.30
Logistic	yes	-4022.11	3	5.12	5.13	37.28	1.63
Normal	yes	-4292.74	3	5.47	5.48	47.53	2.07
Rayleigh	yes	-4383.18	2	5.58	5.59	56.09	1.15
Rician	yes	-4381.45	3	5.58	5.59	56.39	1.17
Exponential	no	-4795.90	1	6.11	6.11	42.09	1.75
t location-scale	no	-5584.53	3	7.11	7.12	31.30	12.10
Logistic	no	-5850.75	2	7.45	7.46	38.46	1.60
Normal	no	-6267.34	2	7.98	7.99	52.61	2.09

Notes: *n* (the number of observations) = 1,571. "No. of param." stays for the number of parameters. WTP is in 2011 EUR.

Germany							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						22.35	4.43
Inverse Gaussian	yes	-3090.28	3	4.42	4.43	25.94	1.82
Birnbaum-Saunders	yes	-3095.65	3	4.43	4.44	27.52	1.63
Log-normal	yes	-3100.70	3	4.44	4.45	26.16	1.85
Log-logistic	yes	-3118.52	3	4.46	4.47	27.57	3.86
Generalized Pareto	yes	-3139.05	4	4.49	4.51	24.79	1.64
Exponential	yes	-3155.12	2	4.51	4.52	25.38	1.34
t location-scale	yes	-3356.39	4	4.80	4.82	19.76	3.14
Logistic	yes	-3480.59	3	4.98	4.99	22.31	1.04
Normal	yes	-3662.08	3	5.24	5.25	29.05	1.36
Rayleigh	yes	-3732.14	2	5.34	5.35	34.34	0.96
Rician	yes	-3732.14	3	5.34	5.35	35.97	0.97
Exponential	no	-4223.77	1	6.04	6.04	25.23	1.04
Logistic	no	-5267.50	2	7.53	7.54	22.92	1.05
Normal	no	-5629.61	2	8.05	8.06	33.29	1.38

Notes: n (the number of observations) = 1,399. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Latvia							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						4.49	1.33
Exponential	yes	-1315.17	2	3.91	3.92	5.31	0.38
Birnbaum-Saunders	yes	-1324.89	3	3.94	3.96	5.58	0.53
Log-normal	yes	-1331.91	3	3.96	3.98	5.92	0.73
Inverse Gaussian	yes	-1335.94	3	3.97	3.99	5.64	0.71
Logistic	yes	-1496.16	3	4.45	4.47	5.00	0.31
Normal	yes	-1544.77	3	4.59	4.61	6.20	0.38
Exponential	no	-1569.69	1	4.66	4.67	5.29	0.27
Rayleigh	yes	-1631.27	2	4.85	4.86	7.26	0.26
Rician	yes	-1631.27	3	4.85	4.87	7.42	0.27
Extreme value	yes	-1694.65	3	5.04	5.06	7.38	0.54
Logistic	no	-2155.43	2	6.40	6.42	4.72	0.28
Normal	no	-2309.65	2	6.86	6.87	7.32	0.40
Extreme value	no	-2677.11	2	7.95	7.96	9.86	0.53

Notes: *n* (the number of observations) = 674. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Lithuania							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						8.30	3.56
Inverse Gaussian	yes	-1232.00	3	4.22	4.24	9.22	0.89
Birnbaum-Saunders	yes	-1238.62	3	4.24	4.27	9.24	0.74
Log-normal	yes	-1244.90	3	4.27	4.29	9.03	0.86
Log-logistic	yes	-1252.87	3	4.29	4.32	9.84	2.60
Exponential	yes	-1263.74	2	4.33	4.34	9.25	0.66
Negative binomial	yes	-1285.12	3	4.40	4.43	9.36	0.81
Negative binomial	no	-1313.82	2	4.50	4.51	9.65	1.17
Logistic	yes	-1427.48	3	4.89	4.91	8.79	0.60
Normal	yes	-1472.78	3	5.05	5.07	11.09	0.72
Rayleigh	yes	-1524.31	2	5.22	5.23	12.74	0.46
Exponential	no	-1576.37	1	5.39	5.40	9.34	0.50
Extreme value	yes	-1626.33	3	5.57	5.59	12.72	1.01
Logistic	no	-2069.07	2	7.08	7.10	8.12	0.50
Normal	no	-2211.37	2	7.57	7.58	12.72	0.69
Extreme value	no	-2558.62	2	8.75	8.77	17.83	1.12

Notes: *n* (the number of observations) = 585. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Poland (CAPI)							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						6.44	2.48
Generalized Pareto	yes	-1585.74	4	3.40	3.42	8.06	3.48
Exponential	yes	-1618.23	2	3.47	3.48	6.90	0.50
Log-normal	yes	-1617.90	3	3.47	3.48	7.72	0.98
Inverse Gaussian	yes	-1620.51	3	3.47	3.49	8.10	0.94
Birnbaum-Saunders	yes	-1620.55	3	3.47	3.49	6.98	0.64
Negative binomial	yes	-1621.12	3	3.47	3.49	7.24	0.78
Negative binomial	no	-1636.34	2	3.50	3.51	7.43	0.97
Logistic	yes	-1823.08	3	3.91	3.92	6.44	0.42
Normal	yes	-1903.44	3	4.08	4.09	8.42	0.53
Rayleigh	yes	-2010.28	2	4.30	4.31	10.72	0.39
Rician	yes	-2010.28	3	4.31	4.32	10.60	0.40
Extreme value	yes	-2096.54	3	4.49	4.51	10.69	0.81
Exponential	no	-2384.61	1	5.10	5.11	7.18	0.32
Logistic	no	-3280.28	2	7.02	7.03	6.27	0.33
Normal	no	-3606.96	2	7.72	7.73	10.96	0.55

