Check for updates

Diagnosing Insensitivity to Scope in Contingent Valuation

Ana Faria Lopes[1] · Gorm Kipperberg[1]

Abstract

Sensitivity to scope is considered a desirable property of contingent valuation studies and often treated as a necessary condition for validity. We first provide an overview of scope *insensitivity* explanations put forth in the environmental valuation literature. Then we analyze data from a contingent valuation survey eliciting willingness-to-pay to prevent oil spills of four different magnitudes in Arctic Norway. In the baseline analysis, the scope inference is ambiguous. There is only statistical difference in willingness to pay to avoid a very large versus small oil spill (NOK 1869 and NOK 1086, respectively). However, further explorations show that several confounding factors suggested in the literature influence the scope inference. The scope sensitivity improves when we control for subjective probabilities of amenity provision, exclude respondents based on the debriefing questions, take into consideration the sample sizes, and impose diminishing marginal utility. Overall, the analysis supports an emerging view in the contingent valuation literature suggesting that statistical scope insensitivity is not a sufficient reason for deeming a study invalid.

Keywords Contingent valuation · Sensitivity to scope · Scope testing · Scope elasticity · Willingness-to-pay

Abbreviations

CV	Contingent valuation
NOAA	National Oceanic and Atmospheric Administration
NOK	Norwegian kroner
SP	Stated preference
WTP	Willingness-to-pay

✉ Ana Faria Lopes
ana.f.lopes@uis.no

[1] UiS Business School, University of Stavanger, 4036 Stavanger, Norway

1 Introduction

Basic microeconomic intuition suggests that "It is reasonable to assume that larger amounts of commodities are preferred to smaller ones" (Mas-Colell, Whinston, and Green 1995). Therefore, it is generally expected that respondents are willing to pay more for preventing a larger damage or receiving a higher quantity or quality of a good (e.g., Smith and Osborne 1996; Carson et al. 2001; Whitehead 2016). This empirical expectation follows formally from the monotonicity (non-satiation) axiom of consumer preferences (e.g., Varian 2014). In Contingent Valuation (CV) studies, this property is known as scope sensitivity.[1,2]

According to the National Oceanic and Atmospheric Administration (NOAA) Blue Ribbon Panel on CV, presence of scope sensitivity is evidence of internal or construct validity, while absence of scope sensitivity puts the validity of the study into question (Arrow et al. 1993). For this reason, the NOAA Panel recommends that welfare measures are tested for sensitivity to scope "…in order to assure reliability and usefulness of the information" (Arrow et al. 1993, p. 34). This recommendation is recently reiterated in the general guidelines for stated preference (SP) research of Johnston et al. (2017).

The scope sensitivity issue has been a point of controversy in non-market valuation for over 30 years, in part because earlier studies failed to find statistically significant increases in willingness-to-pay (WTP) with the magnitude of the good being valued (e.g., Boyle et al. 1994). Some critics go as far as to use examples of scope insensitivity to argue that the CV method is a generally flawed approach to capturing non-market values (Diamond and Hausman 1994; Hausman 2012). Nonetheless, thousands of CV studies have been carried out over this time span (Carson 2012), many of them demonstrating presence of scope effects. As a result, recently a more moderate perspective has emerged, which suggests that failing a statistical scope test is not the ultimate evidence against a CV study's validity (e.g., Heberlein et al. 2005; Banerjee and Murphy 2005; Amiran and Hagen 2010; Desvousges et al. 2012; Whitehead 2016). For one, statistical scope tests can lead to false negatives for a variety of "…reasons that are quite compatible with fundamental economic reasoning and social psychological theory" (Heberlein et al. 2005, p. 3). For example, if utility is sharply diminishing in the quantity or quality of a particular good, it may be difficult to establish statistically significant effects (Rollins and Lyke 1998). Relatedly, the NOAA Panel emphasized that CV studies should demonstrate *adequacy* of scope, not necessarily statistical significance (Arrow et al. 1993, 1994). This is an important subtlety of the NOAA Panel's recommendations, which, unfortunately, is often overlooked. In an addendum to the original report, the Panel warns that inference from statistical scope tests may cause misleading results when the goal is to inform the plausibility or adequacy of scope (Arrow et al. 1994). Here, the NOAA Panel suggests that "…a survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult" (Arrow et al. 1994, p. 123).

---

[1] This paper focus on the concept of scope sensitivity rather than embedding sensitivity. Studying the latter is typically understood as comparing welfare measures for a good on its own with its value when evaluated as part of a broader package of goods (Kahneman and Knetsch 1992). Nonetheless, in our own application, one could think of the largest of the four oil spills as containing or "embedding" each of the three ones. We thank an anonymous reviewer for pointing this out.

[2] The scope sensitivity issue is also discussed in other non-market valuation areas such as health research (e.g. Søgaard et al. 2012) and marketing (e.g., Urminsky and Kivetz 2011) and is relevant for SP research in general (Johnston et al. 2017).

With increasing awareness of the conceptual and empirical complexity of the scope sensitivity issue, researchers have recently shifted their focus away from conventional tests towards defining and testing for adequacy of scope instead (e.g., Amiran and Hagen 2010; Whitehead 2016). Perhaps the most promising approach is the construction of WTP *scope elasticities*.[3] Scope elasticity measures the percentage change in WTP associated with a percentage change in the magnitude of the good, and as such, can be utilized to assess the *economic significance* rather than the statistical significance of scope impacts (Amiran and Hagen 2010; Whitehead 2016). Examples of recent studies that use the scope elasticity concept are Whitehead (2016), Burrows et al. (2017), and Borzykowski et al. (2018). What constitute economic significance scope elasticity magnitudes remain unsettled in this emerging literature. The conceptual analysis in Amiran and Hagen (2010) argue that the scope elasticity should be anywhere between 0 and 1 in order to be consistent with strictly convex neoclassical preferences. Both Whitehead (2016) and Borzykowski et al. (2018) interpret elasticities higher than and statistically different from zero as "plausible" sensitivity to scope, while Burrows et al. (2017) suggest "adequate" scope elasticities thresholds of 0.2 or 0.5.

With the above discussion as a broad motivation, our overall aim is to study the *scope insensitivity* phenomenon in the context of the environmental CV literature. The paper contributes to the literature in the following two ways: The first part of the paper provides a broad overview of explanations for scope insensitivity that have been put forth in previous research. This overview fills a gap in the literature as several authors have called for a thorough investigation of scope-confounding factors, which could potentially lead to *false negatives* (Carson and Mitchell 1995; Whitehead et al. 1998; Heberlein et al. 2005; Desvousges et al. 2012; Whitehead 2016). Some of the common explanations, for example, *diminishing marginal utility* (Arrow et al. 1993; Rollins and Lyke 1998) and *amenity misspecification* (Boyle at al. 1994; Carson and Mitchell 1995), make intuitive sense. However, only a few studies have documented their empirical influence on scope inference (e.g., Bateman et al. 2004; Siikamäki and Larson 2015). Hence, the second part of the paper explores several scope insensitivity explanations in an *ex post* analysis of CV data on WTP for preventing oil spills in Arctic Norway. In particular, the data comes from a survey that included a quadruple split sample experimental design to explore variation in WTP across different oil spill scenarios: small, medium, large, and very large oil spill (see Part 3 for details). Our analysis utilizes variation in preference expressions across survey participants (external scope), rather than multiple responses from each participant (internal scope). We provide baseline results, which pass a statistical scope test only in the case of comparing WTP to avoid the largest versus the smallest oil spill. Then we analyze a series of potentially scope-confounding factors and find that the scope sensitivity improves when we consider several of these. We also compute scope elasticities to assess the economic significance, or adequacy, of the estimated scope impacts. Previous oil spill valuation studies have typically employed an internal CV scope test (Rowe et al. 1992; Van Biervliet et al. 2006; Navrud et al. 2017), investigated sensitivity to scope through choice experiments (e.g., Casey et al. 2008), or not included a scope test (e.g., Loureiro et al. 2009). To our knowledge, only Desvousges et al. (1992), Barton et al. (2003), and Bishop et al. (2017) have utilized an external test before, and only the study by Bishop et al. (2017)

---

[3] An alternative test for adequacy of scope sensitivity is the adding-up test (see Diamond 1996; Desvousges et al. 2012 for an explanation). This test is rarely used as it requires a more complex and costly experimental survey design (Whitehead 2016).

passed the scope test. Analysis of confounding effects and respective impact on scope elasticities has not been carried out in this literature before.

The remainder of the paper is organized as follows: Part 2 classifies and discusses scope insensitivity explanations proposed in the literature. Subsequently, Part 3 presents the context, design and implementation of our case study, while Part 4 presents our empirical scope analysis. Part 5 concludes.

2 Scope Insensitivity

To identify previously proposed scope insensitivity explanations, we conduct a narrative review (Borenstein et al. 2011). We limit our selection to peer-reviewed studies in the field of valuation of environmental goods and services that focus on the contingent valuation method. In summary, we find 13 alternative explanations for presence of scope insensitivity put forth in the literature, which are explored throughout this section and summarized in Table 1.

While there are various ways these could be classified, we do so into four broad categories: (1) explanations related to *microeconomic consumer theory*, (2) explanations related to *how people relate to environmental goods*, (3) explanations related to *survey design and model estimation*, and (4) explanations related to *insights from behavioral economics*.