Notes: *n* (the number of observations) = 935. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Poland (CAWI)							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						16.07	1.90
Exponential	yes	-2385.31	2	5.09	5.10	18.92	1.05
Generalized Pareto	yes	-2379.19	4	5.08	5.10	18.94	1.23
Birnbaum-Saunders	yes	-2385.64	3	5.09	5.11	18.03	1.07
Log-normal	yes	-2385.95	3	5.09	5.11	17.98	1.29
Inverse Gaussian	yes	-2391.31	3	5.11	5.12	18.32	1.33
Log-logistic	yes	-2393.07	3	5.11	5.12	21.20	4.03
Negative binomial	yes	-2396.16	3	5.12	5.13	18.16	1.06
Generalized extreme value	yes	-2393.02	4	5.11	5.13	20.40	2.75
Negative binomial	no	-2462.02	2	5.25	5.26	18.11	1.45
t location-scale	yes	-2547.90	4	5.44	5.46	14.61	3.06
Logistic	yes	-2637.57	3	5.63	5.65	16.79	0.80
Normal	yes	-2720.83	3	5.81	5.82	19.95	0.89
Exponential	no	-2740.43	1	5.85	5.85	18.04	0.81
Rayleigh	yes	-2766.54	2	5.90	5.91	22.31	0.63
Rician	yes	-2766.54	3	5.91	5.92	22.89	0.79
Extreme value	yes	-2962.49	3	6.32	6.34	21.99	1.15
t location-scale	no	-3199.83	3	6.83	6.84	13.55	2.72
Logistic	no	-3332.17	2	7.11	7.12	16.45	0.73
Normal	no	-3470.45	2	7.40	7.41	21.42	0.95
Extreme value	no	-3848.45	2	8.21	8.22	25.52	1.20

Notes: *n* (the number of observations) = 938. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Russia							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						6.61	2.90
Birnbaum-Saunders	yes	-1961.16	3	2.88	2.90	8.66	0.89
Inverse Gaussian	yes	-1966.77	3	2.89	2.90	7.27	1.20
Log-normal	yes	-1990.21	3	2.93	2.94	8.54	1.55
Exponential	yes	-2039.18	2	3.00	3.00	8.01	0.57
Logistic	yes	-2326.87	3	3.42	3.43	6.43	0.47
Normal	yes	-2417.95	3	3.55	3.57	9.63	0.63
Rayleigh	yes	-2644.22	2	3.89	3.89	12.72	0.47
Extreme value	yes	-2641.36	3	3.88	3.89	12.48	1.02
Rician	yes	-2644.22	3	3.89	3.90	13.66	1.04
Exponential	no	-3462.47	1	5.09	5.09	7.75	0.32
Logistic	no	-4902.87	2	7.20	7.21	6.60	0.37
Normal	no	-5471.24	2	8.04	8.04	13.46	1.27

Notes: *n* (the number of observations) = 1,362. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Sweden							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						71.72	15.84
Log-normal	yes	-2304.29	3	4.99	5.00	84.10	5.65
Log-logistic	yes	-2310.48	3	5.00	5.02	91.81	20.75
Inverse Gaussian	yes	-2311.86	3	5.01	5.02	82.06	5.73
Birnbaum-Saunders	yes	-2314.54	3	5.01	5.03	85.35	5.08
Generalized extreme value	yes	-2316.80	4	5.02	5.04	95.68	26.66
Generalized Pareto	yes	-2337.34	4	5.06	5.08	82.84	5.32
Exponential	yes	-2364.73	2	5.12	5.13	82.88	4.14
Negative binomial	yes	-2369.35	3	5.13	5.15	84.53	4.27
Negative binomial	no	-2485.52	2	5.38	5.39	85.52	5.93
t location-scale	yes	-2492.23	4	5.40	5.42	62.96	9.81
Logistic	yes	-2684.49	3	5.81	5.83	72.79	3.40
Exponential	no	-2723.37	1	5.89	5.90	84.17	3.96
Normal	yes	-2906.63	3	6.29	6.31	94.53	4.94
Rayleigh	yes	-3029.65	2	6.55	6.57	124.22	3.33
t location-scale	no	-3048.08	3	6.60	6.61	62.44	8.35
Logistic	no	-3255.88	2	7.04	7.05	73.64	3.22
Normal	no	-3560.42	2	7.70	7.71	130.53	5.52

Notes: n (the number of observations) = 925. “No. of param.” stays for the number of parameters. WTP is in 2011 EUR.

Appendix C. Comparison of different parametric distributions fitted to the binary choice data for the Deepwater Horizon damage assessment

Smaller set of injuries							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						132.36	5.38
BirnbaumSaunders	no	-1166.82	2	1.28	1.28	362.41	61.85
Gamma	no	-1167.95	2	1.28	1.28	479.12	99.64
Exponential	yes	-1168.36	2	1.28	1.28	256.80	32.18
Negative Binomial	no	-1169.89	2	1.28	1.28	371.35	119.20
Gamma	yes	-1166.72	3	1.28	1.29	284.44	103.90
Lognormal	yes	-1168.89	3	1.28	1.29	406.17	107.73
Negative Binomial	yes	-1169.89	3	1.28	1.29	232.00	85.95
Uniform	yes	-1170.64	3	1.28	1.29	202.32	17.99
Rayleigh	yes	-1176.03	2	1.29	1.29	200.65	15.57
Weibull	no	-1176.89	2	1.29	1.29	682.46	178.17
Poisson	yes	-1180.30	2	1.29	1.30	200.10	13.86
Normal	yes	-1184.01	3	1.30	1.30	187.19	10.03
Logistic	yes	-1186.92	3	1.30	1.31	185.73	8.02
Extreme Value	yes	-1187.57	3	1.30	1.31	181.64	6.23
Exponential	no	-1462.80	1	1.60	1.60	152.87	7.23
Poisson	no	-1486.76	1	1.62	1.63	140.40	5.34
Nakagami	no	-1633.44	2	1.78	1.79	170.95	6.14
Uniform	no	-1873.11	2	2.05	2.05	237.32	4.80
Logistic	no	-2157.60	2	2.36	2.36	149.59	5.46
Normal	no	-2165.86	2	2.37	2.37	170.52	5.45
Rayleigh	no	-2353.06	1	2.57	2.57	177.27	4.14
Extreme Value	no	-2439.23	2	2.66	2.67	191.15	5.79

Notes: *n* (the number of observations) = 1,833. “No. of param.” stays for the number of parameters. WTP is in 2013/2014 USD.