2.1 Neoclassical Microeconomic Consumer Theory

One of the most fundamental explanations for scope insensitivity is diminishing marginal utility (Arrow et al. 1993; Boyle et al. 1994; Whitehead 2016). As the size of the environmental good being valued increases, the marginal increments in utility become smaller, leading to apparent insensitivity to scope. In their study about the conservation of the Giant Panda, Kontoleon and Swanson (2003) find a WTP of $0.72 per hectare for the first five hectares, which decreases significantly as the number of hectares increases. At 200 hectares of land, marginal WTP per hectare was estimated at virtually zero. These results illustrate the conclusions of Rollins and Lyke (1998), who suggest that whether sensitivity to scope is found is conditional on the sizes of the environmental good being elicited. To observe statistical scope is challenging (if not impossible) if the researcher is eliciting WTP for high levels of an environmental good.

More broadly, recent research has pointed out that there are utility functions compatible with most preference axioms of neoclassical consumer theory that can exhibit a small degree of sensitivity to scope. Amiran and Hagen (2010) prove that utility functions that are not directionally bounded always yield scope sensitivity. However, directionally bounded utility functions can yield arbitrarily small sensitivity to scope. The Leontief utility function is another example: despite representing regular preferences, it yields scope insensitivity if WTP is measured along the flat segment of the indifference curve (Banerjee and Murphy 2005).

Related to utility functions, the degree of substitutability between market goods and non-market environmental goods affects scope findings (Smith and Osborne 1996; Amiran and Hagen 2010; Whitehead 2016). We illustrate this idea with two extreme examples. In the case of hypersubstitutability, wherein "… the consumer would be willing to forgo nearly all consumption of market goods… in exchange for a sufficiently large increment of the environmental amenity" e.g., Cobb–Douglas utility functions, insensitivity to scope

Table 1 Overview of potentially scope-confounding factors

	Scope insensitivity explanation	First formalization or mentioning in the CV literature
Neoclassical microeconomic consumer theory	Diminishing marginal utility	"WTP increases as the scope of the good increases, but the marginal WTP decreases as more of the good is offered to the consumer" (Rollins and Lyke 1998)
	Utility functions	"Very small sensitivity of scope can be achieved with underlying utility functions that are well-behaved" (Amiran and Hagen 2010)
	Substitutability between market and non-market environmental goods	"The presence of restrictions on substitution are shown to have important implications for the degree of sensitivity to scope" (Amiran and Hagen 2010)
	Incomplete multi-stage budgeting	"Household discretionary budgets may amount to only a modest proportion of household wealth" (Randall and Hoehn 1996)
How people relate to the environmental good	Experience, familiarity, knowledge and/or use	Respondents value environmental goods more "when [they] have knowledge about, experience with, and well-formed attitudes towards the good" (Heberlein et al. 2005)
	Preference Heterogeneity	"False negatives in scope tests can result when individual preference variation and correlation are ignored" (Siikamäki and Larson 2015)
Survey design and model estimation	Survey design	"CV questions have to be posed carefully and in context" (Arrow et al. 1993) with a "description of (…) the goods that respondents understand and a method of provision they find plausible" (Carson and Mitchell 1995)
	Amenity misspecification	Respondents may "make assumptions about the good that they think the interviewer has in mind" if it is vaguely described (Carson and Mitchell 1995)
	Data cleaning	How observations are selected in the data analysis stage could potentially have impacts on scope findings (Whitehead et al. 1998)
	Statistical distribution assumption	"Mean WTP and scope effects are sensitive to the statistical distribution assumption" (Borzykowski et al. 2018)
	Sample size	Sensitivity to scope depends on the sample size: "in small samples, no effects are statistically significant. In large samples, everything is statistically significant" (Arrow et al. 1994)
Insights from behavioral economics	Warm glow	"WTP for public goods is best interpreted as the purchase of moral satisfaction, rather than as a measure of the value associated with a particular public good" (Kahneman and Knetsch 1992)
	Preference reversal theory	Preference reversal refers to circumstances where individuals shift their preference from one good to another depending on the way the good is elicited, for example in "joint and isolated evaluation modes" (Alevy et al. 2011)

should never arise (Amiran and Hagen 2010). On the other extreme, if the market and environmental goods are perfect complements, the consumers' implicit demand for environmental good is irresponsive to positive changes in the provision level (Amiran and Hagen 2010).

The idea that the bid amounts in CV surveys typically represent only a small fraction of the total household budget has led researchers to assume that the budget constraint of respondents "does not bind very tightly" (Hausman 2012) or that median WTP "… is far too small to be severely restrained by wealth" (Kahneman and Knetsch 1992). However, in the short-run, when respondents are asked for their WTP, some household expenditures are fixed. Thus, the budget potentially allocated to the provision of environmental goods being offered might only be a fraction of the total household budget (Randall and Hoehn 1996). Randall and Hoehn (1996) refer to this as incomplete multi-stage budgeting.

2.2 How People Relate to the Environmental Good

Frederick and Fischhoff (1998), Whitehead et al. (1998), Heberlein et al. (2005), and Alevy et al. (2011) suggest that the way individuals relate to the environmental good may influence the scope findings. These individual characteristics may include increased experience, familiarity, knowledge and/or use of the environmental good. For example, users of the environmental good, who are more familiar with it, are more likely to be sensitive to its size when eliciting the WTP compared with nonusers (Frederick and Fischhoff 1998). In their study on biodiversity, Heberlein et al. (2005) find that to know more, to like more, and have more experience at local level (2 counties) rather than at a broader level, lead respondents to value more local diversity than biodiversity in a broader region.

Preference heterogeneity has also been shown to lead to scope insensitive results (Siikamäki and Larson 2015; Giguere et al. 2020). Some individuals might value highly the provision of a single attribute in the bundle of the environmental good provided, while others value all or no attributes in a more balanced manner. One can account for preference heterogeneity by, for example, estimating random parameter models (Siikamäki and Larson 2015), latent class models or using stated attribute non-attendance (Giguere et al. 2020). In their application to water quality improvements in California, Siikamäki and Larson (2015) show that clear sensitivity to scope emerges when accounting for unobserved preference heterogeneity in a mixed logit model. Giguere et al. (2020) find that when accounting for stated attribute non-attendance, the study passes the statistical scope test. Respondents may also shift preferences towards an environmental good at some individual-specific threshold level. For example, in Heberlein et al. (2005)'s study respondents exhibit preferences for the same good in different directions in the case of protecting wolf populations.

2.3 Survey Design and Model Estimation

Poor survey design has been pointed out for some decades to be the main cause for insensitivity to scope (Carson and Mitchell 1995; Carson 1997; Carson et al. 2001; Heberlein et al. 2005; Whitehead 2016). To avoid poor design of a CV study, the NOAA Panel lists several core recommendations, including careful pretesting (Arrow et al. 1993). Survey design flaws that affect scope sensitivity can come in many ways, but their main consequence is on survey consequentiality. Designing a survey that is consequential, i.e. so that the respondents perceive the survey's results as potentially influencing an agency's actions, implies a higher likelihood of finding scope sensitivity (Carson and Groves 2007).

Two examples of how survey design can influence scope findings are: the mode of survey administration and stepwise versus advanced disclosure.[4] Evidence regarding the mode of survey administration is mixed. While Arrow et al. (1993), Carson and Mitchell (1995) and Carson (1997) favor the use of face-to-face interviews over phone, mall interviews or internet panels (Burrows et al. 2017), Whitehead et al. (1998) still find sensitivity to scope when using a phone survey. In the case of stepwise versus advanced disclosure, Bateman et al. (2004) find that the advance disclosure approach yields both internal and external scope sensitivity, while stepwise approaches may not yield scope sensitive results.

Another main factor for scope insensitivity is amenity misspecification (Boyle at al. 1994; Carson and Mitchell 1995; Rollins and Lyke 1998). Rather than how respondents relate to the good, amenity misspecification refers to how respondents perceive the good as described in the survey. This comes in four types: part-whole, metric, probability of provision or symbolic bias. *Part-whole bias* entails that respondents "make assumptions about the good that they think the interviewer has in mind" because it is vaguely described (Carson and Mitchell 1995).

Metric bias occurs if "a respondent values the amenity on a different (and usually less precise) metric or scale than the one intended by the researcher" (Mitchell and Carson 1989). Changes in the size of the environmental good can be described in either relative or absolute terms, or quantitative or qualitative measures (e.g. Boyle et al. 1994 describes changes in bird population in both relative and absolute terms). Ojea and Loureiro (2011)'s meta-analysis suggests that WTP estimates are more sensitive to scope if changes in the environmental good are described quantitatively and in absolute terms.

Probability of provision bias implies that "the perceived probability that the good will be provided differs from the researcher's intended probability" for different sizes of the good (Mitchell and Carson 1989). If respondents subjectively assign a higher probability for provision of a smaller environmental good, respondents may be willing to pay more when compared to a larger but less probable provision level (Carson and Mitchell 1995).

Symbolic bias refers to the case of small damages being perceived as "symbolic for a good of greater magnitude" (Mitchell and Carson 1989; Carson 1997), hence respondents react to the symbolism of the good rather than the size of its provision. For example, Czajkowski and Hanley (2009) find that WTP for protection of forest biodiversity becomes less sensitive to scope when a "natural park" label is included in the valuation exercise.

Data cleaning (i.e. identification of valid responses) could potentially have impacts on scope findings. Observations may be included or excluded based on individuals' responses to debriefing questions, missing data, or any other criterion. However, excluding or including observations has an impact on sample size, which in turn affects the efficiency of the statistical scope test. Valuing water quality improvements, Whitehead et al. (1998) find no impact from inclusion or exclusion of protesters, outliers, nor from the treatment of "don't know" responses.