Larger set of injuries							
Distribution	Zero inflation	Log-likelihood	No. of param.	AIC/n	BIC/n	WTP (mean)	WTP (s.e.)
Lewbel-Watanabe						152.25	5.65
BirnbaumSaunders	no	-1197.21	2	1.32	1.32	530.86	87.86
Gamma	no	-1198.12	2	1.32	1.32	594.60	121.68
Exponential	yes	-1199.39	2	1.32	1.32	318.93	40.32
Negative Binomial	no	-1200.06	2	1.32	1.32	461.55	144.93
Gamma	yes	-1196.98	3	1.32	1.33	400.53	124.67
Weibull	yes	-1197.50	3	1.32	1.33	458.50	176.35
Lognormal	yes	-1200.35	3	1.32	1.33	524.95	127.43
Nakagami	yes	-1200.89	3	1.32	1.33	255.69	37.70
Uniform	yes	-1201.65	3	1.32	1.33	226.08	20.81
Inverse Gaussian	yes	-1202.06	3	1.32	1.33	521.02	119.59
Weibull	no	-1207.46	2	1.33	1.33	800.28	189.09
Rayleigh	yes	-1207.50	2	1.33	1.33	225.30	17.55
Poisson	yes	-1211.35	2	1.33	1.34	222.70	14.59
Normal	yes	-1214.86	3	1.34	1.35	221.87	11.57
Logistic	yes	-1217.36	3	1.34	1.35	212.16	8.51
Extreme Value	yes	-1217.83	3	1.34	1.35	201.89	6.56
Inverse Gaussian	no	-1283.27	2	1.41	1.42	797.14	244.19
Exponential	no	-1462.79	1	1.61	1.61	177.96	7.90
Poisson	no	-1523.08	1	1.67	1.68	159.60	5.68
Nakagami	no	-1607.53	2	1.77	1.77	190.66	6.40
Uniform	no	-1787.51	2	1.96	1.97	244.00	5.22
Normal	no	-2105.65	2	2.31	2.32	189.16	5.85
Logistic	no	-2115.92	2	2.32	2.33	170.65	5.98
Rayleigh	no	-2279.77	1	2.50	2.51	194.30	4.64
Rician	no	-2279.77	2	2.50	2.51	213.19	42.38
Extreme Value	no	-2343.65	2	2.57	2.58	208.21	6.22

Notes: *n* (the number of observations) = 1,823. "No. of param." stays for the number of parameters. WTP is in 2013/2014 USD.

Appendix D. Web vs. personal surveys for elicitation of stated preferences

D.1 Literature review

Stated preference surveys are administered by various modes, which include mail, phone, web and personal (face-to-face) interviews. The prevailing view in the literature is that as long as the samples surveyed via different modes are equivalent with respect to relevant characteristics, a choice of a data collection mode does not affect the survey results significantly. Lindhjem and Navrud (2011) review 17 stated preference studies which compare web and other-mode surveys in the context of environmental goods and environment-related health risks. They conclude that in general, the studies do not evidence important differences in value estimates derived from data collected via different modes, and that data from web surveys is not observed to be of lower quality or validity than data from surveys administered with other modes. Menegaki, Olsen, and Tsagarakis (2016) identify 41 economic valuation studies conducted from 2001 to 2015 that examine differences in value estimates from web and other-mode surveys, and find that the majority of them do not confirm the existence of mode effects. Finally, the contemporary guidance for stated preference studies (Johnston et al. 2017) says that "[r]ecent research suggests that data collection mode does not substantially influence SP [stated preference] study outcomes …", however, the authors add that the results are mixed and specific to a research context. Indeed, a thorough look into studies that inquired this issue reveals that findings on the mode effect are not univocal. A summary provided in Table D1 shows that when the evidence is limited to stated preference valuation studies that compare web and personal modes, the number of studies reporting a significant mode effect is nearly the same as the number of studies reporting this effect to be insignificant.[23]

[23] Table D1 does not list studies that involve web and personal data collection modes but: (1) do not examine differences between the two modes (e.g., Hamzaoui-Essoussi and Zahaf 2012; Reichl, Schmidthaler, and Schneider 2013; Ahtiainen et al. 2014); (2) employ different preference elicitation formats in different modes (e.g., Goethals, Leclercq-Vandelannoitte, and Tütüncü 2012; Sandorf, Aanesen, and Navrud 2016; Ready et al. 2006); (3) do not include a valuation question (e.g., Goldenbeld and de Craen 2013); or (4) evaluate different goods in different modes (e.g., Maier, Wilken, and Dost 2015).

Table D1. Stated preference studies that compared outcomes from web and personal surveys using the same preference elicitation format across the modes

Author(s)	Topic	Preference elicitation format	Difference in value estimates between modes
Balderas Torres et al. (2015)	Carbon offsetting by local forests	Multiple choice sequence (DCE)	Yes (Web < Personal)
Bell, Huber, and Viscusi (2011)	Water quality in rivers, lakes and streams	Binary choice sequence (DCE)	Yes (Web < Personal)
Canavari, Nocella, and Scarpa (2005)	Pesticide ban; Organic apples	Yes-no question and open-ended question (CV); Open-ended question (CV)	No Yes (Web > Personal)
Cardamone, Eboli, and Mazzulla (2014)	Risk of road accidents	Ranking task (DCE)	No
Covey et al. (2010)	Prevention of railway fatalities	Ranking task (DCE)	No
Lee, Kim, and Mjelde (2016)	Nature preservation	Yes-no question (CV)	Yes (Web < Personal)
Lindhjem and Navrud (2011)	Biodiversity protection	Payment card question (CV)	No
Marta-Pedroso, Freitas, and Domingos (2007)	Landscape preservation	Open-ended questions (CV)	Yes (Web < Personal)
Mjelde, Kim, and Lee (2016)	Nature preservation	Multiple choice sequence (DCE)	Yes (Web < Personal)
Mulhern et al. (2013)	Health state	Binary choice sequence (DCE)	No
Nielsen (2011)	Gain in life expectancy in the context of air pollution	Open-ended questions (CV)	No
Ščasný and Alberini (2012)	Reduction of mortality risk attributable to a climate change	Multiple choice sequence (DCE)	No
van der Heide et al. (2008)	Alleviation of negative effects of habitat fragmentation	Double-bounded dichotomous choice question (CV)	Yes (Web < Personal) and No

Notes: Abbreviations CV and DCE are used to refer to the common nomenclature in the stated preference literature: CV stands for contingent valuation and DCE stands for a discrete choice experiment. Notation "Web < Personal" implies that the value estimate from a web survey is statistically significantly lower than its equivalent from a personal survey. "Web > Personal" means the opposite.

Personal interviews have long been acknowledged as the best practice in stated preference research (Arrow et al. 1993; Mitchell and Carson 1989). The NOAA panel (Arrow et al. 1993), suggesting early recommendations for stated preference studies, reason that the in-person mode helps respondents understand complex information, for example, through providing pictures and other visual material, and, hence, the mode fosters collecting data of high quality (that is, data that accurately reflects respondents' preferences). Recent guidance for stated preference research (Johnston et al. 2017) also emphasizes the advantages of using personal interviews, but they point to high cost of employing this mode. Expansion of the internet use allows researchers to administer surveys in a cheaper and faster way, at the same time retaining the possibility of presenting visual material. With the still growing access to the internet, web surveys are gaining popularity. The number of web valuation surveys conducted annually more than tripled in the years 2013-2015 in comparison with the years 2001-2007 (Menegaki, Olsen, and Tsagarakis 2016). Therefore, an essential question is whether, and if so, to what extent, a choice of a data collection mode impinges on survey outcomes.

Discrepancies in value estimates derived from web and personal modes may arise from differences in the populations that are being investigated due to internet access penetration or self-selection bias (Fricker and Schonlau 2002; Stephenson and Crête 2011). In other words, a sample of respondents to a web survey is likely to differ from a sample of respondents to a personal survey. These factors can undermine the representativeness of a web sample, and hence, they may influence the extent to which web-elicited preferences reflect preferences of the population of interest. In addition, a mode itself can alter respondents' stated answers to a survey. This is sometimes referred to as a "pure" mode effect (Jäckle, Roberts, and Lynn 2010). The "pure" mode effect can be attributed to normative/sociological factors or to cognitive/psychological factors (Dillman, Smyth, and Christian 2014). The former involve the influence of social norms on respondents' behavior, and this influence may differ between modes. In particular, the presence of an interviewer in personal surveys is likely to affect respondents' perceptions of (and adherence to) social norms. In this regard, a widely recognized source of the mode effect is social desirability, which means that respondents answer a survey in a way they think they ought to answer because of some social considerations.[24] The cognitive/psychological factors pertain to information processing by respondents. For example, a mode effect in this regard can emerge as a result of satisficing behavior (Manski 2017), which means that respondents make shortcuts and choose a satisfactory answer instead of their best answer.

Overall, as illustrated by the summary presented in Table D1, empirical evidence on the existence of a difference in value estimates derived from web and personal surveys is mixed. Out of the 13 listed studies, 7 reported a significant mode effect. Findings on the sign of this difference are not consistent either, although a majority of the studies observes that web-based data generate lower value estimates than in-person-based data. The observation that many studies find a significant mode effect diverges from the commonly held view that a choice of a

[24] Indeed, some studies observe a stronger social desirability bias in personal interviews than in web surveys (e.g., Lee, Kim, and Mjelde 2016; Mjelde, Kim, and Lee 2016).

data collection mode does not impinge on valuation results. In the face of the inconsistent evidence, in what follows we provide an additional verification, based on a field study, of whether web and personal surveys lead to equivalent value estimates.

D.2 Empirical inquiry

As explained in the main body of the paper, the Baltic Sea survey was administered through different modes (web and personal) in different countries. Poland is the only country where both modes were used, that is, CAWI and CAPI. Thus, based on the data for this country, we verify whether the value estimates differ between the two survey modes. We control for possible differences in socio-demographic characteristics between the mode samples by using weighted maximum likelihood estimation in order to capture the "pure" mode effect. The weighting also allows us for making the samples represent the general population of Poland with respect to the selected characteristics.[25]

Results included in Table D2 show that there is no overlap of 95% confidence intervals for the mean WTP estimates from the best-fitting models for the CAWI and CAPI samples in Poland. On average, CAWI respondents are willing to pay considerably more than CAPI respondents. We further use the estimated relative difference to control for the survey mode effect on the WTP estimates in other countries, thereby providing alternative estimates to those reported by Ahtiainen et al. (2014). These results are shown in Table D3.

The essential finding from this analysis is the extent to which the WTP estimates are affected by the data collection mode. Table D3 illustrates this issue. We observe large discrepancies between the value estimates, which emphasizes how considerably the mode impinges on the valuation results. Importantly, Table D3 displays differences in the average WTP values, while for policy assessments the aggregate value for the entire population is typically used. Aggregation of the average WTP values for the whole population will result in even larger differences in the value estimates derived from the two modes. Consequently, the choice of the mode may affect the evaluation of benefits from a considered policy, which in turn may matter for the decision whether the policy should be introduced or not.