Estimated WTP has been shown to be sensitive to distributional assumptions, namely when comparing parametric versus non-parametric estimation approaches (Haab and McConnell 1997) or estimating data from different elicitation formats, e.g. single versus double-bounded dichotomous choice (Alberini 1995). There is also some evidence

---

[4] Stepwise disclosure means that respondents answering the valuation questions without posterior knowledge about the number of choices; advance disclosure entails informing respondents beforehand about all the questions they will be asked.

suggesting that statistical distribution assumptions may also affect scope sensitivity findings (Whitehead et al. 1997; Berrens et al. 2000; Borzykowski et al. 2018).

Finally, many of the scope insensitive findings are attributed to small sample sizes (Boyle et al. 1994; Carson and Mitchell 1995; Carson 1997; Rollins and Lyke 1998; Carson et al. 2001; Whitehead 2016). If the sample size is small, the experiment will not have enough statistical power to identify the scope effect. Exploring scope sensitivity is traditionally done through statistical tests. However, as Arrow et al. (1994) put it, "The fundamental problem with any purely statistical definition of sensitivity is that it depends (foolishly) on the sample size". Rollins and Lyke (1998) point out that finding scope with small sample sizes is even more challenging if the baseline size of the environmental good is already relatively high.

2.4 Insights from Behavioral Economics

Insights from behavioral economics can prove to be valuable to understand WTP estimates (Kling et al. 2012; Freeman et al. 2014). Poe (2016) summarizes a variety of behavioral anomalies occurring in SP research common to analysis of observed behavior. Insights from observed behavior may help "identifying the cognitive underpinnings of scope effects" (Alevy et al. 2011). If the standard preference axioms of consumer theory (i.e. regular, continuous, strongly monotonic and strictly convex preferences) do not hold, then behavioral anomalies may cause scope insensitivity (Banerjee and Murphy 2005; Whitehead 2016).

One of the most influential papers that initiated the scope debate was Kahneman and Knetsch (1992). Through an embedding experiment, Kahneman and Knetsch (1992) find that the same good is assigned a lower WTP if valued as part of a bundle rather than on its own, leading the authors to conclude that respondents are willing to pay to acquire moral satisfaction rather than revealing their true preferences regarding the environmental good. This effect was also subsequently referred to as "warm glow". If respondents have strong warm glow motivations, then changing the scope of the good "should have little effect on WTP" (Kahneman and Knetsch 1992).

A more recent example of how behavioral economics explains scope findings is preference reversal theory (Alevy et al. 2011). This theory that has also been observed for market goods suggests that individuals may shift their preferences from one good to another depending on the way the good is elicited. Alevy et al. (2011) find that respondents have different preferences towards watershed and farmland preservation when valuing each isolated rather than jointly.

3 The Lofoten Oil Spill Prevention Study

Our data for exploring sensitivity to scope comes from a CV survey focusing on the Norwegian population's WTP for preventing oil spills in the Lofoten Archipelago. This archipelago is an iconic coastal area in Arctic Norway, which is under increasing pressure from economic activities along the coast. Moreover, Norwegian politicians are continuously debating whether to lift the current ban on petroleum exploration outside the Lofoten Archipelago. Estimating the lost non-market values in the case of an oil spill in this area is important for public policy.

3.1 Survey Design and Questionnaire Structure

The survey design was initiated in early 2012 based on several previous oil spill prevention CV surveys (Carson et al. 2003; Loureiro et al. 2009; Carson et al. 2013). A draft survey was then distributed to valuation experts for feedback and subsequently tested in face-to-face interviews with members of the university administrative staff. An updated version was tested in focus groups comprising individuals from the general population. The development of the CV survey was done in collaboration with another team of valuation researchers, which, concurrently, was seeking to study local preferences for preventing oil spills at multiple locations along the Norwegian coast (Navrud et al. 2017). Feedback and comments received during the pretesting stages were incorporated on an ongoing basis before arriving at the final instrument in early 2013.

The CV experiment begins with questions about oil spill knowledge and experience, and reasons why it might be important to prevent oil spills. It then informs that an oil spill will occur as a result of a ship accident with certainty in the Lofoten Archipelago within the next 10 years, if additional preventive and emergency preparedness measures are not implemented. The oil spill scenario is described by an oil spill dispersion map (Appendix 1 in the Online Appendix) and a damage table (Appendix 2 in the Online Appendix). The damages from the oil spill are described quantitatively in terms of bird and seal mortality, kilometers of shoreline soiled, and the recovery time for safe seafood consumption. Preferences are then elicited with a single-bounded, closed-ended referendum question asking about willingness to pay an annual tax increase to prevent the oil spill. Both the tax amounts and the oil spill sizes are randomized across participants in the survey. The tax amounts range from NOK 100 to NOK 2500, while the four oil spill sizes are labeled "small", "medium", "large", and "very large". Next, response certainty (on a 1–10 scale) and up to three reasons for answering yes/no to the proposed tax increase are elicited as debriefing to the referendum question. Lastly the CV part probes subjective oil spill occurrence probabilities, the likelihood that government will use the survey result to design oil spill prevention policies, and the likelihood of having to pay higher taxes.[5]

3.2 Data Collection, Cleaning, and Sample Representativeness

The data collection was executed as a web-based survey in April 2013, employing the pre-recruited national household panel of *NORSTAT*, a leading survey sampling company in Norway.[6] The full dataset consists of 1400 respondents with 500 observations each for the *small* and *very large* oil spill scenarios and 200 observations each for the *medium* and *large* scenarios.[7]

---

[5] In contrast to our study, Navrud et al. (2017) use payment card format to elicit a one-time tax payment for preventing oil spills. Furthermore, they employ an internal scope test by asking the respondents about WTP for each of the four oil spill sizes. Finally, Navrud et al. (2017) focus on WTP for preventing oil spills at different locations along the Norwegian coast, not only the Lofoten Archipelago.

[6] See www.norstat.no. The full survey questionnaire is available as Supplementary material.

[7] The original sampling goal was 300, 200, 200, and 300 responses across the four oil spill scenarios, with the aim of ensuring a relatively higher degree of statistical precision for the welfare estimates associated with the smallest and largest oil spills. When lower-than-expected survey costs permitted sampling of an additional 400 respondents, we decided to have these drawn exclusively from the two extreme scenarios.

Extensive data checking and cleaning routines were carried out prior to the sensitivity to scope analysis. First, 10 respondents who completed the survey in less than 10% of the average completion time (20 min) were removed. Second, debriefing questions were utilized to exclude protesters, strategic bidders, and respondents with lack of belief in the study's consequentiality. Respondents were retained if at least one valid reason for answering yes or no was given.[8] About 25% of the respondents (354) were dropped from further analysis by this criterion. Third, to correct for potential hypothetical bias (Kling et al. 2012; Haab et al. 2013; Loomis 2014) uncertain yes-respondents (scores below 7 out of 10) were re-coded as no-respondents (Champ et al. 2009). In total 149 responses (10.6%) were recoded from "yes" to "no" by this procedure.[9]

Sample representativeness was confirmed as expected given that the respondents were randomly drawn from a pre-recruited web-panel constructed to represent the Norwegian population. For example, average annual income (NOK 718 712) and gender (49% men) are almost identical to the official statistics (NOK 730 800 and 50%, respectively). However, as is often common in social science surveys, the sample was more educated than the general population. The socioeconomic and demographic profile of the reduced sample is similar to that of the full sample (see Appendix 3 in the Online Appendix).

4 Lofoten Scope Analysis

We use the CV survey data described above to conduct an *ex post* exploration of ten of the scope insensitivity factors reviewed in Part 2. In particular, we analyze the impacts of controlling for (1) diminishing marginal utility, (2) incomplete, multi-stage budgeting, (3) experience, familiarity, knowledge and/or use, (4) preference heterogeneity, (5) survey design, (6) amenity misspecification, and (7) warm glow. Furthermore, as part of our baseline estimation, we investigate the impact of (8) assumed statistical distribution by reporting both parametric and non-parametric results, (9) our data cleaning strategy, and (10) sub-sample sizes. Addressing the remaining issues would require additional experimental design modifications or the collection of multiple datasets.

4.1 Analytical Framework

Let X represent a proxy variable that controls for one source of scope insensitivity (e.g., amenity misspecification). We hypothesize that controlling for the scope confounding

---

[8] In other words, respondents were excluded if they *only* reported invalid reasons for voting yes/no to the referendum question. These include reasons indicating warm glow ("I answered yes because the amount was the size that my household tends to give for charitable purposes"; "My household is willing to pay for all good environmental purposes"), survey inconsequentiality ("I answered yes because I do not think the amount will be taxed in any way"; "What I say will not affect whether measures are implemented or not"; "I do not think there will be oil spill in this coastal area"; "I do not trust that the money will go for the right purpose"), or the following reasons: "I felt a commitment to pay because all other households should also contribute"; "It is the shipping companies and the shipping industry that should pay"; "The tax level is already high enough"; "I feel it is not right to value the environment in money"; "The question was too difficult to answer"; "Available public money can be reallocated or used more efficiently".