Table D2. Comparison of WTP estimates from the best-fitting specifications across CAWI and CAPI samples in Poland

Distribution	95% confidence interval for mean WTP

[25] The characteristics included for weights are as per footnote 16.

	Zero inflation	Mean WTP	Standard error of mean WTP	Lower bound	Upper bound
CAWI					
Exponential	yes	18.92	1.05	16.95	21.09
Generalized Pareto	yes	18.94	1.23	16.76	21.34
Birnbaum-Saunders	yes	18.03	1.07	16.00	20.14
CAPI					
Generalized Pareto	yes	8.06	3.48	6.14	13.07
Exponential	yes	6.90	0.50	5.99	7.86
Log-normal	yes	7.72	0.98	6.06	9.93

Table D3. WTP estimates for all countries calibrated by the survey mode effect

	CAWI Mean WTP [95% confidence interval]	**CAPI** Mean WTP [95% confidence interval]
Poland	18.92 (E)	8.06 (GP)
	[16.95 - 21.09]	[6.14 - 13.07]
Denmark	37.42 (BS)	*15.95*
	[32.51 - 42.52]	*[13.85 - 18.12]*
Estonia	28.41 (IG)	*12.11*
	[22.77 - 35.04]	*[9.7 - 14.93]*
Finland	43.4 (LN)	*18.49*
	[38.45 - 50.51]	*[16.38 - 21.52]*
Germany	25.94 (IG)	*11.05*
	[22.62 - 29.77]	*[9.64 - 12.68]*
Sweden	84.1 (IG)	*35.84*
	[73.3 - 95.49]	*[31.23 - 40.69]*
Latvia	*12.46*	5.31 (E)
	[10.69 - 14.31]	[4.56 - 6.1]
Lithuania	*21.64*	9.22 (IG)
	[17.76 - 26.26]	[7.57 - 11.19]
Russia	*20.31*	8.66 (BS)
	[16.51 - 24.61]	[7.03 - 10.49]

Notes: Mean WTP estimates derived from the best-fitting specifications are reported, along with respective 95% confidence intervals in square brackets. The best-fitting distributions are denoted in round brackets: BS is Birnbaum-Saunders, E is exponential, GP is generalized Pareto, IG is inverse Gaussian, and LN is log-normal (each of the best-fitting specifications includes a zero-inflation component). The estimates in italics are calibrated WTP values assuming that the other data collection mode would have been used than the one actually implemented. The calibration is based on a proportional calculation using the ratio of the WTP estimates for Poland from CAWI and CAPI surveys.

References

Ahlvik, Lassi, Petri Ekholm, Kari Hyytiäinen, and Heikki Pitkänen. 2014. "An economic–ecological model to evaluate impacts of nutrient abatement in the Baltic Sea." Review of. *Environmental Modelling & Software* 55 (0):164-75. doi: http://dx.doi.org/10.1016/j.envsoft.2014.01.027.

Ahtiainen, Heini, Janne Artell, Mikołaj Czajkowski, Berit Hasler, Linus Hasselström, Anni Huhtala, Jürgen Meyerhoff, et al. 2014. "Benefits of meeting nutrient reduction targets for the Baltic Sea – a contingent valuation study in the nine coastal states." Review of. *Journal of Environmental Economics and Policy* 3 (3):1-28. doi: 10.1080/21606544.2014.901923.

Alberini, Anna. 1995. "Efficiency vs Bias of Willingness-to-Pay Estimates: Bivariate and Interval-Data Models." Review of. *Journal of Environmental Economics and Management* 29 (2):169-80.

Araña, Jorge E, and Carmelo J León. 2005. "Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogenity." Review of. *Journal of Environmental Economics and Management* 50 (1):170-88.

Arrow, K., R. Solow, Paul R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." Review of. *Federal Register* 58:4601-14.

Ayer, Miriam, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman. 1955. "An Empirical Distribution Function for Sampling with Incomplete Information." Review of. *Ann. Math. Statist.* 26 (4):641-7. doi: 10.1214/aoms/1177728423.

Balderas Torres, Arturo, Douglas C. MacMillan, Margaret Skutsch, and Jon C. Lovett. 2015. "'Yes-in-my-backyard': Spatial differences in the valuation of forest services and local co-benefits for carbon markets in México." Review of. *Ecological Economics* 109 (0):130-41. doi: http://dx.doi.org/10.1016/j.ecolecon.2014.11.008.

Bateman, I. J., Richard T. Carson, Brett Day, Michael W. Hanemann, Nick Hanley, Tannis Hett, Michael Jones-Lee, et al. 2004. *Economic Valuation with Stated Preference Techniques: A Manual*. Northampton, MA: Edward Elgar.

Bateman, Ian J., Ian H. Langford, R. Kerry Turner, Ken G. Willis, and Guy D. Garrod. 1995. "Elicitation and truncation effects in contingent valuation studies." Review of. *Ecological Economics* 12 (2):161-79. doi: https://doi.org/10.1016/0921-8009(94)00044-V.

Bell, Jason, Joel Huber, and W. Kip Viscusi. 2011. "Survey Mode Effects on Valuation of Environmental Goods." Review of. *International Journal of Environmental Research and Public Health* 8 (4):1222.

Bishop, Richard C., and Kevin J. Boyle. 2019. "Reliability and Validity in Nonmarket Valuation." Review of. *Environmental and Resource Economics* 72 (2):559-82. doi: 10.1007/s10640-017-0215-7.