[9] We have tested whether recoding uncertain yes-respondents impacts the results. While the scope findings summarized in Sects. 4.2 and 4.3 remain unchanged, inferences from scope tests and elasticities are less clear-cut when using the original data.
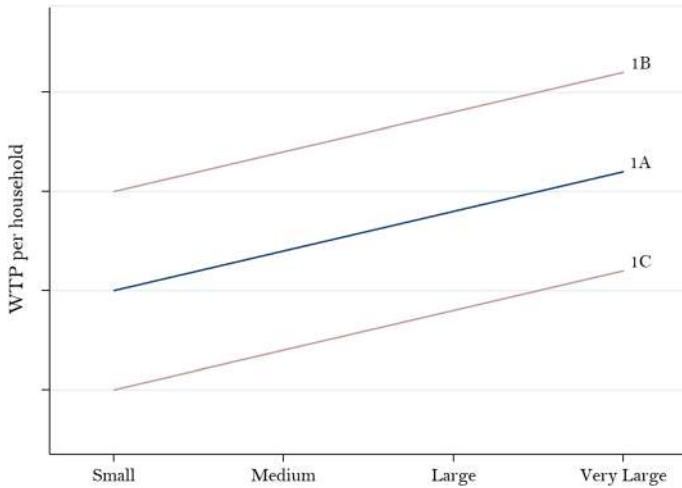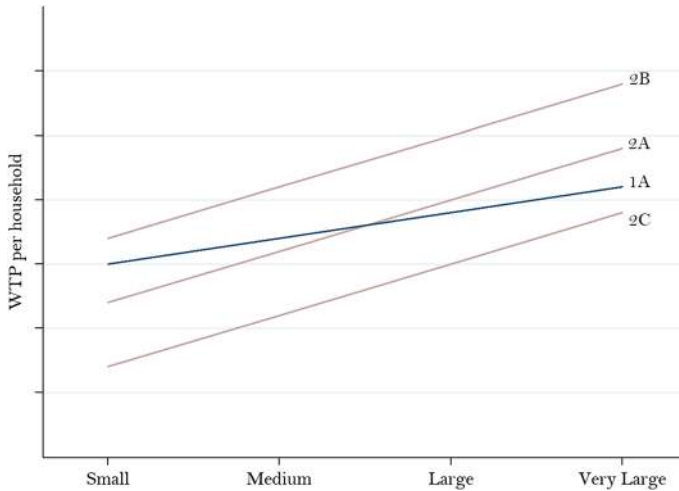
Fig. 1 Conceptualization of no impact on sensitivity to scope

explanation would increase the likelihood of establishing scope sensitivity and decrease the probability of false negatives, all else equal. Conversely, failing to control for this factor is expected to decrease the probability of finding scope effects and increase the chances of false negatives.

The variable X is interacted with dummy variables for the different oil spill sizes, which creates a piecewise linear model with structural breaks. The exception is the case of diminishing marginal utility, where we replace the dummy variables with a quasi-continuous proxy variable for the size of damage.

For each exploration, the estimated WTPs from an uncorrected baseline specification versus a "corrected" *alternative* specification are compared. The analysis is summarized graphically by comparing *scope lines*, which are linear interpolations of the four WTP estimates. All else held constant, a steeper scope line would suggest stronger scope impact. This idea is described conceptually in Figs. 1 and 2. Specifically, scope line 1A represents the uncorrected baseline specification, while 1B, 1C, 2A, 2B, and 2C represent possible corrected alternative specifications. In comparing 1A to either 1B or 1C (Fig. 1), one could say that mean WTP across the four oil spill sizes has changed, but controlling for the source of scope insensitivity does not seem to strengthen the scope inference by yielding a steeper scope line. In contrast, a move from scope line 1A to either 2A, 2B or 2C (Fig. 2) would be indicative of stronger scope sensitivity. For example, a move from 1A to 2B suggests that controlling for the source of scope insensitivity simultaneously strengthens the scope inference and leads to higher mean welfare estimates.

In our analysis below, we assess the presence or absence of sensitivity to scope in the following threefold way: First, we generate *empirical* scope lines for the baseline and alternative specifications, which are visually inspected for upward and monotonically increasing trends. Second, we execute two statistical scope tests on the WTP estimates. The *partial scope test* informs whether WTP to prevent the smallest oil spill is statistically different from WTP to prevent the largest oil spill. The *total scope test* reports statistical difference in WTP across all four oil spill sizes. Both tests are carried out by the *method of*

Fig. 2 Conceptualization of impact on sensitivity to scope

convolution and summarized with *p* values under the null hypothesis of insensitivity to scope.[10] The total scope convolution test is constructed as the average of the six possible partial convolution tests (comparing WTP for avoiding small versus medium oil spill, small versus large, etc.). Third, in order to address the question of adequate or plausible scope effects, we compute corresponding partial and total *scope arc-elasticities* (Whitehead 2016). In general, let WTP_1 and WTP_2 represent two estimates of WTP associated with two levels of oil spill prevention, Q_1 and Q_2, respectively. The scope arc-elasticity is then defined as: $E_{WTP,Q} = \left(\frac{WTP_2 - WTP_1}{Q_2 - Q_1} \right) \left(\frac{\bar{Q}}{\overline{WTP}} \right)$. The partial scope arc-elasticity is based on WTP estimates for the small and very large oil spills, whereas the total scope elasticity reflects the average of six possible scope arc-elasticities. In computing the denominator of the elasticity formula, we use kilometers of soiled coastline from the damage table (Appendix 2 in the Online Appendix) as a proxy variable.[11]

4.2 Baseline Results

Parametric and non-parametric baseline results are summarized in Figs. 3 and 4, respectively. The parametric estimation was carried out under an arbitrary assumption of normality following the direct WTP approach suggested by Cameron (1988). We use a piecewise linear functional form for the different oil spill sizes, with the smallest oil spill scenario as reference category. The parametric WTP estimates are plotted in Fig. 3. Regression outputs are provided in Appendix 4 in the Online Appendix.

---

[10] The method of convolution is a generalized test for statistical difference between two distributions. Poe et al. (2005) propose testing external scope by this method. In our adaptation, 5000 replications are generated for each WTP (Jeanty 2008).

[11] The scope elasticity results presented in Sect. 4.3 are robust with respect to the damage measure used in the denominator (i.e. kilometers of coastline affected, number of birds or number of seals dead).
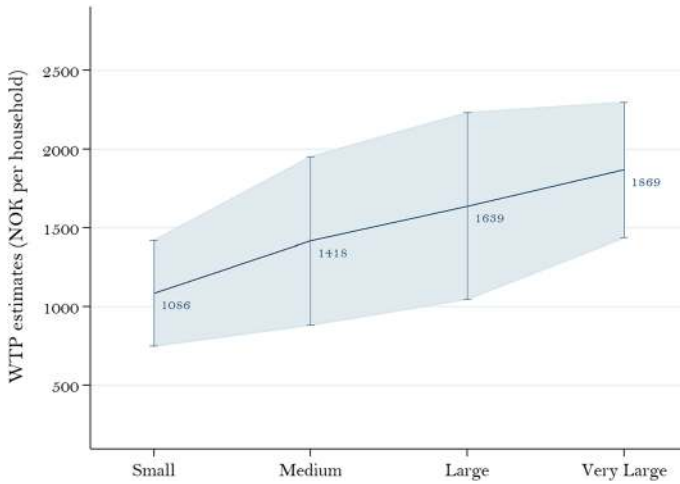
Fig. 3 Baseline parametric WTP welfare estimates (mean estimates and 95% confidence intervals)
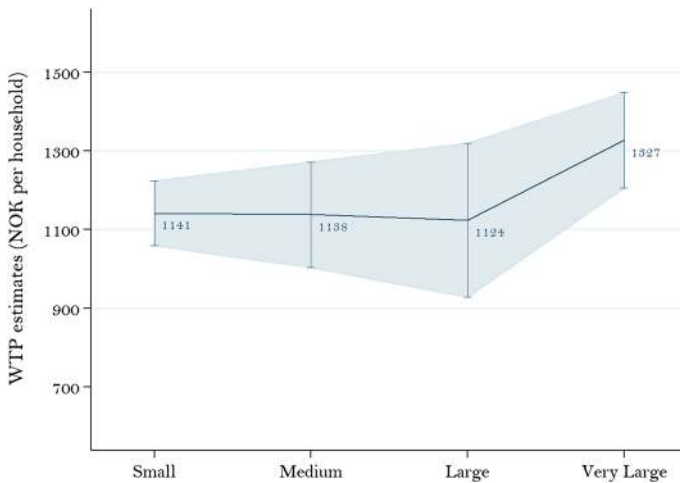

Fig. 4 Non-parametric welfare estimates (mean estimates and 95% confidence intervals)

The estimated annual household WTPs are NOK 1 086, 1 418, 1 639, and 1 869 to prevent a small, medium, large, and very large oil spill, respectively.[12] Estimated WTPs are almost identical to those reported for the same study site by Navrud et al. (2017). The corresponding empirical scope line is monotonically increasing. Furthermore, the 95% confidence intervals for the smallest and largest oil spills do not overlap, while other comparisons are overlapping. Second, formal convolution scope tests reported in the first row of Table 2 support the graphical analysis. The *p* value for partial scope is 0.0023, whereas the

---

Table 2 Baseline results and sensitivity analysis (scope convolution tests, scope arc-elasticities, and difference in WTP)

Confounding factor	Proxy variable(s)	Question from survey	P values for partial scope test	P values for total scope test	Partial scope arc-elasticity	Total scope arc-elasticity	P values for WTP diff.
Baseline estimations			0.0023	0.1446	0.27	0.18	
Diminishing marginal utility	Logarithm of Coastline affected in kilometers	Damage to coastline as described in damage table	0.0012	0.0709	0.27	0.18	0.92
Incomplete multi-stage budgeting	Dummy for income greater than 450 000 kr	Q33: "What would you estimate as your household's total gross disposable income per year?"	0.0216	0.2788	0.23	0.14	0.56
Experience, familiarity, knowledge and/or use	Dummy for previous experience with oil spills (in Norway or abroad)	Q18: "Have you ever seen damage from oil spills?"	0.0624	0.2683	0.28	0.17	0.14
	Dummy for past visit	Q8: "Have you ever been in the Vestfjord area in Nordland county?"	0.0318	0.2229	0.22	0.15	0.00
Preference Heterogeneity	Stated importance of avoiding environmental damages (scale 1–5)	Q19: "How important is each of these reasons for you? [Prevent long-term environmental damage]	0.0018	0.1453	0.25	0.17	0.00
Survey Design	Dummy for Consequentiality	Q21 and Q22 Responses: "I answered yes because I do not think the amount will be taxed in any way"; "What I say will not affect whether measures are implemented or not"; "I do not think there will be oil spill in this coastal area"; "I do not trust that the money will go for the 'right purpose'"	0.0035	0.1735	0.27	0.17	0.00
Amenity Misspecification	Perceived Probability (%) of oil spill happening	Q23: "How likely do you think that each of the four environmental damage levels will occur in Vestfjorden during the next 10 years?"	0.00	0.0921	0.41	0.30	0.07
Warm Glow	Dummy for Warm Glow	Q21 Responses: "I answered yes because the amount was the size that my household tends to give for charitable purposes"; "My household is willing to pay for all good environmental purposes"	0.0039	0.1587	0.38	0.27	0.00
Statistical Distribution Assumption	Non-parametric Estimator (Kriström)		0.00	0.1968	0.08	0.06	0.02

p value for total scope is 0.1446. Third, partial and total scope elasticities are also reported in Table 2. For the baseline parametric estimation, both elasticity estimates, 0.27 for partial and 0.18 for total scope, imply inelastic WTP with respect to the magnitude of oil spill damage. Nonetheless, these scope elasticity estimates are within the range of what has been discussed as adequate scope sensitivity in the literature (e.g., Amiran and Hagen 2010; Whitehead 2016). Combined, this threefold analysis suggests presence of partial, but not total, scope sensitivity.

Before presenting results from our main explorations, we briefly discuss our findings in relation to explanations 8–10 mentioned above: statistical distribution assumption, data cleaning strategies and sample size.

First, as is typical in CV studies (e.g., Carson et al. 2003), we also report non-parametric welfare estimates (see Fig. 4 and the last row of Table 2). This relates to the issue of statistical distribution assumption. In particular, we generate non-parametric WTPs using Kriström's method (Kriström 1990). This yields WTP estimates of NOK 1 141 to prevent the smallest oil spill and NOK 1327 to prevent the largest oil spill. As seen in Fig. 4, the non-parametric scope line is not monotonously increasing across the four oil spill scenarios. Similarly to the parametric baseline model, the non-parametric estimates pass the partial scope test (*p* value of 0.00) but not the total scope test (*p* value of 0.1968). The corresponding scope elasticities are smaller than the parametric ones at 0.08 for partial scope and 0.06 for total scope. Unlike the analysis in Borzykowski et al. (2018), the non-parametric approach yields weaker scope sensitivity in our case.

Second, as described in Sect. 3.2, we exclude 25% of the respondents from our analysis on the basis of their answers to debriefing questions. This relates to the issue of data cleaning strategies. If we re-estimate the baseline model by including these respondents, the *p* values of the partial and total statistical scope tests increase to 0.03 and 0.25, respectively. Moreover, the *p* values associated with specifications controlling for other scope confounding factors (reported in 4.3 below) also increase and the scope elasticities are lower. We thus conclude that excluding respondents based on the debriefing questions strengthens scope findings.

Third, sub-sample sizes for each oil spill scenario were set given budget constraints. Consequently, our quadruple split-sample CV experiment might not have enough power to find full scope sensitivity in the piecewise linear specification we employ.[13] This relates to the issue of sample size in scope sensitivity testing. Modest sub-sample sizes make it more difficult to pass statistical scope tests as pointed out by Arrow et al. (1994) and Rollins and Lyke (1998). Since the two extreme oil spill sizes (small vs. very large) have the largest sub-samples (N = 314 and N = 295, respectively), this is likely to be the comparison with the highest statistical power, while remaining comparisons are more likely to suffer from lack of power.[14] Consequently, there is a higher probability of passing a partial scope test than the total scope test.

---

[13] We thank the editor and an anonymous reviewer for drawing attention to the issue of statistical power.

[14] In general, the power of an experiment is the probability of finding a statistically significant effect for a given sample size (Duflo et al. 2007).
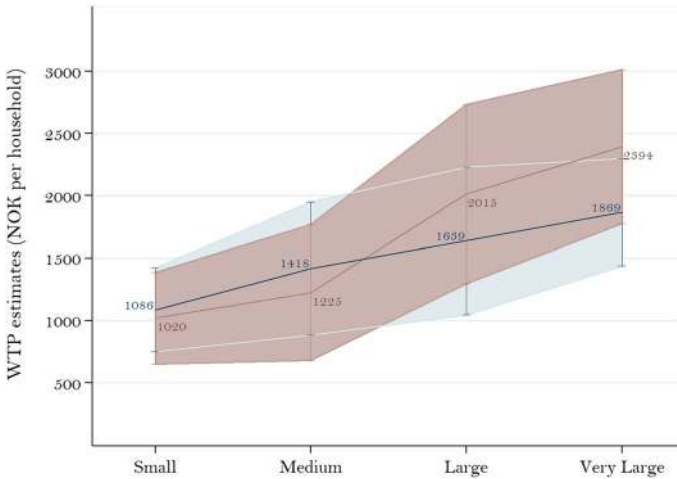
Fig. 5 Controlling for subjective oil spill probabilities (amenity misspecification)
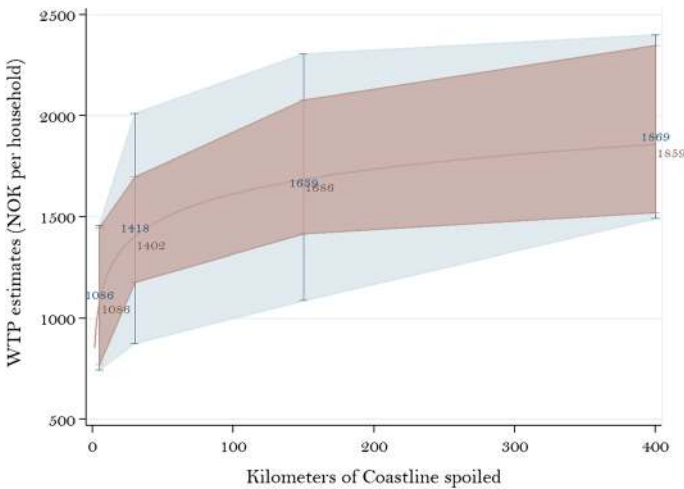

Fig. 6 Controlling for diminishing marginal utility

4.3 Sensitivity to Scope Analysis

Results from the scope diagnostics are presented in Figs. 5, 6, 7, 8, 9, 10, 11 and 12. Each figure represents *ceteris paribus* exploration of one scope insensitivity explanation and contains two estimated scope lines. The baseline scope line (blue) represents the parametric baseline results presented above.

Each scope line has a corresponding 95% confidence band. Confidence intervals are calculated using the Krinsky-Robb simulation (Krinsky and Robb 1986). The red confidence band is associated with the scope line from the estimation that controls for the potentially scope confounding factor, while the blue band belongs to the baseline scope line. Table 2
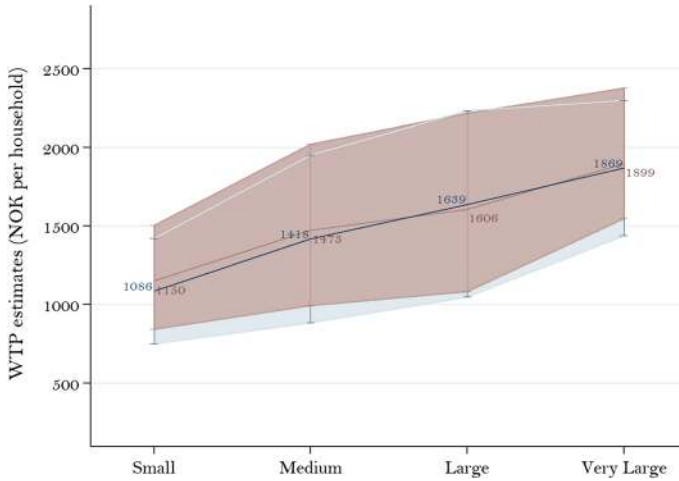
Fig. 7 Controlling for subjective importance of preventing oil spills (preference heterogeneity)
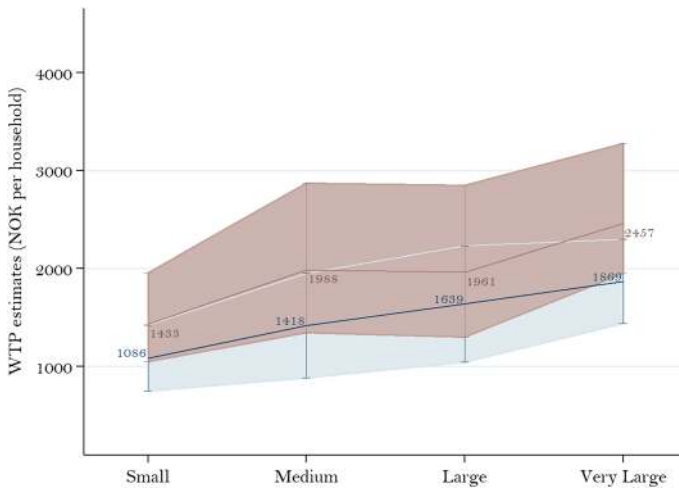
Fig. 8 Controlling for consequentiality (survey design)

describes how the control variable was constructed for each sensitivity analysis and reports corresponding convolution tests and scope elasticity estimates. Table 3 reports the WTP to prevent a small, medium, large or very large oil spill for the baseline estimation and for the various sensitivity analyzes.

In summary, accounting for amenity misspecification (Fig. 5), and imposing diminishing marginal utility (Fig. 6) have positive impacts on the scope inference. In contrast, accounting for preference heterogeneity (Fig. 7), consequentiality (Fig. 8), experience, familiarity, knowledge and/or use (Figs. 9, 10), controlling for incomplete multi-stage budgeting (Fig. 11), and warm glow (Fig. 12) do not appear to affect statistical scope inference in our case. Interestingly, several of the explorations imply differences in mean welfare estimates.
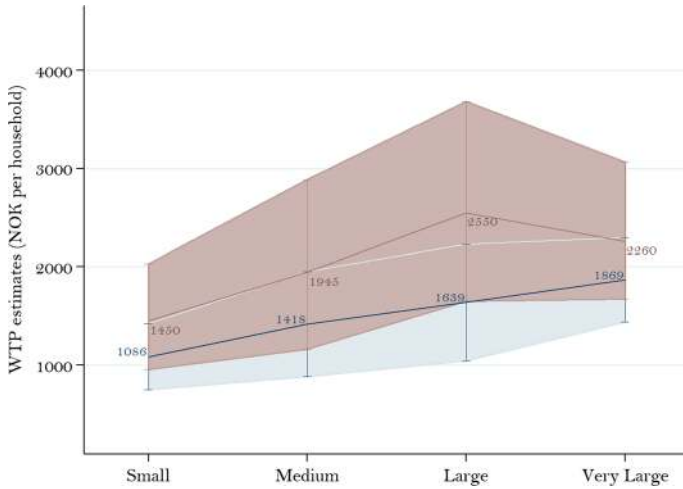
Fig. 9 Controlling for prior use of the Lofoten Archipelago (experience, familiarity, knowledge and/or use)
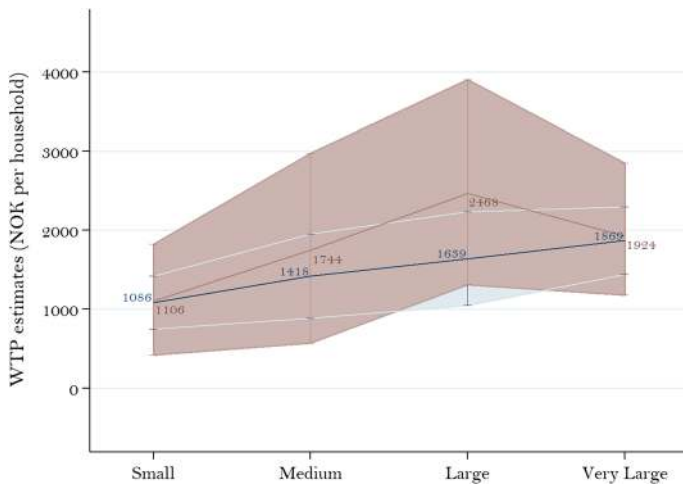

Fig. 10 Controlling for previous experience with oil spills (experience, familiarity, knowledge and/or use)

For example, controlling for perceived consequentiality (Fig. 8) and prior recreational use of the Lofoten Archipelago (Fig. 9) lead to higher WTP estimates. Next we discuss each exploration in further details. The underlying regression results are provided in Appendix 4 in the Online Appendix.

4.3.1 Controlling for Perceived Oil Spill Probabilities (Fig. 5)

As discussed in Part 2, the probability of provision bias is a type of amenity misspecification (Mitchell and Carson 1989). This issue may be of particular relevance in oil spill prevention studies, as participants are likely to bring subjective risk assessments into the
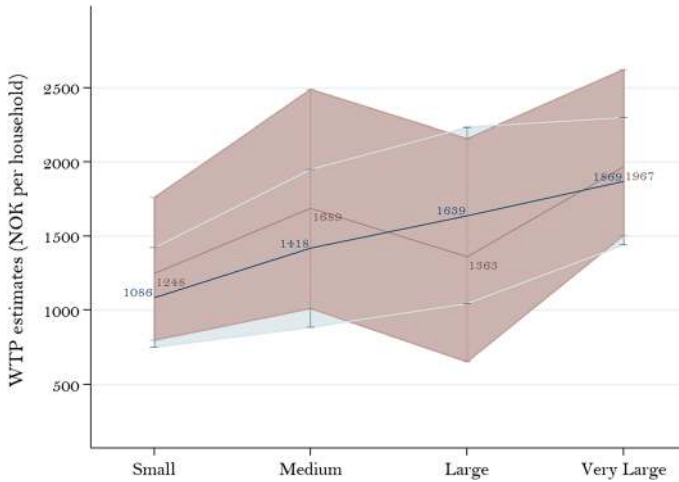
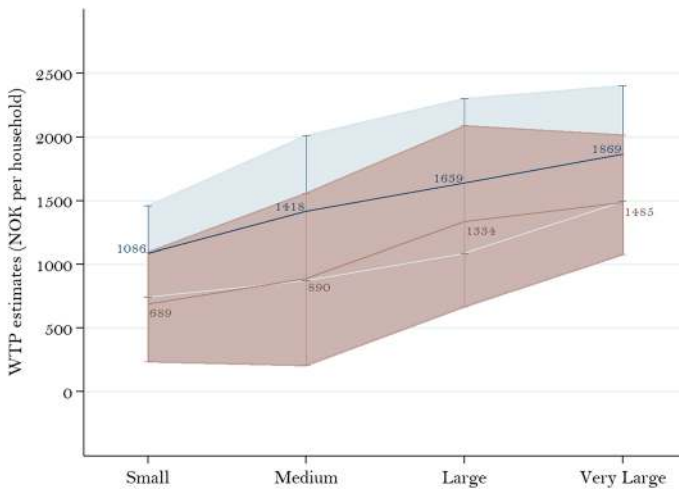**Fig. 11** Controlling for budget constraints

Fig. 12 Controlling for altruistic motivations (warm glow)

valuation exercise. In the Lofoten survey, the participants were told that an oil spill would happen with certainty, that is, with implied 100% probability, within the next 10 years. However, debriefing questions eliciting perceived probabilities of oil spills revealed that respondents considered larger oil spills less likely than smaller ones. The average perceived probability across all oil spill sizes was 0.39, while the average perceived probability was 0.54, 0.43, 0.33 and 0.27 for the small, medium, large and very large oil spill scenario, respectively.

We test whether such priors influence the scope inference by estimating a model that interacts perceived oil spill probabilities with the size dummies. The estimated WTP is subsequently computed at equalized (corrected) probabilities across different oil spill sizes.

Table 3 WTP estimates for baseline estimation and sensitivity analyzes

	Small	Medium	Large	Very large
Baseline estimations	1086	1418	1639	1869
	[750; 1421]	[885; 1951]	[1047; 2232]	[1440; 2298]
Diminishing marginal utility	1086	1402	1686	1859
	[770; 1444]	[1178; 1696]	[1417; 2076]	[1521; 2347]
Incomplete multi-stage budgeting	1248	1689	1363	1967
	[798; 1760]	[1012; 2489]	[654; 2156]	[1505; 2622]
Experience with oil spills	1106	1744	2468	1924
	[421; 1819]	[566; 2973]	[1307; 3905]	[1180; 2847]
Prior use of Lofoten Archipelago	1450	1945	2550	2260
	[957; 2030]	[1162; 2887]	[1651; 3685]	[1674; 3066]
Preference heterogeneity	1150	1473	1606	1899
	[842; 1503]	[993; 2022]	[1084; 2218]	[1547; 2378]
Survey design	1433	1988	1961	2457
	[1050; 1958]	[1346; 2871]	[1301; 2851]	[1950; 3277]
Amenity misspecification	1020	1225	2015	2394
	[652; 1389]	[681; 1769]	[1297; 2733]	[1776; 3013]
Warm glow	689	890	1334	1485
	[237; 1094]	[206; 1561]	[670; 2090]	[1077; 2015]
Statistical distribution assumption	1141	1138	1124	1327
	[1059; 1223]	[1004; 1272]	[929; 1319]	[1206; 1449]

Confidence intervals are calculated using the Krinsky–Robb simulation (Krinsky and Robb 1986)

Estimated WTP are in 2013 NOK per household per year. As of 18/02/2020: 1 Euro=NOK 10.1015; 1 USD=NOK 9.3324 (*Source*: https://www.bloomberg.com/markets/currencies)

As seen in Fig. 5, the estimated scope line is steeper in the corrected case. We conclude that controlling for amenity misspecification has a positive impact on scope findings. The visual observation is corroborated by the statistical scope tests and scope elasticities reported in Table 2. After correcting for the differences in perceived oil spill probabilities, the *p* value of the total scope test decreases from 0.1446 to 0.0921 and the total scope elasticity increases from 0.18 to 0.30.

4.3.2 Controlling for Diminishing Marginal Utility (Fig. 6)

The idea that diminishing marginal utility confounds sensitivity to scope has been proposed by many authors including Boyle et al. (1994) and Whitehead (2016). We explore this issue by converting the oil spill size dummies into a single, quasi-continuous variable denoting the logarithm of kilometers of soiled coastline. Using this variable as a damage proxy imposes diminishing marginal utility, rather than allowing for it through the piecewise linear specification. As seen in Appendix 2 in the Online Appendix, the small oil spill implies 5 km of coastline soiled, while the very large oil spill implies 400 km. The empirical scope line in red in Fig. 7 is produced from the specification with the logarithmic size variable as reported in Appendix 4 in the Online Appendix. The resulting WTP estimates are almost identical to those of the baseline model. Furthermore, all measures of fit improve with the logarithmic specification. Imposing diminishing marginal utility leads to smaller *p* values for the statistical scope tests, but does not change the scope elasticities. The *p* value of the total scope convolution test is 0.0709 (vs. 0.1446 for the baseline).

4.3.3 Controlling for Preference Heterogeneity (Fig. 7)

According to Siikamäki and Larson (2015), failure to account for unobserved preference heterogeneity can mask scope sensitivity. We explore the role of preference heterogeneity by accounting for stated importance of avoiding long-term environmental damage.[15] Figure 6 compares the scope line from the resulting model estimation with the baseline specification. The scope lines are almost identical, suggesting no impacts on sensitivity to scope nor overall willingness to pay. This is supported by the unchanged convolution tests and scope elasticities reported in Table 2. However, the p value of the partial scope test decreases (improves) slightly (from 0.0023 to 0.0018).

4.3.4 Controlling for Consequentiality (Fig. 8)

The idea that lack of consequentiality as a survey design issue may adversely affect scope findings was proposed by Carson and Groves (2007). If, for example, respondents believe that the probability of policy implementation or the probability of having to pay is zero, then preference expressions could be invariant to the size of the elicited good. To explore this issue, we compare results from an estimation that accounts for belief in the study's consequentiality with the baseline model. As observed in Fig. 8, it appears that accounting for consequentiality has no effect on scope sensitivity, though the scope line is no longer strictly increasing. The p values for both the partial and total scope tests reported in Table 2 are similar to those of the baseline. However, consequentiality does seem to have a positive effect on overall WTP (p value of 0.00).[16]

4.3.5 Controlling for Prior Use of the Lofoten Archipelago (Fig. 9)

Being a user rather than a non-user is an important dimension of how people relate to environmental goods and was mentioned by Whitehead et al. (1997) as a factor likely to influence scope sensitivity. We explore this hypothesis by interacting the size dummies with an indicator for use of the Lofoten Archipelago, defined as having visited or residing there. As can been seen in Fig. 9, respondents classified as users have higher WTP for all oil spill sizes. However, the scope lines do not indicate a clear difference in scope sensitivity. The statistical tests reported in Table 2 corroborate these observations. A partial scope finding is retained for both segments, while total scope is not supported. The p value for a difference in overall WTP is 0.00.

4.3.6 Controlling for Previous Experience with Oil Spill (Fig. 10)

Experience with oil spills is another case-specific dimension of how people relate to the environmental good. Heberlein et al. (2005) hypothesize that having previous experience might lead to clearer scope sensitive findings. Figure 10 compares estimation results for respondents with prior experience with oil spills relative to the baseline model. As

---

[15] We would like to thank the editor and an anonymous reviewer for their input regarding how to test for preference heterogeneity. We estimated a random parameter logit and latent class models without getting further insights.

[16] We also tested alternative measures of consequentiality using other information from the debriefing questions. The results were similar to those reported here.

indicated by the test statistics in Table 2, this does not improve the scope inference nor lead to statistically different WTP estimates.

4.3.7 Controlling for Incomplete Multi-stage Budgeting (Fig. 11)

Short-term restrictions on how households allocate their income could impact scope sensitivity according to Randall and Hoehn (1996). As a way to test for this kind of confounding factor, we create a dummy variable identifying households with yearly household income lower than 450 000 NOK (that is, the 25th percentile of the income distribution). We hypothesize that respondents less constrained by income would be more sensitive to scope. However, as seen in Fig. 11 and Table 2, we only find an impact on WTP, not on sensitivity to scope. This result is similar to the finding in Randall and Hoehn (1996).

4.3.8 Controlling for Warm Glow (Fig. 12)

Kahneman and Knetsch (1992) claim that WTP estimates are mainly driven by warm glow preferences. This suggests that accounting for warm glow could yield clearer scope findings. To explore this possibility, we identified respondents indicating such motivation from the debriefing questions. As illustrated by the exploration in Fig. 12 and corroborated by the test statistics in Table 2, controlling for warm-glow preferences seems to negatively affect overall WTP (p value of 0.00). However, the scope inference is ambiguous. One the one hand, the p values of the statistical scope tests do not improve relative to the baseline model. On the other hand, the scope elasticity estimates are higher at 0.38 for partial scope and 0.27 for total scope.

5 Concluding Remarks

In this paper, we make two primary contributions to the environmental CV literature. First, we give an overview of the sensitivity to scope issue and review a number of theoretical and empirical explanations for why WTP estimates sometimes are found to be insensitive to the scope of the environmental good being valued. Then we investigate the validity of many of these explanations in the context of valuing the prevention of oil spills in Arctic Norway.

The literature review uncovers 13 distinct explanations for insensitivity to scope in CV studies. These are placed in four categories according to whether they relate to (1) *microeconomic consumer theory*, (2) *how people relate to the environmental good*, (3) *survey design and model estimation*, *or* (4) *insights from behavioral economics*. The literature analysis answers a repeated call for an overview of scope-confounding factors (Carson and Mitchell 1995; Whitehead et al. 1998; Desvousges et al. 2012; Whitehead 2016). Despite being suggested by several researchers, few studies have actually carried out explorations of scope-confounding factors in specific case analyzes. The second part of the paper addresses this research gap by testing empirically a subset of the explanations proposed in the literature.

Based on our review, it is clear that failing statistical scope tests do not invalidate single studies, certainly not the CV method in general. At least thirteen factors can mask scope sensitivity. Some of these factors can be controlled for *ex post* (e.g., degree of experience with the environmental good), while others are best dealt with *ex ante* (e.g., ensuring

incentive-compatible and consequential survey instruments). Nonetheless, some *ex ante* considerations must be made to ensure validity of a study. Namely, if survey design and/or amenity specification are not adequate and the study does not follow best practices, then the statistical scope test is likely to fail and WTP estimates are not valid. Therefore, we strongly advise researchers to contextualize empirically how one or more of the reviewed explanations may affect the scope inference of their study.

Our baseline estimation indicates *partial* scope sensitivity, defined as a statistically significant difference in WTP for avoiding the smallest versus the largest oil spill. The estimated WTP to prevent the smallest and largest oil spills are NOK 1 086 and NOK 1 869, respectively. Accounting for scope-confounding factors strengthens the scope inference. These include: excluding problematic respondents based on debriefing questions, taking sample sizes into consideration, accounting for subjective probabilities of amenity provision, and imposing diminishing marginal utility. Controlling for the last two factors improve the inference from partial to *total* external scope at the 10% significance level, the latter defined as an overall difference in WTP across oil spill sizes.

Furthermore, scope elasticity estimates indicate presence of economically significant, adequate effects. In the baseline specification, the partial and total scope elasticities are 0.27 and 0.18, respectively. Our scope diagnostics show that controlling for confounding factors generally leads to higher scope elasticity estimates, with the highest estimates found in the specification accounting for amenity misspecification. Here, the partial and total scope elasticities are 0.41 and 0.30, respectively. While the literature has not reached consensus with respect to what constitutes adequate scope, we judge a scope elasticity estimate of 0.2 to be of adequate and *plausible* magnitude. Such estimate indicates inelastic WTPs and conforms to the explanation of diminishing marginal utility from avoiding damages to environmental goods. This magnitude is also in line with scope elasticity estimates reported by Whitehead (2016), Spencer-Cotton et al. (2018) for the case of coastal areas in Australia, and by Borzykowski et al. (2018) for the case of protected forest areas in Switzerland.

Overall WTP for oil spill prevention in the Lofoten Archipelago seems to be inelastic with respect to the damage size. This observation is consistent with the notion of sharply diminishing marginal utility for preventing oil spills in Arctic areas. The Norwegian population views Lofoten as an exceptional coastal area when it comes to natural and cultural amenities (Kaltenborn and Linnell 2019). Therefore, exposing it to any kind of non-trivial industrial accident such as an oil spill could be seen as fundamentally damaging: once the Lofoten Archipelago is *soiled*, its non-market economic value is *spoiled*—the size of the oil spill may not matter so much.

Finally, the set of scope insensitivity explanations addressed in this paper is not necessarily exhaustive. Other factors are likely to emerge from current or future studies, perhaps particularly related to research in behavioral economics. Furthermore, the empirical observations made in this paper regarding factors that influence the scope inference do not necessarily generalize or carry over to other study contexts. Nonetheless, this paper lends support to the sentiment expressed by several other authors (Arrow et al. 1994; Heberlein et al. 2005; Banerjee and Murphy 2005; Amiran and Hagen 2010; Desvousges et al. 2012; Whitehead 2016: Johnston et al. 2017), to wit, that standard statistical scope tests can be uninformative and potentially misleading if taken at face value. We therefore advise CV practitioners to pursue their own case-specific scope sensitivity diagnostics using our review as a starting point. We also advise future CV applications to emphasize whether their scope findings are adequate and/or plausible by computing and reporting scope elasticities and other effect size measures.

Compliance with Ethical Standards

References

Alberini A (1995) Testing willingness-to-pay models of discrete choice contingent valuation survey data. Land Econ 71(1):83–95

Alevy J, List J, Adamowicz W (2011) How can behavioral economics inform nonmarket valuation? An example from the preference reversal literature. Land Econ 87(3):365–381

Amiran E, Hagen D (2010) The scope trials: variation in sensitivity to scope and WTP with directionally bounded utility functions. J Environ Econ Manag 59(3):293–301

Arrow K, Solow R, Portney PR et al (1993) Report of the NOAA panel on contingent valuation. Fed Reg 58(10):4601–4614

Arrow K, Leamer E, Schuman H et al (1994) Appendix D in "Comments on proposed NOAA/DOI regulations on natural resource damage assessment". U.S. Environmental Protection Agency, New York

Banerjee S, Murphy JH (2005) The scope test revisited. Appl Econ Lett 12(10):613–617

Barton D, Navrud S, Siripala N et al (2003) Procedures for economic valuation of resources and accounting of oil spill damages. Institutional strengthening of the oil spill contingency management project. Marine Pollution Prevention Authority, Sri Lanka

Bateman IJ, Cole M, Cooper P et al (2004) On visible choice sets and scope sensitivity. J Environ Econ Manag 47(1):71–93

Berrens RP, Bohara AK, Silva CL et al (2000) Contingent values for New Mexico instream flows: with tests of scope, group-size reminder and temporal reliability. J Environ Manag 58(1):73–90

Bishop RC, Boyle KJ, Carson RT et al (2017) Putting a value on injuries to natural assets: the BP oil spill. Science 356(6335):253–254

Borenstein M, Hedges LV, Higgins JP et al (2011) Introduction to meta-analysis. Wiley, New York

Borzykowski N, Baranzini A, Maradan D (2018) Scope effects in contingent valuation: Does the assumed statistical distribution of WTP matter? Ecol Econ 144:319–329

Boyle KJ, Desvousges WH, Johnson FR et al (1994) An investigation of part-whole biases in contingent-valuation studies. J Environ Econ Manag 27(1):64–83

Burrows J, Newman R, Genser J et al (2017) Do contingent valuation estimates of willingness to pay for non-use environmental goods pass the scope test with adequacy? A review of the evidence from empirical studies in the literature. In: McFadden D et al (eds) Contingent valuation of environmental goods. Edward Elgar Publishing, New York

Cameron TA (1988) A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. J Environ Econ Manag 15(3):355–379

Carson RT (1997) Contingent valuation surveys and tests of insensitivity to scope. In: Kopp RJ (ed) Determining the value of non-marketed goods. Springer, Dordrecht, pp 127–163

Carson RT (2012) Contingent valuation: a comprehensive bibliography and history. Edward Elgar Publishing, New York

Carson RT, Groves T (2007) Incentive and informational properties of preference questions. Environ Resour Econ 37(1):181–210

Carson RT, Mitchell RC (1995) Sequencing and nesting in contingent valuation surveys. J Environ Econ Manag 28(2):155–173

Carson RT, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence. Environ Resour Econ 19(2):173–210

Carson RT, Mitchell RC, Hanemann M et al (2003) Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. Environ Resour Econ 25(3):257–286

Carson RT, Conaway MB, Hanemann WM et al (2013) Valuing oil spill prevention: a case study of California's central coast, vol 5. Springer, Berlin

Casey JF, Kahn JR, Rivas AA (2008) Willingness to accept compensation for the environmental risks of oil transport on the Amazon: a choice modeling experiment. Ecol Econ 67(4):552–559

Champ PA, Moore R, Bishop RC (2009) A comparison of approaches to mitigate hypothetical bias. Agric Resour Econ Rev 38(2):166–180

Czajkowski M, Hanley N (2009) Using labels to investigate scope effects in stated preference methods. Environ Resour Econ 4(4):521

Desvousges WH, Johnson FR, Dunford RW et al (1992) Measuring nonuse damages using contingent valuation: an experimental evaluation of accuracy. Research Triangle Park, NC

Desvousges W, Mathews K, Train K (2012) Adequate responsiveness to scope in contingent valuation. Ecol Econ 84:121–128

Diamond P (1996) Testing the internal consistency of contingent valuation surveys. J Environ Econ Manag 30(3):337–347

Diamond PA, Hausman JA (1994) Contingent valuation: Is some number better than no number? J Econ Perspect 8(4):45–64

Duflo E, Glennerster R, Kremer M (2007) Using randomization in development economics research: a toolkit. Discussion Paper Series 6059

Frederick S, Fischhoff B (1998) Scope (in) sensitivity in elicited valuations. Risk Decis Policy 3(2):109–123

Freeman AM III, Herriges JA, Kling CL (2014) The measurement of environmental and resource values: theory and methods. Routledge, London

Giguere C, Moore C, Whitehead JC (2020) Valuing hemlock woolly adelgid control in public forests: scope effects with attribute nonattendance. Land Econ 96(1):25–42

Haab TC, McConnell KE (1997) Referendum models and negative willingness to pay: alternative solutions. J Environ Econ Manag 32(2):251–270

Haab TC, Interis MG, Petrolia DR et al (2013) From hopeless to curious? Thoughts on Hausman's "dubious to hopeless" critique of contingent valuation. Appl Econ Perspect Policy 35(4):593–612

Hausman J (2012) Contingent valuation: from dubious to hopeless. J Econ Perspect 26(4):43–56

Heberlein TA, Wilson MA, Bishop RC et al (2005) Rethinking the scope test as a criterion for validity in contingent valuation. J Environ Econ Manag 50(1):1–22

Jeanty PW (2008) WTPCIKR: Stata module to estimate Krinsky and Robb confidence intervals for mean and median willingness to pay. Statistical software components S456965, Boston College Department of Economics

Johnston RJ, Boyle KJ, Adamowicz W et al (2017) Contemporary guidance for stated preference studies. J Assoc Environ Resour Econ 4(2):319–405

Kahneman D, Knetsch JL (1992) Valuing public goods: the purchase of moral satisfaction. J Environ Econ Manag 22(1):57–70

Kaltenborn BP, Linnell JD (2019) Who are legitimate stakeholders? National and local perceptions of environmental change in the Lofoten islands, Norway. Polar Geogr 42(4):236–252

Kling CL, Phaneuf DJ, Zhao J (2012) From Exxon to BP: Has some number become better than no number? J Econ Perspect 26(4):3–26

Kontoleon A, Swanson T (2003) The willingness to pay for property rights for the giant panda: Can a charismatic species be an instrument for nature conservation? Land Econ 79(4):483–499

Krinsky I, Robb AL (1986) On approximating the statistical properties of elasticities. Rev Econ Stat 68:715–719

Kriström B (1990) A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. Land Econ 66(2):135–139

Loomis JB (2014) 2013 WAEA keynote address: strategies for overcoming hypothetical bias in stated preference surveys. J Agric Resour Econ 39(1):34–46

Loureiro ML, Loomis JB, Vázquez MX (2009) Economic valuation of environmental damages due to the Prestige oil spill in Spain. Environ Resour Econ 44(4):537–553

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory, vol 1. Oxford University Press, New York

Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resour Fut

Navrud S, Lindhjem H, Magnussen K (2017) Valuing marine ecosystem services loss from oil spills for use in cost-benefit analysis of preventive measures. Handb Econ Manag Sustain Oceans 124–137

Ojea E, Loureiro ML (2011) Identifying the scope effect on a meta-analysis of biodiversity valuation studies. Resour Energy Econ 33(3):706–724

Poe GL (2016) Behavioral anomalies in contingent values and actual choices. Agric Resour Econ Rev 45(2):246–269

Poe GL, Giraud KL, Loomis JB (2005) Computational methods for measuring the difference of empirical distributions. Am J Agric Econ 87(2):353–365

Randall A, Hoehn JP (1996) Embedding in market demand systems. J Environ Econ Manag 30(3):369–380

Rollins K, Lyke A (1998) The case for diminishing marginal existence values. J Environ Econ Manag 36(3):324–344

Rowe RD, Shaw WD, Schulze W (1992) Nestucca oil spill. In: Ward KM, Duffield JW (eds) Natural resource damages: law and economics. Wiley, New York, pp 527–554

Siikamäki J, Larson DM (2015) Finding sensitivity to scope in nonmarket valuation. J Appl Econom 30(2):333–334

Smith VK, Osborne LL (1996) Do contingent valuation estimates pass a "scope" test? A meta-analysis. J Environ Econ Manag 31(3):287–301

Søgaard R, Lindholt J, Gyrd-Hansen D (2012) Insensitivity to scope in contingent valuation studies. Appl Health Econ Health Policy 10(6):397–405

Spencer-Cotton A, Kragt ME, Burton M (2018) Spatial and scope effects: valuations of coastal management practices. J Agric Econ 69(3):833–851

Urminsky O, Kivetz R (2011) Scope insensitivity and the "mere token" effect. J Mark Res 48(2):282–295

Van Biervliet K, Le Roy D, Nunes PA (2006) An accidental oil spill along the Belgian Coast: results from a CV Study

Varian HR (2014) Intermediate microeconomics: a modern approach. WW Norton & Company, New York

Whitehead JC (2016) Plausible responsiveness to scope in contingent valuation. Ecol Econ 128:17–22

Whitehead JC, Haab TC, Huang JC (1997) Part-whole bias in contingent valuation: Will scope effects be detected with inexpensive survey methods? Working Paper

Whitehead JC, Haab TC, Huang JC (1998) Part-whole bias in contingent valuation: Will scope effects be detected with inexpensive survey methods? South Econ J 65(1):160–168