Bishop, Richard C., Kevin J. Boyle, Richard T. Carson, David Chapman, W. Michael Hanemann, Barbara Kanninen, Raymond J. Kopp, et al. 2017. "Putting a value on injuries to natural assets: The BP oil spill." Review of. *Science* 356 (6335):253-4. doi: 10.1126/science.aam8124.

Cameron, Trudy Ann. 1988. "A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression." Review of. *Journal of Economics and Management* 15 (3):355-79.

Cameron, Trudy Ann, and Daniel D. Huppert. 1989. "OLS versus ML estimation of non-market resource values with payment card interval data." Review of. *Journal of Environmental Economics and Management* 17 (3):230-46.

Cameron, Trudy Ann, and John Quiggin. 1994. "Estimation Using Contingent Valuation Data From a "Dichotomous Choice With Follow-up" Questionnaire." Review of. *Journal of Environmental Economics and Management* 27 (3):218–34.

Canavari, Maurizio, Giuseppe Nocella, and Riccardo Scarpa. 2005. "Stated Willingness-to-Pay for Organic Fruit and Pesticide Ban." Review of. *Journal of Food Products Marketing* 11 (3):107-34. doi: 10.1300/J038v11n03_07.

Cardamone, AS, L Eboli, and Gabriella Mazzulla. 2014. "Drivers' road accident risk perception. A comparison between face-to-face interview and web-based survey." Review of. *Advances in Transportation Studies* (33):59-72.

Carson, Richard T., and M. Czajkowski. 2014. "The Discrete Choice Experiment Approach to Environmental Contingent Valuation." In *Handbook of choice modelling*, edited by Stephane Hess and Andrew Daly. Northampton, MA: Edward Elgar.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and informational properties of preference questions." Review of. *Environmental and Resource Economics* 37 (1):181-210.

Carson, Richard T., Theodore Groves, and John A. List. 2014. "Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice." Review of. *Journal of the Association of Environmental and Resource Economists* 1 (1/2):171-207.

Carson, Richard T., and Jordan Louviere. 2011. "A Common Nomenclature for Stated Preference Elicitation Approaches." Review of. *Environmental and Resource Economics* 49 (4):539-59. doi: 10.1007/s10640-010-9450-x.

Carson, Richard T., Leanne Wilks, and David Imber. 1994. "Valuing the Preservation of Australia's Kakadu Conservation Zone." Review of. *Oxford Economic Papers* 46:727-49. doi: 10.2307/2663496.

Champ, P. A., K. J. Boyle, and T. C. Brown. 2017. *A Primer on Nonmarket Valuation*. Amsterdam: Springer.

Covey, Judith, Angela Robinson, Michael Jones-Lee, and Graham Loomes. 2010. "Responsibility, scale and the valuation of rail safety." Review of. *Journal of Risk and Uncertainty* 40 (1):85-108. doi: 10.1007/s11166-009-9082-0.

Cragg, John G. 1971. "Some statistical models for limited dependent variables with application to the demand for durable goods." Review of. *Econometrica* 39 (5):829-44.

Creel, Michael, and John Loomis. 1997. "Semi-nonparametric distribution-free dichotomous choice contingent valuation." Review of. *Journal of Environmental Economics and Management* 32 (3):341-58.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4 ed: Wiley.

DWH Total Value Team. 2015. "MEMORANDUM DWH-AR0290345. Appendix 1.4: Approaches to Estimating Mean Willingness to Pay and the Effects of Covariates." In.

Freeman, A Myrick, Joseph A Herriges, and Catherine L Kling. 2014. *The measurement of environmental and resource values: Theory and methods*: Routledge.

Fricker, Ronald D., and Matthias Schonlau. 2002. "Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature." Review of. *Field Methods* 14 (4):347-67. doi: 10.1177/152582202237725.

Goethals, Frank, Aurélie Leclercq-Vandelannoitte, and Yazgi Tütüncü. 2012. "French consumers' perceptions of the unattended delivery model for e-grocery retailing." Review of. *Journal of Retailing and Consumer Services* 19 (1):133-9. doi: https://doi.org/10.1016/j.jretconser.2011.11.002.

Goldenbeld, C., and S. de Craen. 2013. "The comparison of road safety survey answers between web-panel and face-to-face; Dutch results of SARTRE-4 survey." Review of. *Journal of Safety Research* 46 (Supplement C):13-20. doi: https://doi.org/10.1016/j.jsr.2013.03.004.

Greene, William H. 2011. *Econometric Analysis*. 7 ed. Upper Saddle River, NJ: Prentice Hall.

Gurmu, Shiferaw, and Pravin K. Trivedi. 1996. "Excess Zeros in Count Models for Recreational Trips." Review of. *Journal of Business & Economic Statistics* 14 (4):469-77. doi: 10.2307/1392255.

Haab, T, and K. McConnell. 2003. *The Econometrics Of Non-Market Valuation*. Edited by Wallace E. Oates and Henk Folmer, *Valuing Environmental And Natural Resources*. Northampton, MA: Edward Elgar.

Hamzaoui-Essoussi, Leila, and Mehdi Zahaf. 2012. "Canadian Organic Food Consumers' Profile and Their Willingness to Pay Premium Prices." Review of. *Journal of International Food & Agribusiness Marketing* 24 (1):1-21. doi: 10.1080/08974438.2011.621834.

Hanemann, W. M. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." Review of. *American Journal of of Agricultural Economics* 71:1057-61.

Hanley, Nick, and Mikołaj Czajkowski. 2019. "The Role of Stated Preference Valuation Methods in Understanding Choices and Informing Policy." Review of. *Review of Environmental Economics and Policy* 13 (2):248-66. doi: 10.1093/reep/rez005.

HELCOM. 2013. "Taking Further Action to Implement the Baltic Sea Action Plan - Reaching Good Environmental Status for a healthy Baltic Sea." In. Copenhagen, Denmark: HELCOM Copenhagen Ministerial Declaration.

———. 2014. "Eutrophication status of the Baltic Sea 2007-2011 - A concise thematic assessment." Review of. *Baltic Sea Environment Proceedings* 143.

Jäckle, Annette, Caroline Roberts, and Peter Lynn. 2010. "Assessing the Effect of Data Collection Mode on Measurement." Review of. *International statistical review* 78 (1):3-20. doi: 10.1111/j.1751-5823.2010.00102.x.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, et al. 2017. "Contemporary Guidance for Stated Preference Studies." Review of. *Journal of the Association of Environmental and Resource Economists* 4 (2):319-405.

Kerr, G. N. 2000. "Dichotomous choice contingent valuation probability distributions." Review of. *Australian Journal of Agricultural and Resource Economics* 44 (2):233-52.

Kiirikki, Mikko, Arto Inkala, Harri Kuosa, Heikki Pitkänen, Minna Kuusisto, and Juha Sarkkula. 2001. "Evaluating the effects of nutrient load reductions on the biomass of toxic nitrogenfixing cyanobacteria in the Gulf of Finland, Baltic Sea." Review of. *Boreal Environment Research* 6:1-16.

Kiirikki, Mikko, Jouni Lehtoranta, Arto Inkala, Heikki Pitkänen, Susanna Hietanen, Per O. J. Hall, Anders Tengberg, Jorma Koponen, and Juha Sarkkula. 2006. "A simple sediment process description suitable for 3D-ecosystem modelling — Development and testing in the Gulf of Finland." Review of. *Journal of Marine Systems* 61 (1–2):55-66. doi: http://dx.doi.org/10.1016/j.jmarsys.2006.02.008.

Krinsky, Itzhak, and A. L. Robb. 1991. "Three methods for calculating the statistical properties of elasticities: A comparison." Review of. *Empirical Economics* 16 (2):199-209.

Krinsky, Itzhak, and A. Leslie Robb. 1986. "On approximating the statistical properties of elasticities." Review of. *The Review of Economics and Statistics* 68 (4):715-9.

Kriström, B. 1997. "Spike Models in Contingent Valuation." Review of. *American Journal of Agricultural Economics* 79 (3):1013-23.

Langford, Ian H., Areti Kontogianni, Mihalis S. Skourtos, Stavros Georgiou, and Ian J. Bateman. 1998. "Multivariate Mixed Models for Open-Ended Contingent Valuation Data: Willingness To Pay For Conservation of Monk Seals." Review of. *Environmental and Resource Economics* 12 (4):443-56. doi: 10.1023/A:1008286001085.

Layton, David F, and Klaus Moeltner. 2005. "The cost of power outages to heterogeneous households." In *Applications of Simulation Methods in Environmental and Resource Economics*, 35-54. Springer.

Lee, Choong-Ki, Tae-Kyun Kim, and James W. Mjelde. 2016. "Comparison of preservation values between Internet and interview survey modes: the case of Dokdo, South Korea." Review of. *Journal of Environmental Planning and Management* 59 (1):22-43. doi: 10.1080/09640568.2014.980900.

Lewbel, Arthur. 2000. "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." Review of. *Journal of Econometrics* 97 (1):145-77. doi: https://doi.org/10.1016/S0304-4076(00)00015-4.

Lindhjem, Henrik, and Ståle Navrud. 2011. "Are Internet surveys an alternative to face-to-face interviews in contingent valuation?" Review of. *Ecological Economics* 70 (9):1628-37. doi: http://dx.doi.org/10.1016/j.ecolecon.2011.04.002.

Maar, Marie, Eva Friis Møller, Jesper Larsen, Kristine Skovgaard Madsen, Zhenwen Wan, Jun She, Lars Jonasson, and Thomas Neumann. 2011. "Ecosystem modelling across a salinity gradient from the North Sea to the Baltic Sea." Review of. *Ecological Modelling* 222 (10):1696-711. doi: http://dx.doi.org/10.1016/j.ecolmodel.2011.03.006.

Maier, Erik, Robert Wilken, and Florian Dost. 2015. "The double benefits of consumer certainty: combining risk and range effects." Review of. *Marketing Letters* 26 (4):473-88. doi: 10.1007/s11002-014-9282-5.

Manski, Charles F. 2017. "Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment." Review of. *Theory and Decision* 83 (2):155-73. doi: 10.1007/s11238-017-9592-1.

Mariel, Petr, David Hoyos, Jürgen Meyerhoff, Mikolaj Czajkowski, Thijs Dekker, Klaus Glenk, Jette Bredahl Jacobsen, et al. 2020. *Environmental Valuation with Discrete Choice Experiments. Guidance on Design, Implementation and Data Analysis*: Springer.

Markowska, A., and T. Żylicz. 1999. "Costing an International Public Good: The Case of the Baltic Sea." Review of. *Ecological Economics* 30:301-16.

Marta-Pedroso, Cristina, Helena Freitas, and Tiago Domingos. 2007. "Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews." Review of. *Ecological Economics* 62 (3):388-98. doi: https://doi.org/10.1016/j.ecolecon.2007.02.005.

McFadden, D. . 1974. "Conditional Logit Analysis of Qualititative Choice Behaviour." In *Frontiers in Econometrics*, edited by P. Zarembka, 105-42. New York, NY: Academic Press.

Menegaki, Angeliki N., Søren Bøye Olsen, and Konstantinos P. Tsagarakis. 2016. "Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys." Review of. *Journal of Choice Modelling* 18 (Supplement C):18-50. doi: https://doi.org/10.1016/j.jocm.2016.04.005.

Mitchell, R.C., and R.T. Carson. 1989. *Using Surveys to Value Public Goods: the Contingent Valuation Method*. Washington, DC: Resources for the Future.

Mjelde, James W., Tae-Kyun Kim, and Choong-Ki Lee. 2016. "Comparison of Internet and interview survey modes when estimating willingness to pay using choice experiments." Review of. *Applied Economics Letters* 23 (1):74-7. doi: 10.1080/13504851.2015.1051648.

Mulhern, Brendan, Louise Longworth, John Brazier, Donna Rowen, Nick Bansback, Nancy Devlin, and Aki Tsuchiya. 2013. "Binary Choice Health State Valuation and Mode of Administration: Head-to-Head Comparison of Online and CAPI." Review of. *Value in Health* 16 (1):104-13. doi: https://doi.org/10.1016/j.jval.2012.09.001.

Nielsen, Jytte Seested. 2011. "Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews." Review of. *Resource and Energy Economics* 33 (1):119-29. doi: https://doi.org/10.1016/j.reseneeco.2010.01.006.

Pakalniete, Kristine, Juris Aigars, Mikołaj Czajkowski, Solvita Strake, Ewa Zawojska, and Nick Hanley. 2017. "Understanding the distribution of economic benefits from improving coastal and marine ecosystems." Review of. *Science of The Total Environment* 584-585:29-40.

Ready, Richard, Ann Fisher, Dennis Guignet, Richard Stedman, and Junchao Wang. 2006. "A pilot test of a new stated preference valuation method: Continuous attribute-based stated choice." Review of. *Ecological Economics* 59 (3):247-55. doi: https://doi.org/10.1016/j.ecolecon.2005.11.004.

Reichl, Johannes, Michael Schmidthaler, and Friedrich Schneider. 2013. "The value of supply security: The costs of power outages to Austrian households, firms and the public sector." Review of. *Energy Economics* 36 (Supplement C):256-61. doi: https://doi.org/10.1016/j.eneco.2012.08.044.

Reusch, Thorsten B. H., Jan Dierking, Helen C. Andersson, Erik Bonsdorff, Jacob Carstensen, Michele Casini, Mikolaj Czajkowski, et al. 2018. "The Baltic Sea as a time machine for the future coastal ocean." Review of. *Science Advances* 4 (5). doi: 10.1126/sciadv.aar8195.

Roach, B., K. J. Boyle, and M. P. Welsh. 2002. "Testing Bid Design Effects in Multiple Bounded Contingent Valuation." Review of. *Land Economics* 78 (1):121-31.

Rowe, R. D., W. D. Schulze, and W. S Breffle. 1996. "A Test for Payment Card Biases." Review of. *Journal of Environmental Economics and Management* 31 (2):178-85.

Sandorf, Erlend Dancke, Margrethe Aanesen, and Ståle Navrud. 2016. "Valuing unfamiliar and complex environmental goods: A comparison of valuation workshops and internet panel surveys with videos." Review of. *Ecological Economics* 129 (Supplement C):50-61. doi: https://doi.org/10.1016/j.ecolecon.2016.06.008.

Scarpa, Riccardo, Mara Thiene, and Kenneth Train. 2008. "Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps." Review of. *American Journal of Agricultural Economics* 90 (4):994-1010. doi: 10.1111/j.1467-8276.2008.01155.x.

Ščasný, Milan, and Anna Alberini. 2012. "Valuation of Mortality Risk Attributable to Climate Change: Investigating the Effect of Survey Administration Modes on a VSL." Review of. *International Journal of Environmental Research and Public Health* 9 (12):4760.

Söderqvist, T., and H. Scharin. 2000. "The regional willingness to pay for a reduced eutrophication in the Stockholm archipelago." In. Beijer Discussion Paper Series No. 128: The Beijer Institute.

Stephenson, Laura B., and Jean Crête. 2011. "Studying Political Behavior: A Comparison of Internet and Telephone Surveys." Review of. *International Journal of Public Opinion Research* 23 (1):24-55. doi: 10.1093/ijpor/edq025.

Strazzera, E., R. Scarpa, P. Calia, G. D. Garrod, and K. G. Willis. 2003. "Modelling zero values and protest responses in contingent valuation surveys." Review of. *Applied Economics* 35:133-8.

Tuhkanen, Heidi, Evelin Piirsalu, Tea Nõmmann, Aljona Karlõševa, Sulev Nõmmann, Mikołaj Czajkowski, and Nick Hanley. 2016. "Valuing the benefits of improved marine environmental quality under multiple stressors." Review of. *Science of The Total Environment* 551-552:367-75.

Turnbull, Bruce W. 1976. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data." Review of. *Journal of the Royal Statistical Society. Series B (Methodological)* 38 (3):290-5.

van der Heide, C. Martijn, Jeroen C. J. M. van den Bergh, Ekko C. van Ierland, and Paulo A. L. D. Nunes. 2008. "Economic valuation of habitat defragmentation: A study of the Veluwe, the Netherlands." Review of. *Ecological Economics* 67 (2):205-16. doi: https://doi.org/10.1016/j.ecolecon.2008.04.012.

Vossler, Christian A., and J. Scott Holladay. 2018. "Alternative value elicitation formats in contingent valuation: Mechanism design and convergent validity." Review of. *Journal of Public Economics* 165:133-45. doi: https://doi.org/10.1016/j.jpubeco.2018.07.004.

Vossler, Christian A., and Ewa Zawojska. 2020. "Behavioral Drivers or Economic Incentives? Toward a Better Understanding of Elicitation Effects in Stated Preference Studies." Review of. *Journal of the Association of Environmental and Resource Economists* 7 (2):279-303. doi: 10.1086/706645.

Watanabe, Masahide. 2010. "Nonparametric Estimation of Mean Willingness to Pay from Discrete Response Valuation Data." Review of. *American Journal of Agricultural Economics* 92 (4):1114-35. doi: 10.1093/ajae/aaq034.

Werner, Megan. 1999. "Allowing for zeros in dichotomous-choice contingent-valuation models." Review of. *Journal of Business & Economic Statistics* 17 (4):479-86.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl