Essays on the Validity and Reliability of Non-Market Valuation Methods

by

Ana Faria Lopes

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

PhD programme in Social Sciences
UiS Business School
2020

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

ISBN: 978-82-7644-939-6
ISSN: 1890-1387
PhD: Thesis UiS No. 532

Acknowledgments

This research was funded by the Norwegian Research Council in the context of the COAST-BENEFIT project (project number 255777). The financial support from the Norwegian Research Council is gratefully acknowledged.

Doing a PhD in Norway was a "triple threat". Besides the already challenging experience of doing a PhD, I struggled with my identity as an expat, creating a new social network as well as learning a new language. Nonetheless, the PhD road was an enriching experience and I developed an immense toolkit while feeding my inquisitive and restless mind.

I feel tremendously blessed to have had the opportunity to work and study at the University of Stavanger. All the support I received from the institution and my peers allowed me to build a scientific network and acquire specific skills which I would not have been able to if I had done my PhD anywhere else.

I am extremely grateful for all the support, help and openness of my supervisor Gorm Kipperberg. It was rewarding but also fun to knock on his door to bounce off research ideas or just to vent about obstacles in the PhD research. He taught me many valuable lessons that helped me grow into a more mature researcher.

I would like to extent some thankful words to my co-advisor, Henrik Lindhjem, opponent for the 90% seminar, Maria Loureiro, and co-authors, Petr Mariel and John Whitehead. Thanks to John's invitation, I was able to spend an extremely rewarding semester at the Appalachian State University, where I was surrounded by new ideas and peers.

I would also like to thank my colleagues at the University of Stavanger for their support, friendliness and interesting conversations at the lunch table. A special thank you to my PhD colleagues for keeping my sane throughout the PhD.

Living far from family is a struggle, especially from a meddlesome yet loving family that I am blessed to have. I want to thank my family for all the video calls, chaotic Christmas dinners and words of encouragement I have received for the last three years.

And the biggest thank you to Thomas, who turned out to be the most important support I did not know I needed

Summary

This dissertation makes important contributions to the literature of non-market valuation methods. A vast bulk of the literature has focused on ensuring the validity and reliability of non-market valuation methods, but many challenges remain. Throughout the dissertation, I focus on three of these challenges: 1) addressing high multicollinearity in revealed preference data, 2) providing an answer to the scope insensitivity phenomenon, and 3) tackling misspecification when estimating revealed or stated preference data. I address these challenges in order to obtain both valid and reliable estimates of welfare change. More importantly, my dissertation shows that non-market valuation methods are themselves both valid and reliable.

Table of Contents

Chapter 3 - Essay 2: “Diagnosing Insensitivity to Scope in Contingent Valuation”

Chapter 4 - Essay 3: “Estimating the Ex-ante Recreational Loss of an Oil Spill using Revealed Preference Site Selection and Multinomial Stated Preference Data”

1 Introduction

Throughout this thesis, I estimate the economic value of environmental goods and services (hereafter referred to simply as environmental goods). Examples of environmental goods include recreational sites, clean water or natural resources. Examples of environmental services include carbon sequestration, recreation possibilities or biodiversity conservation. The availability of these goods affects the well-being of the individuals who depend on and use them. Hence changes in their associated quality or quantity will impact individual welfare.

The common framework for the above-mentioned concepts of economic value, welfare and well-being is utilitarian economic theory. In his seminal work published in “The Wealth of Nations”, Adam Smith (1776) recognized the duality of the concept of value as both referring to the intrinsic value of a good and its market price. In utility theory, the intrinsic value of a good for an individual is the total utility or welfare derived from it. Hence, total utility (i.e., value of a good) is the sum of the good’s market price and its excess value, referred to as “relative utility” by Dupuit (1844), or more commonly known as “consumer surplus” (Marshall, 1920). Hence, the market price of a good is not the true measure of its value, but it is rather a lower-bound estimate (Dupuit, 1844). The true measure of the value of a good is the maximum price an individual would agree to pay given a budget constraint (Dupuit, 1844; Vatin et al., 2016).

In many cases, however, there are no readily available prices nor estimates of the lower-bound values of environmental goods. For some environmental goods, prices in the market economy are absent as these goods are not transacted in the same manner as market goods. The reason for this is the public and common good nature of environmental goods, which relate to absence of property rights.

Policy makers need to allocate limited resources in the most efficient manner for society. Many policy decisions involve changing the quantity or quality of environmental or market goods available to society.

However, the two cannot be fairly compared if the price for environmental goods is absent, thus increasing uncertainty for policy making, and ultimately an inefficient allocation of resources compared with the socially optimal one (Bradshaw and Borchers, 2000; Knights et al., 2014). Therefore, providing estimates of the change in value following a change in environmental good provision may improve policy making.

While obtaining these estimates may be useful for policy makers, these should be both valid and reliable. Ensuring the validity of estimates means these should be unbiased, while ensuring the reliability means minimizing the variation of the estimates, i.e., improving precision. Relevant methods for estimating value change are generally known as non-market valuation methods. The literature on these methods focuses to a great extent on ensuring the validity and reliability of value change estimates. Several authors show that failing to account for various factors related to individual preferences may result in invalid or unreliable value estimates (e.g., Lew and Wallmo, 2017; Li et al., 2015). Therefore, a fundamental question is as follows: *How can we ensure that environmental value estimates are valid and reliable*?

The overarching goal of this dissertation is to contribute to our understanding of the validity and reliability of non-market valuation methods. In particular, I address three methodological challenges that have complicated the validity and reliability of non-market valuation methods: 1) addressing high multicollinearity in revealed preference data; 2) providing an answer to the scope insensitivity phenomenon, and 3) avoiding misspecification when separately estimating revealed preference (RP) or stated preference (SP) data. I address these challenges in the following essays:

- Essay 1: “Estimating the economic benefits of coastal quality change: An Application to Beach Recreation in Norway”
- Essay 2: “Diagnosing Insensitivity to Scope in Contingent Valuation”

- Essay 3: "Estimating the Ex-ante Recreational Loss of an Oil Spill using Revealed Preference Site Selection and Multinomial Stated Preference Data"

Essay 1 improves reliability of RP methods by ensuring proper identification of welfare estimates using simulation prior to estimation. Essay 2 investigates validity in SP methods, by showing that failure to pass a scope test does not imply invalidity of SP methods. Finally, Essay 3 addresses validity of both SP and RP methods by illustrating the validity gains of combining different data sources.

The remainder of this chapter is structured as follows. Section 2 summarizes the conceptual framework for non-market valuation. Section 3 presents the methodological problem and then situates the three essays within the current debate on the validity and reliability of non-market valuation methods. Section 4 summarizes the policy problem and this dissertation's policy contributions. Section 5 concludes.

2 Conceptual Framework

The ultimate goal of the non-market valuation researcher is to estimate the change in welfare associated with an increase or decrease in quantity or quality of environmental goods. Welfare estimates can be expressed in terms of willingness to pay (WTP) or willingness to accept (WTA). The following discussion uses the WTP concept, but the implications can be extended to the WTA measure.

The researcher can obtain welfare estimates by observing people's choices in markets or surveys to understand their preferences (Champ et al., 2003, p. 2). By observing the choices people make when deciding amongst market and non-market goods, the researcher can analyze the implicit trade-offs and estimate WTP (Bishop and Boyle, 2019).

Microeconomic consumer theory provides the foundation for studying individual choices and preferences. Assume that an individual has preferences towards a bundle of market goods, denoted by $\boldsymbol{X}$, and a bundle of non-market goods, denoted by $\boldsymbol{Q}$. Hence, I can represent the individual's utility function as dependent on the individual's endowment of market and non-market goods:

$$U = U(\boldsymbol{X}, \boldsymbol{Q}). \tag{1}$$

Utility is assumed to increase in the quantity or quality of the market and non-market goods, that is $U_X > 0$ and $U_Q > 0$, respectively. This utility function exists so long as the preference relation between market and non-market goods is continuous (Mas-Colell et al., 1995). Continuity does not hold if, for example, the good became a bad at some threshold level.

Let U_0 represent an initial level of utility. An indifference curve represents all possible combinations of $\boldsymbol{X}$ and $\boldsymbol{Q}$ that yield the same utility level (U_0) for the individual. Figure 1 illustrates the indifference curve that yields utility U_0 and an initial endowment level ($\boldsymbol{X}_0, \boldsymbol{Q}_0$).

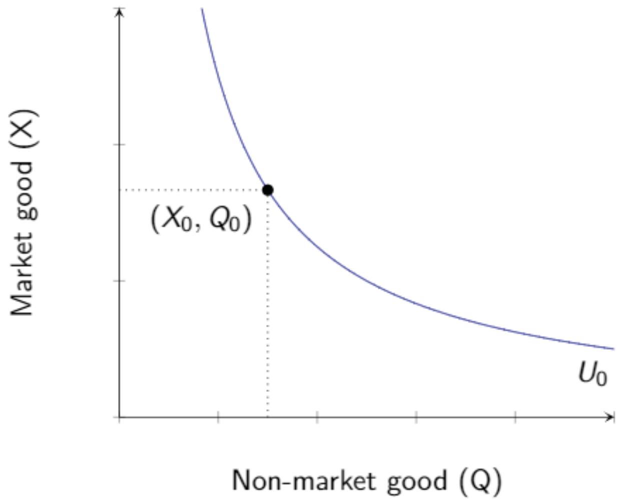


Figure 1 – Indifference Curve for the individual

An individual would be indifferent to being at the initial endowment level $(\boldsymbol{X}_0, \boldsymbol{Q}_0)$ or any other point along the line U_0. This means that the individual would be willing to trade-off some of his/her endowment of market goods to obtain a slightly higher quantity of the non-market goods. Such a movement would be a move along the indifference curve to the right. The amount of market goods that the individual would be willing to give up to obtain a marginal increase in the quantity of non-market goods is defined in microeconomics as the marginal rate of substitution (MRS), in this case $MRS_{\boldsymbol{X},\boldsymbol{Q}}$ (Mas-Colell et al., 1995).

Assume that the bundle of market goods, denoted by $\boldsymbol{X}$, is a numeraire market good expressed in money terms. If so, the MRS is interpreted as the money amount an individual would be willing to give up to obtain a marginal increase in the quantity of the non-market good. This is the marginal WTP, that is, the measure of welfare change I want to estimate.

Even though the individual would be willing to trade-off the market and non-market good, (s)he is not able to decide how much to consume of the non-market good. The public agent (e.g., State) determines the

supply of non-market goods, due to their nature as public goods (Champ et al., 2003, p. 28).
Let us assume that the endowment of the non-market good is exogenously increased from $\boldsymbol{Q}_0$ to $\boldsymbol{Q}_1$. Because $U_{\boldsymbol{Q}} > 0$ and the initial endowment of market goods remains unchanged at $\boldsymbol{X}_0$, the individual's utility will increase to U_1, hence the current situation is now on a different indifference curve. In Figure 2, I illustrate how the individual's situation changes from the initial endowment $(X_0, \boldsymbol{Q}_0)$ to the final endowment $(X_0, \boldsymbol{Q}_1)$, which is a point on a higher indifference curve.

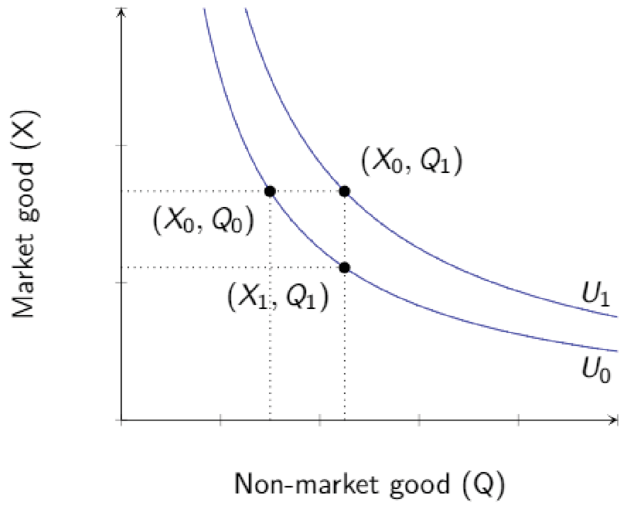


Figure 2 – Increase in exogenous supply of non-market good

One can analyse along the initial indifference curve U_0 to find the amount of money the individual would be willing to pay that would put him/her back at the initial utility level. Because the numeraire market good is expressed in money terms, the distance $X_0 - X_1$ is interpreted as the maximum amount of money the individual would be willing to pay to obtain this quantity of nonmarket good, i.e. his/her WTP. The corresponding welfare measure is compensating surplus.[1]

Given the utility function in (1), the researcher formalizes the willingness to pay with the following expression:

$$U(X_0, Q_0) = U(X_0 - WTP, Q_1). \qquad (2)$$

Hence, to obtain the WTP for a higher quantity or quality of non-market goods, such as environmental goods, the researcher needs to model a utility function that depends both on the quantity of non-market good and the composite money good.

[1] If one would compare two points at the final utility level, the corresponding welfare measure would be equivalent surplus. The corresponding expression in Equation 2 would be $U(X_0, Q_1) = U(X_0 + WTP, Q_0)$.

3 Methodological Problems & Contributions

This dissertation makes important contributions to the literature of validity and reliability of non-market valuation methods. In this Section, I introduce the two concepts of validity and reliability, and explain how these have been addressed in the non-market valuation literature. I then summarize the contributions of this dissertation in light of the methodological problems.

3.1 Methodological Problems

Non-market valuation methods include two types: revealed and stated preference methods. Revealed preference (RP) methods use primary or secondary market data to analyze how observed choices reveal the individuals' implicit preferences towards changes in environmental quality or quantity (Freeman et al., 2014). Stated preference (SP) methods involve asking individuals in carefully constructed and hypothetical scenarios about their willingness to trade-off money for environmental goods (Freeman et al., 2014). In my dissertation, Essay 1 applies RP methods, Essay 2 applies SP methods, and Essay 3 combines both methods.

Let j denote an application of non-market valuation methods characterized by one sample drawn from the population of interest and a set of procedures. In application j the researcher obtains an estimate of WTP, denoted by WTP_j. However, WTP_j is not the same concept as the true, albeit latent WTP as described in the conceptual framework (Equation 2). Bishop and Boyle (2019) formalized the relationship between the WTP estimate (WTP_j) and true WTP as:

$$WTP_j = WTP + e_j, \tag{3}$$

wherein e_j is an error term. While it is unrealistic to expect that $e_j = 0$, in which case the WTP estimate would be equal to the true WTP, it is essential that $E(e_j) = 0$. That is, if the study was replicated an infinite number of times, the expected value of WTP estimates should be the same as the true value. If so, WTP_j is an unbiased, that is, valid estimate of WTP. Ensuring the validity of the WTP estimates is the primary concern of the researcher by selecting the most appropriate methods, assumptions and procedures when conducting a study.

Individual applications of non-market valuation methods can be assessed to ensure the resulting WTP is valid. Since researchers cannot observe true WTP, there is no measure with which to compare WTP_j in order to ensure validity (Hoyos and Mariel, 2010). Instead, to assess a study's validity, one can consider three criteria: construct, content and criterion validity (Bishop and Boyle, 2019). Content validity testing implies making sure the procedures followed in the study comply with best practices. Construct validity testing involves understanding whether results conform to expectations from theory. Criterion validity testing implies comparing the WTP estimates obtained using different methods, especially if one method is accepted as having a higher level of validity.

Various authors have contributed towards ensuring the validity of non-market valuation methods. Many factors may lead to biased WTP estimates in RP studies, for example choice set misspecification (Li et al., 2015), assuming independence across choice occasions (English, 2010), not accounting for multiple purpose trips (Bin et al., 2007), the naïve inclusion of endogenous variables such as congestion (Bujosa et al., 2015; Bujosa, 2010; Hindsley et al., 2007), not allowing for substitution across activities (Cutter et al., 2007), or measurement error in the opportunity cost of leisure time (Czajkowski et al., 2019). In SP studies, past research has focused on obtaining unbiased WTP estimates when, for example, facing attribute non-attendance (Colombo et al., 2013), defining the available choice sets as perceived by respondents (DeShazo et al., 2009), investigating income effects (e.g., Jacobsen and Hanley, 2009), or solving the disparity between WTP and WTA measures (e.g., Tunçel and Hammitt, 2014). However, many threats to validity remain unresolved.

A second concern of the researcher is to ensure reliability of WTP estimates. Reliability concerns minimizing the variation of the error term rather than its bias. In other words, to obtain a reliable measure of WTP, the researcher should minimize the standard error of its estimate. While some of the dispersion of the standard deviation arises due to the natural variation within the sample, the researcher can minimize the standard deviation by choosing the appropriate methods, assumptions and procedures.

Bishop and Boyle (2019) argue that assessing a study's reliability involves inferring how all of the study's steps influence the magnitude of the error term, including all econometric assumptions. They suggest replicating the study and comparing WTP_j and WTP_{j+1}, wherein $j+1$ refers to the replication study.

One can replicate a study by surveying either the same individuals or a random sample drawn from the same population (Rakotonarivo et al., 2016). Tests of reliability have focused mainly on testing the temporal stability of preferences by administering the same study at two or more points in time. Studies focusing on temporal stability of preferences include Lew and Wallmo (2017) and Schaafsma et al. (2014) for SP data, and Mkwara et al. (2015) and Parsons and Stefanova (2009) for RP data. Reliability may also be tested by assessing the effect (if any) of slightly different survey designs (Rakotonarivo et al., 2016).

In conclusion, researchers strive to ensure their WTP estimates are valid and reliable by choosing the most appropriate methods, procedures and assumptions in each context. To this end, many guidelines and examples of the best practices are available for RP and SP methods. Examples of recent state-of-the-art applications include English et al. (2018) for RP methods and Bishop et al. (2017) for SP methods. The NOAA panel developed a series of guidelines regarding the application of SP methods (Arrow et al., 1993), recently revised by Johnston et al.

(2017), that most studies take into account. Regarding RP methods, however, no such detailed guidelines exist.

3.2 Methodological Contributions

This dissertation specifically tackles three challenges associated with RP and SP data: 1) ensuring reliability when using RP data and facing multicollinearity; 2) ensuring validity when using SP data due to insensitivity to scope; 3) ensuring validity by combining RP and SP methods. This dissertation makes important contributions to the literature on non-market valuation methods by addressing both their reliability and validity.

Essay 1 focuses on the reliability of RP methods. While it is typically tested by replicating studies, the underlying cause of a study's unreliability may be the data itself. In that case, infinite replications of the same study will not improve its reliability. Instead, the nature of the data may already hint at whether resulting WTP estimates are reliable or not, namely, if high multicollinearity is present in the data. Multicollinearity refers to the case of "high (but not perfect) correlation between two or more independent variables" (Wooldridge, 2016). If high multicollinearity is present in the data, the WTP estimate may have such a large standard deviation that the corresponding confidence interval is too broad, resulting in a statistically insignificant, hence unreliable WTP estimate.

The phenomenon of multicollinearity is common when handling RP or observed data (Adamowicz et al., 1994; Ben-Akiva et al., 2002; Earnhart, 2002; Whitehead et al., 2012, p. 2), especially if the environmental goods available are rather homogenous. For example, two recreational sites may not differ sufficiently in terms of their observable characteristics if they are physically close to each other. An additional challenge is the potential lack of variation in the RP data, due to all alternatives being too similar. Given these disadvantages, one would consider discarding RP data altogether. However, RP data does have its advantages: these represent observed past, rather than stated choices. In Essay 1, I use RP data and show that my data does indeed suffer from lack of variation and high multicollinearity, which complicates the proper identification of the parameters needed to estimate WTP.

Two strategies have been proposed to solve the problem of identification in RP data: combine RP with SP data to break the multicollinearity (von Haefen and Phaneuf, 2008), or ensure proper identification by using Murdock (2006)'s two-stage strategy. However, I do not have access to SP data on the attributes of interest, nor enough alternatives to apply Murdock (2006)'s strategy. Instead, I propose using simulation to tackle identification prior to estimation. I do this in three steps: 1) define possible population parameters and error term assumptions; 2) given step 1, find the alternative that yields the highest utility (i.e., the choice); and 3) predict the choice given the data. The parameters retrieved in step 3 should be the same as those defined in step 1. I thus choose a functional form for the utility function (Equation 1) that reduces the multicollinearity in the data by avoiding highly correlated explanatory variables. I then estimate WTP_j for these environmental quality variables. The proposed solution to the identification problem expands the toolkit of practitioners that wish to explain observed choices among similar goods with few alternatives (less than 30).

The remaining two Essays (Essays 2 and 3) focus on validity. While RP methods are generally considered to yield valid estimates of WTP, many researchers have questioned the validity of SP methods (e.g., Diamond and Hausman, 1994). More recently, Hausman (2012) and McFadden and Train (2017) argue that the contingent valuation method (one of the SP methods) provides flawed measures of WTP. If the contingent valuation method does have a persistent error outside of the researcher's control, as they argue, then $E(e_j) \neq 0$ and the method should be discarded. I show in Essay 2, however, that one of the arguments used by Hausman (2012) and McFadden and Train (2017), that is scope insensitivity, is not a method-specific error and can be accounted for by the researcher.

The existence of a method-specific error hinges on the premise that SP methods occasionally fail construct validity tests. Specifically, one can test for construct validity by ascertaining whether results conform to expectations from economic theory. For example, one would expect that an individual should always prefer a higher quantity of an environmental good over a smaller quantity. This should translate into higher WTP estimates for larger sizes of Q. Such a property is known in the non-market valuation literature as scope sensitivity. However, when applying the contingent valuation method, this property does not always hold. For example, Boyle et al. (1994) found that survey respondents were indifferent between the prevention of 2000, 20,000 or 200,000 bird deaths. In Essay 2, I am confronted with the same artifact in the baseline analysis, as the WTP_{Q_0} is not statistically different from WTP_{Q_1}, where $Q_1 > Q_0$.

However, many explanations that also conform to economic theory have been proposed to explain this phenomenon. These include diminishing marginal utility (Rollins and Lyke, 1998) and poor survey design (Carson and Mitchell, 1995). Given the plenitude of explanations put forth in previous research, several authors have called for a thorough review of the various explanations that may confound scope (Carson and Mitchell, 1995; Desvousges et al., 2012; Heberlein et al., 2005; Whitehead et al., 1998; Whitehead, 2016).

Essay 2 answers this need for a review by focusing on the scope issue: what are the various reasons previously identified in the literature that may lead to scope insensitivity? How does each of these reasons affect scope findings in an empirical example? In Essay 2, I first identify 13 different reasons proposed in the last 40 years as to why insensitivity to scope occurs. I then use data to analyze how controlling for these reasons affects my findings. I find four reasons out of the thirteen that lead to more plausible scope in my empirical application. I conclude that scope insensitivity is not a sufficient reason for deeming a study nor a method invalid, as there are multiple explanations for false negatives. Essay 2 has implications beyond environmental valuation, as scope insensitivity is an artifact that occurs in other fields such as marketing or health economics that also use SP methods. I propose that practitioners perform their scope diagnostics when facing scope insensitivity, using the review as a helping guide.

The validity of non-market valuation methods may also be assessed using criterion validity tests. These involve comparing WTP estimates using different methods, for example, using RP (Essay 1) and SP (Essay 2). Since SP methods imply the construction of hypothetical markets wherein respondents state their preferences, it is useful to compare these with observed (RP) outcomes.

However, it is often the case that RP and SP methods do not yield the same WTP. Moreover, using either RP or SP data may not allow the researcher to fully capture the value of environmental damage. Instead of using the two datasets separately, some authors suggest combining them (Whitehead et al., 2008). The data are complementary: SP data are hypothetical, while RP is based on observed choices; RP data suffers from lack of variation and high multicollinearity in attribute data, while SP data can be experimentally varied in the survey. The combination of the RP and SP data generally results in a better fit for the models.

Nonetheless, the researcher should pay attention to whether the data should be combined in the first place. If RP and SP data elicit distinct preferences, then the data may not be combined. Even if RP and SP data come from the same underlying preferences, the researcher should still account for the possibility of scale parameters differing across datasets, namely if SP scenarios are less familiar to respondents and their stated choices are more random (Huang et al., 1997).

In Essay 3, I combine actual (RP) and hypothetical (SP) data to estimate the WTP given a change in environmental quality Q. I face a scenario where there is a simultaneous drop in environmental quality and the number of available alternatives for the consumer (i.e., reduction in the choice set). Capturing a reduction in quality is possible by using SP data, while the loss associated with a reduced choice set is possible by using RP data. I argue that jointly estimating RP and SP data has important validity gains. Estimating RP or SP models separately leads to misspecification since not all relevant parameters can be identified. Moreover, combining RP and SP data has several advantages in terms of welfare analysis: using only RP data leads to low welfare losses, while using only SP data leads to high welfare losses.

Moreover, past combinations of RP and SP data have oversimplified the hypothetical scenarios when eliciting stated choices. In the surveys, respondents are given the options of visiting one or more recreational sites within the same study area (e.g. Truong et al., 2018); opting-out i.e., "staying at home" (Yi and Herriges, 2017); or postponing the trip (Parsons and Stefanova, 2011). In reality, other options may exist, such as going to a different recreational site or engaging in different recreational activities. Omitting available options when modeling choices can result in biased measures of welfare (Stafford, 2018). I allow for more flexible patterns of substitution by presenting an alternative formulation of the hypothetical scenario.

To sum up, in Essay 3, I make two significant contributions to the state-of-the-art of RP and SP data combinations. I first illustrate why it is important to combine RP and SP to jointly simulate a reduction in environmental quality and reduction of the choice set, as well as why using a single data source leads to biased estimates of welfare loss. Secondly, I propose a formulation of a hypothetical scenario that allows for broader patterns of substitution.

4 Policy Problem & Contributions

Quantifying the value of changes in the provision of environmental goods is especially relevant when considering public policies. A relevant example is the case of oil drilling in Northern Norway. While the benefits of allowing for oil drilling can be inferred by looking at the value added to the oil industry in national accounts, the costs due to loss of welfare for the Norwegian population are less tangible. Losses may arise due to, for example, changes in welfare from knowing pristine ecosystems are damaged.

Changes in environmental quality or quantity imply both use and non-use value changes for the economic agents affected. Use values suggest some direct or indirect human interaction with the environment (Barbier et al., 2011), and non-use values (also known as existence or passive-use values) refer to cases in which people assign value to an environmental good despite not using it directly or indirectly. Motivations for non-use values include the "mere existence" of an environmental good (Krutilla, 1967), bequest, altruistic reasons or maintaining a future use-option (Millennium Ecosystem Assessment, 2005). RP methods are tailored to estimate use value estimates, while SP methods can estimate changes in both use and non-use values (Eom and Larson, 2006; Perman et al., 2003).

All three essays estimate the value of coastal environmental goods in Norway. In Essays 1 and 3, the environmental good in question is beach recreation, while Essay 2 focuses on ocean conservation (i.e., preventing an oil spill accident). Essays 1 and 3 estimate changes in use values, while Essay 2 focuses on changes in both use and non-use values. Two study areas are considered: the Lofoten archipelago in Northern Norway (Essay 2) and the *Jæren* coast in Southwestern Norway (Essays 1 and 3). Both study areas are illustrated in Figure 3.

.

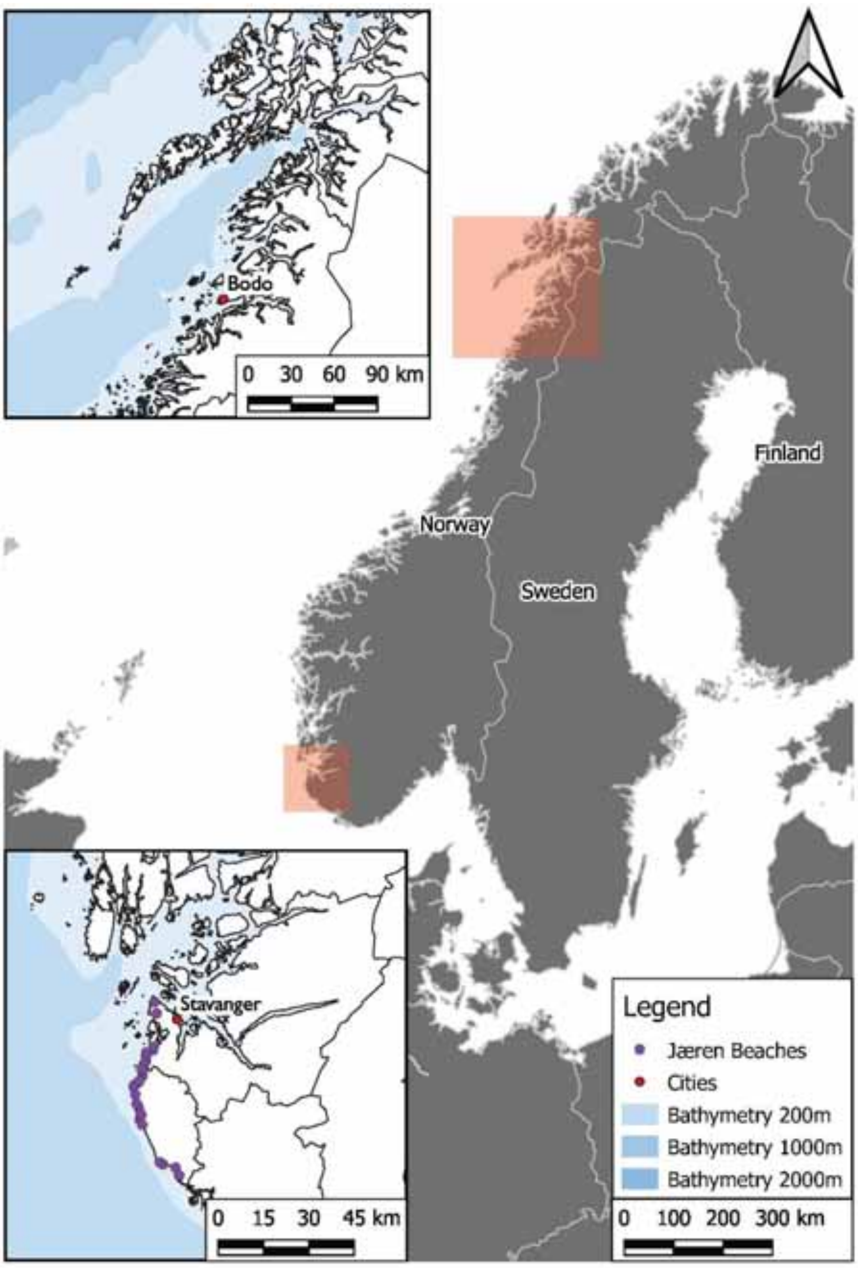


Figure 3 – Map illustrating the two study areas considered in the dissertation

Each of the three essays contributes to coastal policy in distinct ways. The estimated WTP pertains to changes in beach quality in Essay 1 and the lost value due to an oil spill accident in Essays 2 and 3. The aggregate values estimated in each paper are summarized in Table 1.

Table 1 – Overview of Policy Contributions (Aggregate Value Changes)

	Type of Data	**Environmental good**	**Policy Focus**	**Population affected**	**Aggregate Value estimates**
Essay 1	RP	Beach recreation	Recreational gains due to beach facility improvements	Rogaland (county) residents	+38.6 million NOK
			Recreational losses due to dune erosion		-98.9 million NOK
			Recreational gains due to bus route creation		+30.9 million NOK
Essay 2	SP	Oil Spill Prevention	Total value loss due to oil spill	Norwegian population	-2.6 billion NOK (Small)
					-4.5 billion NOK (Very Large)
Essay 3	RP&SP	Beach recreation	Recreational losses due to oil spill	Rogaland (county) residents	-368 million NOK (Small)
					-718 million NOK (Very Large)

Essay 1 estimates the WTP for changes in beach quality. Three beach quality scenarios are considered: improvement in beach facilities (parking and toilet facilities), loss of dunes due to erosion, and the creation of a bus route. I find that the loss of dunes generates a welfare loss of 13.6 Norwegian kroner per visit and group, while the creation of a bus route increases recreational welfare by 4.3 Norwegian kroner per visit and group. These estimates of recreational benefits may be useful for policy makers to assess whether beach quality improvements should be implemented by comparing its costs and benefits. In Essay 1, I illustrate that improving beach facilities generates 38.6 million Norwegian kroner in recreational benefits, which exceeds by seven the estimated investment costs of 5 million kroner. In the case of the erosion of dunes, aggregate recreational losses are in the order of 98.9 million Norwegian kroner. Finally, the recreational gains regarding the creation of a bus route are estimated to be 30.9 million Norwegian kroner.

The remaining two essays focus on the welfare loss due to an oil spill accident. If an oil spill were to occur, Norwegian society would incur substantial losses in both use and non-use values. If the magnitude of the costs due to a specific oil spill is known before the accident, policymakers may assess the cost-effectiveness of implementing preventive measures to avoid the oil spill and associated welfare losses.

Essay 3 focuses on the use losses (recreational impact) of an oil spill in the *Jæren* coast. Heavier ship traffic along the *Jæren* coast in Norway increases the risk of a ship grounding and, consequently, an oil spill occurring in the area. While no oil spill accident was recorded in the Jæren coast, since 2011, the Norwegian Maritime Authority has recorded a total of 132 cargo ship accidents in the jurisdiction of the county. I find a welfare loss of 123 Norwegian kroner per visit solely attributed to recreational (use) value losses. I estimate WTP to avoid three other oil spill scenarios: 188, 262 and 289 Norwegian kroner to prevent a medium, large and very large oil spill. This corresponds to aggregate recreational losses in the order of 368 million NOK in the case of a small oil spill, which would increase to 718 million NOK in case of a very large oil spill.

In Essay 2, I estimate the total loss due to an oil spill (i.e., both use and non-use value losses). Opening the area near the Lofoten islands for oil exploration is being considered by the Norwegian government. The area should only be open for oil exploration if the societal benefits exceed the costs. An increase in oil production will likely lead to an increase in ship traffic, thus increasing the risk of an oil spill. If an oil spill were to occur, the welfare for the Norwegian population would decrease, mainly due to losses in non-use value. I estimate the WTP to implement oil spill preventive measures in the Lofoten islands per Norwegian household. 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 spills, respectively. When considering 2.4 million Norwegian households, this amounts to an aggregate welfare loss ranging from 2.6 billion in the case of a small oil spill to 4.5 billion Norwegian kroner due to a very large oil spill.

The value estimates in question when changing environmental quality are relatively large, in the order of million or billion Norwegian kroner. This dissertation shows that not only do individuals have preferences towards the provision of environmental goods, but their economic value is also substantial.

5 Conclusions

This dissertation makes important contributions to the literature of non-market valuation methods. A vast bulk of the literature has focused on ensuring the validity and reliability of non-market valuation methods, but many challenges remain. Throughout the dissertation, I address three important challenges related to the validity and reliability of these methods. There are as follows: 1) addressing high multicollinearity in RP data, 2) providing an answer to the scope insensitivity phenomenon in SP data, and 3) tackling misspecification when estimating revealed or preference data. Table 2 summarizes the three essays included in this dissertation.

Table 2 – Overview of the Three Essays in Dissertation

	Type of Data	**Methodological Focus**	**Authorship**	**Status**
Essay 1	RP	Reliability	Lopes & Mariel	In review (Coastal Management)
Essay 2	SP	Validity	Lopes & Kipperberg	Accepted for publication (Environmental and Resource Economics)
Essay 3	RP&SP	Validity	Lopes & Whitehead	Manuscript

In my dissertation, I show how to tackle these challenges in order to obtain both valid and reliable estimates of welfare change. More importantly, my dissertation shows that non-market valuation methods are themselves both valid and reliable. Nonetheless, practitioners should not overlook the challenges associated with the application of these methods. While this dissertation makes important contributions towards the validity of SP and RP methods and reliability of RP methods, many important challenges remain.

From a policy standpoint, I show how substantial the value of environmental goods is for Norwegian society. Changes in the quality or quantity of environmental goods may impact social welfare in the order of millions or billions of Norwegian kroner. To avoid any losses in welfare from materializing, policymakers should recognize the value of existing environmental goods and promote their preservation.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. J. Environ. Econ. Manag. 26, 271–292. https://doi.org/10.1006/jeem.1994.1017

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

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R., 2011. The value of estuarine and coastal ecosystem services. Ecol. Monogr. 81, 169–193.

Ben-Akiva, M., McFadden, D., Morikawa, T., 2002. Discrete choice models incorporating revealed preferences and psychometric data, in: Advances in Econometrics, Advances in Econometrics. Emerald Group Publishing Limited, pp. 29–55. https://doi.org/10.1016/S0731-9053(02)16003-8

Bin, O., Dumas, C., Poulter, B., Whitehead, J.C., 2007. Measuring the impacts of climate change on North Carolina coastal resources, Final Rep. for National Commission on Energy Policy.

Bishop, R.C., Boyle, K.J., 2019. Reliability and Validity in Nonmarket Valuation. Environ. Resour. Econ. 72, 559–582. https://doi.org/10.1007/s10640-017-0215-7

Bishop, R.C., Boyle, K.J., Carson, R.T., Chapman, D., Hanemann, W., Kanninen, B., Kopp, R.J., Krosnick, J.A., List, J., Meade, N., Paterson, R., Presser, S., Smith, V.K., Tourangeau, R., Welsh, M., Wooldridge, J.M., DeBell, M., Donovan, C., Konopka, M., Scherer, N., 2017. Putting a value on injuries to natural assets: The BP oil spill. Science 356, 253–254. https://doi.org/10.1126/science.aam8124

Boyle, K.J., Desvousges, W.H., Johnson, F.R., Dunford, R.W., Hudson, S.P., 1994. An Investigation of Part-Whole Biases in Contingent-Valuation Studies. J. Environ. Econ. Manag. 27, 64–83. https://doi.org/10.1006/jeem.1994.1026

Bradshaw, G.A., Borchers, J., 2000. Uncertainty as Information: Narrowing the Science-policy Gap. Conserv. Ecol. 4. https://doi.org/10.5751/ES-00174-040107

Bujosa, A., Riera, A., Hicks, R.L., McConnell, K.E., 2015. Densities Rather than Shares: Improving the Measurement of Congestion in Recreation Demand Models. Environ. Resour. Econ. 61, 127–140. https://doi.org/10.1007/s10640-014-9785-9

Bujosa, A.R., 2010. Congestion and site choice: beach recreation in the Mediterranean. Presented at the AERE Inaugural Summer Conference, Seattle.

Carson, R.T., Mitchell, R.C., 1995. Sequencing and Nesting in Contingent Valuation Surveys. J. Environ. Econ. Manag. 28, 155–173. https://doi.org/10.1006/jeem.1995.1011

Champ, P.A., Boyle, K.J., Brown, T.C., Peterson, L.G., 2003. A primer on nonmarket valuation. Springer.

Colombo, S., Christie, M., Hanley, N., 2013. What are the consequences of ignoring attributes in choice experiments? Implications for ecosystem service valuation. Ecol. Econ. 96, 25–35.

Cutter, W.B., Pendleton, L., Deshazo, J.R., 2007. Activities in Models of Recreational Demand. Land Econ. 83, 370–381.

Czajkowski, M., Giergiczny, M., Kronenberg, J., Englin, J., 2019. The Individual Travel Cost Method with Consumer-Specific Values of Travel Time Savings. Environ. Resour. Econ. 74, 961–984. https://doi.org/10.1007/s10640-019-00355-6

DeShazo, J.R., Cameron, T.A., Saenz, M., 2009. The Effect of Consumers' Real-World Choice Sets on Inferences from Stated Preference Surveys. Environ. Resour. Econ. 42, 319–343. https://doi.org/10.1007/s10640-008-9250-8

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

Diamond, P.A., Hausman, J.A., 1994. Contingent Valuation: Is Some Number Better than No Number? J. Econ. Perspect. 8, 45–64. https://doi.org/10.1257/jep.8.4.45

Dupuit, J., 1844. On the measurement of the utility of public works. Int. Econ. Pap. 2, 83–110.

Earnhart, D., 2002. Combining Revealed and Stated Data to Examine Housing Decisions Using Discrete Choice Analysis. J. Urban Econ. 51, 143–169. https://doi.org/10.1006/juec.2001.2241

English, E., 2010. Discrete Choice, Recreation Demand, And Consumer Surplus. Cornell University.

English, E., von Haefen, R.H., Herriges, J., Leggett, C., Lupi, F., McConnell, K., Welsh, M., Domanski, A., Meade, N., 2018. Estimating the value of lost recreation days from the Deepwater Horizon oil spill. J. Environ. Econ. Manag. 91, 26–45. https://doi.org/10.1016/j.jeem.2018.06.010

Eom, Y.-S., Larson, D.M., 2006. Improving environmental valuation estimates through consistent use of revealed and stated preference information. J. Environ. Econ. Manag. 52, 501–516. https://doi.org/10.1016/j.jeem.2005.11.001

Freeman, A.M., Herriges, J.A., Kling, C.L., 2014. The measurement of environmental and resource values: theory and methods, 3rd ed. ed. Resources for the Future, Washington, DC.

Hausman, J., 2012. Contingent Valuation: From Dubious to Hopeless. J. Econ. Perspect. 26, 43–56. https://doi.org/10.1257/jep.26.4.43

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

Hindsley, P., Landry, C., Bin, O., Vogelsong, H., 2007. Site congestion in recreation choice models: A generated regressors approach to beach site selection.

Hoyos, D., Mariel, P., 2010. Contingent valuation: Past, present and future. Prague Econ. Pap. 4, 329–343.

Jacobsen, J.B., Hanley, N., 2009. Are There Income Effects on Global Willingness to Pay for Biodiversity Conservation? Environ. Resour. Econ. 43, 137–160. https://doi.org/10.1007/s10640-008-9226-8

Johnston, R.J., Boyle, K.J., Adamowicz, W. (Vic), Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C.A., 2017. Contemporary Guidance for Stated Preference Studies. J. Assoc. Environ. Resour. Econ. 4, 319–405. https://doi.org/10.1086/691697

Knights, A.M., Culhane, F., Hussain, S.S., Papadopoulou, K.N., Piet, G.J., Raakær, J., Rogers, S.I., Robinson, L.A., 2014. A step-wise process of decision-making under uncertainty when implementing environmental policy. Environ. Sci. Policy 39, 56–64. https://doi.org/10.1016/j.envsci.2014.02.010

Krutilla, J.V., 1967. Conservation Reconsidered. Am. Econ. Rev. 57, 777–786.

Lew, D.K., Wallmo, K., 2017. Temporal stability of stated preferences for endangered species protection from choice experiments. Ecol. Econ. 131, 87–97. https://doi.org/10.1016/j.ecolecon.2016.08.009

Li, L., Adamowicz, W., Swait, J., 2015. The effect of choice set misspecification on welfare measures in random utility models. Resour. Energy Econ. 42, 71–92. https://doi.org/10.1016/j.reseneeco.2015.07.001

Marshall, A., 1920. Principles of economics. Lond. Mcmillan.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic theory. Oxford university press New York.

McFadden, D., Train, K., 2017. Contingent Valuation of Environmental Goods: A Comprehensive Critique. Edward Elgar Pub, Cheltenham, UK.

Millennium Ecosystem Assessment, M.E.A., 2005. Ecosystems and human well-being. Synthesis.

Mkwara, L., Marsh, D., Scarpa, R., 2015. Testing the stability of welfare estimates in travel cost random utility models of recreation: an application to the Rotorua Lakes, New Zealand.

Murdock, J., 2006. Handling unobserved site characteristics in random utility models of recreation demand. J. Environ. Econ. Manag. 51, 1–25. https://doi.org/10.1016/j.jeem.2005.04.003

Parsons, G., Stefanova, S., 2009. Temporal stability of recreation choices. Presented at the 2009 Annual meeting, July, pp. 26–28.

Parsons, G.R., Stefanova, S., 2011. Gauging the Value of Short-Term Site Closures in a Travel-Cost RUM Model of Recreation Demand With a Little Help from Stated Preference Data, in: Preference Data for Environmental Valuation: Combing Revealed and Stated Preference Approaches, Routledge Explorations in Environmental Economics. Routledge, Taylor & Frances Group.

Perman, R., Ma, Y., McGilvray, J., Common, M., 2003. Natural resource and environmental economics. Pearson Education.

Rakotonarivo, O.S., Schaafsma, M., Hockley, N., 2016. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. J. Environ. Manage. 183, 98–109. https://doi.org/10.1016/j.jenvman.2016.08.032

Rollins, K., Lyke, A., 1998. The Case for Diminishing Marginal Existence Values. J. Environ. Econ. Manag. 36, 324–344. https://doi.org/10.1006/jeem.1998.1045

Schaafsma, M., Brouwer, R., Liekens, I., De Nocker, L., 2014. Temporal stability of preferences and willingness to pay for natural areas in choice experiments: A test–retest. Resour. Energy Econ. 38, 243–260.

Smith, A., 1776. The wealth of nations. na.

Stafford, T.M., 2018. Accounting for outside options in discrete choice models: An application to commercial fishing effort. J. Environ. Econ. Manag. 88, 159–179. https://doi.org/10.1016/j.jeem.2017.10.006

Truong, T., Adamowicz, W., Boxall, P.C., 2018. Modelling the Effect of Chronic Wasting Disease on Recreational Hunting Site Choice Preferences and Choice Set Formation over Time. Environ. Resour. Econ. 70, 271–295. https://doi.org/10.1007/s10640-017-0120-0

Tunçel, T., Hammitt, J.K., 2014. A new meta-analysis on the WTP/WTA disparity. J. Environ. Econ. Manag. 68, 175–187. https://doi.org/10.1016/j.jeem.2014.06.001

Vatin, F., Simonin, J.-P., Marco, L., 2016. The Works of Jules Dupuit, Classical texts in Economics and Management. Edi-Gestion.

von Haefen, R.H., Phaneuf, D.J., 2008. Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference approach. J. Environ. Econ. Manag. 56, 19–32. https://doi.org/10.1016/j.jeem.2008.01.002

Whitehead, J., Haab, T., Huang, J.-C., 2012. Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches. Routledge.

Whitehead, J., Haab, T., Huang, J.-C., 1998. Part-Whole Bias in Contingent Valuation: Will Scope Effects Be Detected with Inexpensive Survey Methods? South. Econ. J. 65, 160–168.

Whitehead, J.C., 2016. Plausible responsiveness to scope in contingent valuation. Ecol. Econ. 128, 17–22.

Whitehead, J.C., Pattanayak, S.K., Houtven, G.L.V., Gelso, B.R., 2008. Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science. J. Econ. Surv. 22, 872–908. https://doi.org/10.1111/j.1467-6419.2008.00552.x

Wooldridge, J.M., 2016. Introductory econometrics: A modern approach. Nelson Education.

Yi, D., Herriges, J.A., 2017. Convergent Validity and the Time Consistency of Preferences: Evidence from the Iowa Lakes Recreation Demand Project. Land Econ. 93, 269–291. https://doi.org/10.3368/le.93.2.269

Estimating the economic benefits of coastal quality change: An Application to Beach Recreation in Norway

Ana Faria Lopes[a] and Petr Mariel[b]

Abstract: Coastal managers are continuously considering improving the quality of coastal sites. To identify the quality change scenario that yields the highest economic welfare, it is useful to know the implicit value of the site attributes and thus determine the policy change that yields the most benefits. However, multicollinearity and lack of variation of site attributes complicate the task of estimating the implicit value of site attributes. To this end, we first develop an identification strategy relying on simulation and then apply the discrete choice model to explain recreational beach site choice. To the best of our knowledge, this is the first paper that tackles identification using RP data alone when few alternatives are available. We uncover preference heterogeneity by relying on observable group characteristics, namely group size and number of children. We illustrate the policy-relevance of our approach by provide welfare estimates for three scenarios currently being considered by Norwegian beach managers.

[a] University of Stavanger, UiS Business School; ana.f.lopes@uis.no.
[b] University of the Basque Country UPV/EHU; petr.mariel@ehu.eus.

The first author thanks Gorm Kipperberg for his input throughout the whole survey design process and John Whitehead for his input. The first author is also grateful to the participants of a stakeholder seminar held at the University of Stavanger, including Magnus Utsogn for the information provided to improve attribute data, and Per Kristian Austbø on his input on survey design. The first author thanks Jette Bredahl Jacobsen, John Rolfe, Ståle Navrud and Henrik Lindhjem for additional helpful comments on the survey design.

1 Introduction

Managers of recreational sites are responsible for improving the sites over which they have jurisdiction towards enhancing visitors' experiences. To this end, they should consider increasing the quality of these sites if the benefits of their improvement exceed the costs of implementing those changes. Consider the example of introducing a new bus route connecting major nearby cities to coastal recreational sites. While the costs of setting up such a bus route are retrieved from existing market prices (e.g., labor and gas costs), the recreational benefits of a new bus route are less evident. This is due to the public good nature of recreational sites, being non-excludable and non-rival. For recreational sites, no market prices exist and information predicting how visitation changes given policy scenarios is scarce, requiring economists to rely on non-market valuation methods to estimate benefits and costs.

The use of the travel cost method (TCM) applied to recreation is an example of a non-market valuation method. The TCM is a revealed preference method wherein the price to recreate at a site is the travel cost incurred to reach that site (Parsons, 2017). The analysis of recreational choices has both a participation and a site selection component. Our strategy is to apply the discrete choice model to understand recreational site choices. Analyzing site selection rather than participation frequency has some advantages: it allows for substitution across sites, we estimate the implicit prices of site attributes in a more straightforward manner, and we can account for preference heterogeneity (Parsons, 2017; Phaneuf and Requate, 2017).[1]

However, a challenge arises when operationalizing a discrete choice model of site choice. That is, if lack of variation and multicollinearity are present in the data, proper identification of the parameters of interest is

[1] A third analytical framework to analyze both participation and site selection of recreational activities is the use of corner solutions, or Kuhn-Tucker models. For a review of these models, see Phaneuf and Requate (2017).

challenging. That is the case of our data. The phenomenon of multicollinearity is common when handling RP or observed data (Adamowicz et al., 1994; Ben-Akiva et al., 2002; Earnhart, 2002), especially if the environmental goods available are rather homogenous. An additional challenge is the potential lack of variation in the RP data, due to recreational alternatives being too similar.

Two strategies have been proposed to solve the problem of identification in RP data: either combining RP with SP data to break the multicollinearity (von Haefen and Phaneuf, 2008), or ensuring proper identification by using Murdock (2006)'s two-stage strategy. However, nor do we have access to SP data, nor enough alternatives to apply Murdock (2006)'s strategy. Instead, we propose using simulation to tackle identification prior to estimation. We do this in three steps: 1) define possible population parameters and error term assumptions; 2) given step 1, find the alternative that yields the highest utility (i.e., the choice); and 3) predict the choice given the data. The parameters retrieved in step 3 should be the same as those defined in step 1. We thus select a functional form for the utility function that reduces the multicollinearity in the data by avoiding highly correlated explanatory variables. To the best of our knowledge, this is the first paper that tackles identification using RP data alone when few alternatives are available. The proposed solution to the identification problem expands the toolkit of practitioners that wish to explain observed choices among similar goods.

Accounting for preference heterogeneity is also relevant in the context of recreational choices. However, we find that multicollinearity and lack of variation do not allow the identification of the additional parameters needed to account for unobserved preference heterogeneity. Instead, we opt for controlling for observable characteristics through the introduction of interaction effects in the model.

Our case study pertains to the *Jæren* beaches in Norway. The *Jæren* beaches are located on the west-southern coast of Norway and are some of the most visited natural attractions in the country with at least 600.000

visitors per year (Sveen, 2018). The vast majority of these visits are day trips, making beach recreation in *Jæren* a pertinent case for the application of the TCM.

Coastal managers in Norway aim at increasing the quality of these beaches by for example improving facilities or maintaining their natural attributes. Some changes in beach quality have recently been concluded (e.g., improvement in the parking facilities in *Bore* beach); some are scheduled soon (e.g., improvement of parking facilities in *Brusand* beach); others remain under consideration (e.g., new bus route connecting *Jæren* to urban centers). Not only do we estimate the recreational benefits of these quality changes, but we also illustrate how welfare estimates can be used in a benefit-cost analysis.

Around ten studies have previously applied the travel cost method in Norway (Kipperberg et al., 2019; Navrud, 2001). However, none of these studies focus on site selection (Kipperberg et al., 2019; Navrud, 2001). To the best of our knowledge, this is the first study to apply a site choice model to recreational choices in Norway. In the realm of beach recreation, there are only two surveys conducted in Europe that analyze site choice: Mallorca in Spain (e.g., Bujosa et al., 2015), and West Brittany in France (LePlat et al., 2018), with our study being the third.

The remainder of this paper is structured as follows. Section 2 describes the theoretical framework and identification strategy. Section 3 describes the survey design process and data. Section 4 presents the results. Section 5 presents welfare change measures from three scenarios and a benefit-cost analysis of one welfare change scenario. Section 6 concludes.

2 Methods

Discrete choice modeling is a useful tool to analyze revealed preference (RP) data (e.g., English et al., 2018). RP surveys can elicit individuals' choices, resulting in a series of discrete outcomes. The Random Utility Model (RUM) is one possible model to analyze such choices. In the context of site choice, the RUM framework models the probability of selecting a site given the available choice set, the bundle of site attributes and the associated travel cost (Freeman et al., 2014; Haab and McConnell, 2002; Phaneuf and Requate, 2017). By observing the implicit trade-offs done by the respondents, researchers estimate the marginal utility of site attributes. They can then estimate welfare measures in the face of varying quality of one or more sites, changes in the probability of visitation across sites, or welfare losses in the case of site closure.

The remainder of this section follows to a great extent the theoretical framework established in Haab and McConnell (2002), Parsons (2017), Phaneuf and Requate (2017), and Freeman et al. (2014). Suppose an individual has decided to go to the beach but has yet to decide which beach to visit. For the sake of simplicity, assume an individual i has two possible beaches to choose from: *Sola* and *Bore*.

Each of the two beaches corresponds to certain level of quality, q_j, as well as a cost of travel associated with getting there, C_{ij}, where $j = \{Sola, Bore\}$. Beach quality q_j is measured through attributes, which are the same across respondents but differ for each beach (e.g., length of the beach, water quality, or presence of dunes).

Given the its quality and the cost incurred to get there, $Sola$ beach has utility $U_{i,SOLA}$:

$U_{i,SOLA} = U(M_i - C_{i,SOLA}, q_{SOLA}),$ (1)
where the individual's available income is denoted by M_i.

We separate the individual's utility into an observable component V_{ij} that is the indirect utility function, and an unobservable error term ε_{ij}:

$U_{ij} = V_{ij}(M_i - C_{ji}, q_j) + \varepsilon_{ij}\ ,\ j = \{Sola, Bore\},\ i = 1,2, \ldots, I.$ (2)

Utility is expected to increase with desirable beach attributes (e.g., water quality), and decrease with undesirable beach attributes (e.g., beach litter). If individual i is rational, when faced with the choice of either going to *Sola* or *Bore* beach, (s)he chooses the beach that yields the highest utility. An individual chooses *Sola* beach if equation (3) holds:

$U_{i,SOLA} \geq U_{i,BORE}.$ (3)

However, researchers do not observe utility, which is in nature a latent variable, but rather the discrete choice made by the individual and can thus model the probability of observing that choice. The probability of choosing $Sola$ beach rather than $Bore$ beach is:

$Prob(SOLA) = Prob(U_{i,SOLA} \geq U_{i,BORE}).$ (4)

Using the indirect utility function from (2), the probability of choosing *Sola* is rewritten as:

$$Prob(SOLA) = Prob(V_{i,SOLA} + \varepsilon_{i,SOLA} \geq V_{i,BORE} + \varepsilon_{i,BORE})$$
$$= Prob(V_{i,SOLA} - V_{i,BORE} \geq \varepsilon_{i,BORE} - \varepsilon_{i,SOLA}). \quad (5)$$

The example above can be generalized for a non-empty set of N beaches. Instead of two alternatives, a rational individual chooses the beach that gives him the highest utility from the group of N recreational sites available to him, i.e., the *choice set*, represented by $n = \{1,2, \ldots, j, \ldots, N\}$. In equation (5), the utility of choosing *Sola* beach must exceed the utility associated with any of the other alternatives in n.

To operationalize the RUM, we need two further assumptions: 1) the functional form of V_{ij} and 2) the distribution of ε_{ij}. The literature on discrete choice modeling commonly assumes linearity in parameters in the functional form of the indirect utility function. If so, the individual's indirect utility function is:

$$V_{ij} = -\beta_M C_{ij} + \beta_q q_j, \tag{6}$$

wherein the individual's income is omitted. β_M and β_q represent the marginal utility of money and beach quality, respectively.

The distribution of the error terms can be assumed to follow different distributions. If ε_n are identically and independently distributed type I extreme values, then the difference $\varepsilon_n - \varepsilon_j$ is logistically distributed. The probability of choosing beach j can be written as:

$$Prob(j) = \frac{e^{-\beta_M C_{ij} + \beta_q q_j}}{\sum_{n=1,\dots j\dots,N}^{N} e^{-\beta_M C_{in} + \beta_q q_n}}. \tag{7}$$

This probability results in the well-known conditional logit model. However, this model is based on relatively restrictive assumptions that include fixed preferences for all individuals. Other more complex models, such as the mixed logit and latent class models, allow for preference heterogeneity (Hensher and Greene, 2010), but they require the estimation of more parameters when compared with the conditional logit model.

In our application of the RUM, we analyze a single choice occasion (i.e., last visited beach by each respondent) using the conditional logit model given individual- and site-specific travel costs and site-specific quality. A popular alternative to the single choice occasion model would be to estimate a repeated logit model by using the visitation data for the entire summer season (e.g. English, 2010). However, we opt for analyzing a single choice occasion to use the detailed information we have regarding this visit (i.e. information on group size, mode of transportation, and time spent at the beach). We then calculate the welfare changes associated with different policies given estimated parameters by maximum likelihood, β_M and β_q.

2.1 Identification of relevant parameters

Discrete choice modeling has become mainstream to model stated preference (SP) data (e.g., Hoyos, 2010). In SP data, the variation of the attribute data is generated by the experimental design that assures identification. Identification pertains to the unambiguous determination of the coefficients of the model (Lancsar and Louviere, 2008). The concept of identification in a discrete choice model is usually related to the definition of the error term ε in (2) and (5). The assumption of a specific distribution and its parameters allows for identification of the parameters under the assumption of sufficient variation of the matrix of attributes.

In RP studies, however, data on attributes are often collected objectively by researchers based on direct observation or existing data (Adamowicz et al., 1997). The levels of attributes in RP studies cannot be experimentally varied. As a result, many attributes either do not have enough variation (e.g., a dummy for presence of lifeguard taking the same value across beaches) or suffer from high multicollinearity (e.g., number of toilets would be highly or perfectly correlated with other attributes). Our data are a prime example of this. Our data include information on 15 attributes: number of parking spaces, dummy for area protected for bird species, water quality index, beach length and width, presence of rocks, dunes, marina, recycling bins, bike paths nearby and camping possibilities, number of toilets, public access points to beach and food amenities (bars, restaurants and kiosks), and congestion.[1] Our data suffer from both lack of variation and high multicollinearity. For example, lack of variation is present in the water quality variable: although the scale ranges from 1 to 5 (very bad to very good quality, respectively), the observed water quality along the study site only takes the value 3 (moderate) or 4 (good quality). High multicollinearity is also

[1] We have more attributes than the average in site choice models applied to beach recreations (average of 9.69 attributes in 39 studies). The number of attributes in past studies ranges from 2 (Chen and Lupi, 2013; Hicks and Strand, 2000; Whitehead et al., 2008a) to 30 (Pendleton, 2012)

present in our data. For example, the correlation coefficient between the attribute levels for camping and food amenities is relatively high (0.72). Multicollinearity and lack of variation can complicate the identification and precise estimation of parameters of interest, occasionally resulting in estimated coefficients having counter-intuitive signs or being statistically (in)significant.

In fact, lack of variation and multicollinearity in RP data is one of the main motivations to combine RP and SP data (von Haefen and Phaneuf, 2008; Whitehead et al., 2008b). Some studies, such as Adamowicz et al. (1994), Ben-Akiva et al. (2002), and Earnhart (2002), combine RP and SP data to reduce the collinearity present in the attribute levels and thus allow for the strong identification of attribute coefficients.

In the absence of SP data, Murdock (2006) proposes a two-stage strategy to identify unbiased parameters of attributes with RP data. In the first stage, a site choice model is estimated given travel costs, any interaction of individual and attributes characteristics, and a full set of alternative specific constants (ASCs). In a second stage, the estimates of the ASCs of the first stage become the dependent variable in an ordinary least squares regression given observed site attributes. The number of observations in this second-stage is equal to the number of available sites. Some site choice model applications already apply this strategy (e.g., Timmins and Murdock, 2007).

However, we cannot apply the proposed model in our context. Since our aim is to estimate the welfare change given changes in attributes, we need to estimate the parameters of site attributes, which would require the second-stage regression proposed by Murdock (2006). Yet, our choice set only includes twenty different beaches.

Instead, our strategy is to identify a subset of attributes and ASCs whose parameters can be identified. To the best of our knowledge, this is the first study to tackle identification using only RP data. We do this by simulation. Our approach is summarized in Figure 1.

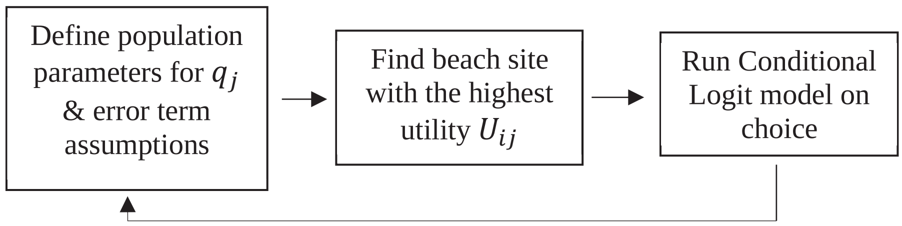


Figure 1 – Summary of identification strategy (Steps 1, 2 and 3)

We assume that the data generation process comes from an underlying RUM model, in which the errors fulfill all underlying assumptions. Our objective is to identify the specification where the β parameters can be properly retrieved.

In Step 1, we calculate the indirect utility function with a defined set of parameters set to a priori values ($\widehat{\beta_M}$ and $\widehat{\beta_q}$) as obtained from our pilot data. To do so, we estimate the utility of k different specifications as follows:

$$U_{ijk} = -\widehat{\beta_M} C_{ij} + \widehat{\beta_q} \boldsymbol{q}_j + \widehat{\beta_{ASC}} ASC_j + \varepsilon_{ij}, \tag{8}$$

where $\boldsymbol{q}_j$ is a vector representing various beach attributes. Eleven beach attributes are considered. Each U_{ijk} depends on a different combination of beach attributes, specified in $\boldsymbol{q}_j$, and alternative specific constants. We consider a total of 572 combinations of attributes and alternative specific constants in the utility function in (8).

The errors ε in (8) are assumed to be identically and independently Gumbel distributed (type I extreme value distribution with location parameter zero and scale one). In Step 2, given the utility estimates obtained from (8), we find the beach site j for each individual that gives him/her the highest utility. This should be the site choice that the individual makes. We use this choice as the dependent variable in the conditional logit model (step 3) to estimate β_M, β_q and β_{ASC}.

We analyze the empirical distribution of all estimated parameters that depend critically on the properties of the matrix gathering levels of all included attributes and ASCs. Table 1 reports descriptive statistics of average difference between the a priori values ($\widehat{\beta_M}$, $\widehat{\beta_q}$ and $\widehat{\beta_{ASC}}$) and the conditional logit model estimates for a subset of the attributes.

Table 1 – Descriptive Statistics for the absolute difference between a priori values and estimated parameters

	Mean	**SD**	**Min**	**Max**
β_M	0.001	0.000	0.000	0.002
$\beta_{Parking}$	0.000	0.000	0.000	0.002
$\beta_{Congestion}$	0.820	1.533	0.000	20.037
$\beta_{BirdProtected}$	0.092	0.072	0.000	0.340
β_{Length}	0.000	0.000	0.000	0.001
β_{Width}	0.003	0.003	0.000	0.019
β_{Rocks}	0.144	0.137	0.001	1.286

We find that the distribution of the estimated parameters differs from the assumed parameter values in some of the specifications. For example, the difference between the a priori value and the estimated parameter of $\boldsymbol{\beta_{Congestion}}$ may be as high as 20. Furthermore, we find that the exclusion of the Bird Protected dummy improves identification for many of the parameters. We identify one specification where parameters can be correctly identified, including 7 out of the 11 attributes. These are parking, length, width, rocks, dunes, toilet and food amenities. Since the research focus is to estimate welfare change from differing site quality, we opt for the inclusion of attributes to explain site choice in detriment of ASCs. The chosen specification is applied in Section 4.

We also simulate whether the variation of the attribute levels in our sample is sufficient to identify parameters in more complex models such as mixed logit or latent class model that allow modeling of unobserved preference heterogeneity. The results indicate that the variation of the attribute levels in our dataset is not sufficient to retrieve the additional true parameters. That is why we opt for adding flexibility to our conditional logit model by interacting the attribute coefficients with the observed group characteristics.

3 Survey Area and Data

Our study area, the *Jæren* coast, is located in the county of Rogaland (southwestern coast of Norway). It is the host of thirteen signaled beaches, whose length varies from 10 to 650 meters (Sveen, 2018). These beaches are located in a 70-kilometer stretch from *Tungenes* in the North to *Ogna* in the South (see Figure 2). The area is classified as a nature conservation area since 1977 due to its geological, botanical, zoological and cultural heritage value. The beaches have white sand, dunes and many rare species and vegetation systems. The coast provides areas for birds to find shelter and nest.

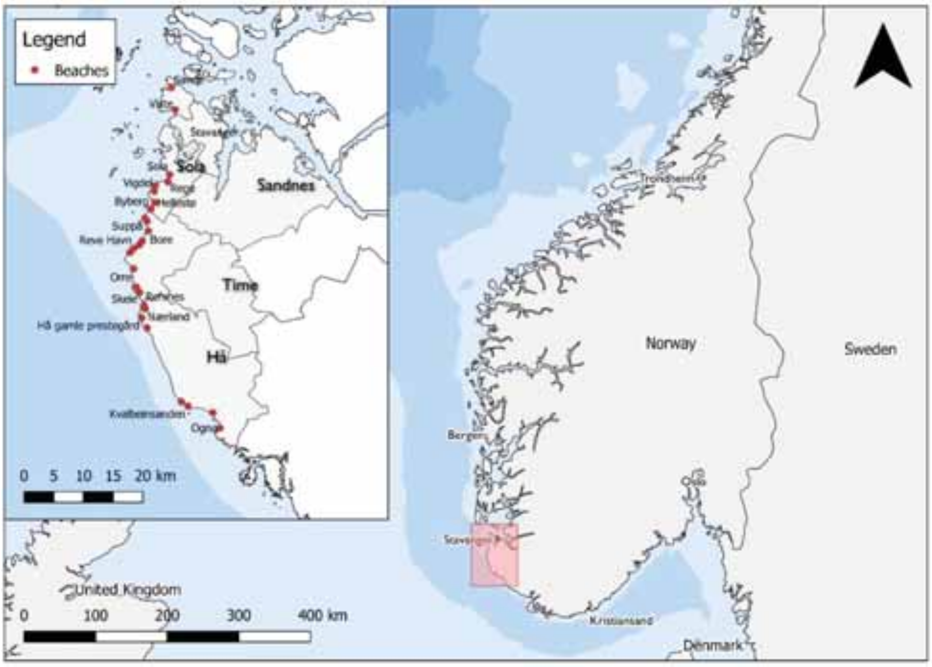


Figure 2 - Map of the study area: the Jæren coast and its beaches

3.1 Survey Design

Survey design started in January 2017. Students carried out three pilot studies: one in Easter 2017 (Bui and Sæland, 2017) and two in Easter

2018 (Gilje, 2018; Kleppe and Jensen, 2018). Sampling for the pilot studies was done on-site at four beaches. We were able to identify the most relevant attributes, the activity engaged in by respondents, and obtain the first estimates of consumer surplus.

We based the design of the survey on nine previous state-of-the-art studies that resulted in a site choice model application (e.g. Bin et al., 2007; Bujosa et al., 2015; Chen, 2013; Hicks and Strand, 2000; Leggett et al., 2014; Lew and Larson, 2008; Matthews et al., 2018; Parsons et al., 1999; Yeh et al., 2006). Three national environmental economics experts commented on the design of the survey, specifically to reduce recall bias. We consulted coastal managers, namely from *Jæren Friluftsråd* and *Fylkesmannen i Rogaland*, who helped expand the list of beach names, and identify coastal threats and relevant policy scenarios.

In order to gather data to design the questionnaire, we conducted one focus group in March 2018. The eight participants, who were employees at the university, were not informed about the topic of the discussion before the meeting. The focus group included a discussion concerning motivations for choosing a particular beach, identification of the coastal threats, and ranking of beach attributes.

To test the survey, we conducted six personal interviews in September 2018. We first asked participants to fill out the survey without assistance. We then asked them some debriefing questions about general comprehension of the survey and various aspects related to their last visit (e.g., the relevance of overnight trips, and identification of appropriate substitute sites).

3.2 Data

We collected data during October and November of 2018 using a web panel from a survey company (*Norstat*). Whereas most TCM data are collected on-site (e.g., Bin et al., 2007), we conducted the survey off-site by sampling residents in the Rogaland county of Norway. We collected 982 responses, resulting in a response rate of 25.9%.

Nearly all respondents (98.3%) reported knowing or having heard of at least one of the beaches in *Jæren*. On average, respondents took 29 minutes to respond to the survey and a median time of 16 minutes.

To ensure that our sample is representative of the Rogaland population, we compare key statistics of the population with the sample means in Table 2. Respondents were randomly selected, which implies that every member of our population of interest (residents of the Rogaland county) has the same probability of being selected to answer the survey. Respondents were also not informed about the topic of the survey prior to answering it. We conclude that the sample is representative, as most sample means are not statistically different from the population means (see Table 2)[1]. Respondents are on average more educated than the population, as is common in Internet-based surveys (e.g., Lindhjem, 2011). We replace missing data on income with the population's mean income, adjusted for the number of household members.

Table 2 - Comparison of Descriptive Statistics between Population of county residents and Respondent Sample (N = 965)

		Respondent Sample	**Population**
Continuous Variables		**Mean**	**Mean**
Household Gross Income (NOK per year)		808 333	874 400
Household Size		2.56	2.32
Age		47.28	37.62
Dummy Variables		**Proportion**	**Proportion**
Education Attainment	Primary school	4.49%	25.70%
	High school	36.15%	39.20%
	Vocational or university education	59.36%	35.10%
Gender (% of women)		54.30%	49.20%

Source: SSB (Statistics Norway) for population means for the year 2016. As of 12/06/2019: 1 Euro = NOK 9.7710; 1 USD = NOK 8.6318 (Source: https://www.bloomberg.com/markets/currencies)

[1] Although respondents are on average 10 years older than the population, this is because we excluded people under 18 years of age from answering the survey.

Our survey elicited both the respondent's general visitation pattern, and detailed information on the last beach visit during the summer season of 2018. Around 68% of the sample reported having at least one visit to the Jæren beaches in the summer season of 2018. Therefore, the final sample size to analyze the choice of the last beach visited consists of 657 respondents. The thirteen main beaches represent 89.6% of the visitation. The most visited beaches are *Sola* (32.7%) and *Ølberg* (15.1%), followed by *Bore*, *Orre*, and *Hellestø* (see Figure 3). Norwegians use beaches differently from traditional beach users: the intention upon visiting for the majority of the respondents is to go on walks or to relax.

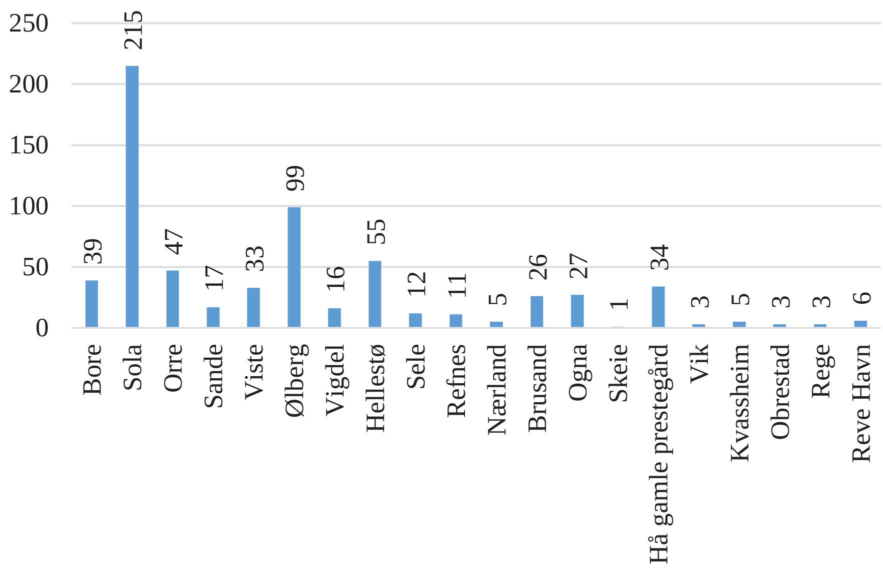


Figure 3 - Distribution of last visited beach reported along the twenty Jæren beaches

The respondent's travel cost represents the various costs incurred to visit the beach. The calculation of the travel cost is conditional on the mode of transportation, which we elicited for each respondent. The majority of respondents traveled by diesel car (40%) and by petrol car (34%). The remainder traveled by electric car (7%), hybrid car (9%), bicycle (3.8%), public transportation (2.4%) and on foot (3.2%).

The travel cost C_{ij} to beach j of group i is given by:

$$C_{ij} = \left(p_d d_{ij} + f_j + 2 * p_j * g_i + w_i t_{ij}\right) * \delta_i. \tag{9}$$

where p_d denotes the per kilometer cost of travel, and d_{ij} the round-trip distance traveled in kilometers. For groups traveling by car, $p_d d_{ij}$ is the roundtrip distance traveled times the money cost (in Norwegian kroner) per kilometer. We measure the distance traveled between the respondent's zip code and the beach's parking lot coordinates using the google maps API tool.

Groups traveling by diesel, petrol or hybrid cars also incur a toll fee, denoted by f, of 20 NOK. For groups traveling by bus or train, we multiply the ticket price, denoted by p_j (35 and 70 NOK, respectively) by the group size g_i irrespective of the distance traveled.

The round-trip travel time spent (in hours) t_{ij} is calculated using the google maps API tool, and it is conditional on the group's mode of transportation. If groups are free to choose the number of hours worked at a given wage rate, then the opportunity cost of time, w_i, simplifies to the group's wage rate (Freeman et al., 2014). w_i is assumed to be one third of the group's net hourly wage rate, given an average of 1950 hours of work per year. We adjust for multiple-purpose trips following the method proposed by Yeh et al. (2006), and thus weigh the travel cost variable with the term δ_i, which denotes the percentage of the travel reported to have been spent in that beach.

We collect data on various beach attributes related to parking, accessibility, water, land cover nearby, protected area status, physical characteristics, natural characteristics, facilities, and litter. In the site choice modeling literature, the most common attributes are beach length, the presence of a park and parking (Bujosa et al., 2015; Hilger, 2006; Lew and Larson, 2008; Massey and Parsons, 2007). Other beach attributes considered by previous literature are beach width (Bin et al., 2007), urbanized area (Bujosa et al., 2015), presence of litter (Leggett et al., 2014), level of congestion (Cushman et al., 2004), tree coverage (Font, 2000), and water quality (Hicks and Strand, 2000), to name a few.

As explained in Section 2.1, we select a subset of the attributes as explanatory variables. The results of the focus group also guided the subset of the relevant attributes to explain beach choice. Table 3 summarizes these attributes.

Table 3 - Beach Attributes' Description (name, description, data source, average, minimum and maximum attribute level for all 20 sites)

Name of Variable	**Description**	**Source**	**Mean**	**Min**	**Max**
Parking Spaces	Number of public parking spaces available	Coastal Managers (Jæren Friluftsråd)	123.45	0	360
Length	Length of the beach (in meters)	Spatial data (Google maps satellite images)	805.45	0	2810
Width	Width of the beach (in meters)	Spatial data (Google maps satellite images)	32.07	0	68
Rocks	Dummy: 1 if the beach has rocks or cobblestones; 0 if only white sand	Spatial data (Google maps satellite images)	0.40	0	1
Dunes	Dummy: 1 if the beach has dunes; 0 otherwise	Coastal Managers (Fylkesmannen i Rogaland)	0.60	0	1
Toilets	Number of toilets	Coastal Managers (Jæren Friluftsråd)	1.70	0	4
Food amenities	Number of restaurants, bars and kiosks nearby	Coastal Managers (Jæren Friluftsråd) & Visitor Reviews (Trip Advisor)	0.75	0	3

4 Results

We analyze the choice of the last visited beach in the 2018 summer season along the *Jæren* coast. These choices are conditional on individual and site-specific travel costs, and site-specific attributes: the number of parking spaces, toilet, and food amenities, and whether the area has rocks or cobblestones, or dunes, as well as length and width of the beach. We have twenty beaches along Jæren that respondents reported as their last visited beach.

We do not include congestion as an explanatory variable. The well-known challenge of including congestion in discrete choice models is endogeneity: the same unobserved factors that drive the site choice of the individual also influence congestion at each site (Hindsley et al., 2007). Most authors account for the endogenous nature of congestion using an instrumental variables approach (e.g., Boxall et al., 2005; Timmins and Murdock, 2007) in the two-stage model proposed by Murdock (2006).

We expect higher travel costs to decrease the probability of visiting a beach, leading respondents to be more likely to visit the beach sites closest to them. We also expect utility to increase with the number of parking spaces, toilets and food amenities. The remaining four attributes (*Length, Width, Dunes,* and *Rocks*) can be either considered *a priori* an amenity or disamenity by visitors.

As mentioned in Section 2.1, we focus on preference heterogeneity by accounting for observed characteristics of the individuals. Common variables that can explain preference heterogeneity include gender, age, group size, number of children in the group, and income. However, the beach choice is the result of a group-based decision process, rather than an individual decision. There is no guarantee that individual characteristics help explain preference regarding attributes; rather it should be group characteristics that better explain beach choice. Indeed, when interacting beach attributes with several individual characteristics (i.e., age, gender, membership to an environmental or touristic organization, and perceived knowledge about coastal fauna and flora), we do find that group characteristics explain beach choice better than individual characteristics.[1] Kaoru (1995) also find evidence that group composition influences recreational decisions. We explain beach choice by adding interaction effects of beach attributes and observable group characteristics.

Given data availability, we use two group characteristics to disentangle the observed preference heterogeneity: group size and the number of children. The median group consists of two people, while the average group comprises of 3.13 visitors. Most groups do not include children. Another candidate variable to uncover preference heterogeneity is the activity engaged in by the group (e.g., sunbathing, running, fishing, walking, and relaxing). However, the fit of the specifications interacting the activities engaged in with the beach attributes is inferior to those of the specifications using the group composition variables.[2]

We estimate a conditional logit model with this set of explanatory variables (results are reported in Table 4). As predicted, travel cost has a negative impact on utility, and hence on the probability of visitation.[3] At the mean, the number of toilets and food amenities is welfare-enhancing and significant, as expected. Respondents also prefer longer beaches (i.e., the coefficient is positive and significant). Adding the interaction

[1] These results are available upon request.
[2] These results are available upon request.
[3] We do sensitivity analysis on the travel cost variable by: 1) not adjusting for multiple-purpose trips (Yeh et al., 2006) hence assuming that δ is one for all respondents; 2) using the self-reported departure coordinates to calculate distances and times instead of the postal code; 3) using a different percentage (50%) of the wage rate as the opportunity cost of time. While the coefficients of all attributes (except travel cost) remain unchanged, the fit of the models deteriorate in all the sensitivity analyzes. Therefore, we choose to keep the adjustment for multiple purpose trips as proposed by Yeh et al. (2006), the postal codes as the departure coordinates, and 33% of the wage rate as the opportunity cost of time.

effects improves the fit of the model when compared with the model omitting any interactions (AIC decreases from 3007.616 to 2906.081).[4]

Table 4 - Estimation Results for the Conditional Logit Model

	Conditional Logit Model		
***Dep. Var.:* Beach Choice**	**Mean Effect**	**Interaction effects with Group Size**	**Interaction effects with the number of children**
Travel Cost	-0.013*** (0.001)	0.001*** (0.0001)	-0.001*** (0.0002)
Parking Spaces	0.001 (0.001)	0.001 (0.0005)	-0.002* (0.001)
Length	0.0002* (0.0001)	0.0001 (0.000)	-0.0002* (0.0001)
Width	0.006 (0.008)	-0.004 (0.003)	-0.010 (0.005)
Rocks	-0.456 (0.308)	-0.090 (0.119)	-0.156 (0.172)
Dunes	0.239 (0.336)	-0.050 (0.125)	0.682** (0.239)
Toilets	0.335** (0.102)	-0.083* (0.041)	0.114* (0.054)
Food Amenities	0.386** (0.137)	-0.026 (0.053)	0.028 (0.067)
Number of observations	657		
Log-likelihood	-1429		
AIC	2906.081		
BIC	3013.785		

Note: *** denotes statistical significance at the 0.1% level, ** at the 1% level, and * at the 5% l evel.

Preferences given group size (third column) differ in what concerns the number of toilets and travel costs. Preference heterogeneity regarding the travel cost variable is fairly intuitive: the larger the group, the more the group shares the costs of travel, and thus are less sensitive to the

[4] These results are available upon request.

travel cost variable. This result is also suggested in Kaoru (1995). Larger groups also appear to place less importance on the number of toilets available at each beach, but the net effect for groups up to three people is still positive.

The preferences of groups with children suggest additional preference heterogeneity patterns. The more children in the group, the more sensitive the group is to the travel cost incurred to reach the beach. Groups with more children also have stronger preferences for beaches with dunes, more toilets, fewer parking spaces and shorter beaches, when compared to groups without children.

4.1 Marginal Willingness to Pay

While the estimates in Table 4 are informative to understand preferences, they are not directly interpretable, since these are in utility-space. To compare benefits to costs of beach quality change, it is useful to convert the relative importance of each attribute into a money metric. We can calculate the marginal Willingness to Pay (WTP) for attributes as:

$$WTP = -\frac{\beta_q}{\beta_M}. \tag{10}$$

Welfare estimates are expressed per visit and per group, rather than per person.

In the previous section, we uncover some patterns of preference heterogeneity. Hence, the marginal WTP for each attribute varies across groups of different size and composition. Two of the most common group compositions in our sample are 1) two adults and no children (i.e., group size of two); and 2) two adults and two children (i.e., group size of four). We estimate marginal WTP using the mean group composition, as well as these two common group compositions.

Table 5 reports the estimated marginal WTPs. The standard errors of the WTPs are computed by the delta method.

Table 5 - Marginal WTP (in NOK) for beach attributes in Jæren beaches

	[1] Mean Group Size (3.12) and children (0.68)	**[2] Two adults**	**[3] Two adults and two children**
Parking Spaces	0.16	0.19*	-0.08
Length	0.04***	0.04***	0.01
Width	-0.95	-0.11	-1.49*
Rocks	-69.44**	-52.50**	-68.50**
Dunes	43.73	11.48	108.49***
Toilets	12.59	14.05	28.79***
FoodAmenities	26.78*	27.61**	28.18**

Note: *** denotes statistical significance at the 0.1% level, ** at the 1% level, and * at the 5% level. As of 12/06/2019: 1 Euro = NOK 9.7710; 1 USD = NOK 8.6318 (Source: https://www.bloomberg.com/markets/currencies)

For the average group [1], the most valuable attributes seem to be the absence of rocks, followed by the number of food amenities and the length of the beach. The average group would be willing to pay 27 Norwegian kroners for an additional bar, restaurant or kiosk in their chosen beach. The average group would also be willing to pay almost 70 kroners to avoid a beach with rocks or cobblestones.

The highest marginal WTP across all three group compositions considered is for the dunes attribute (108.5 NOK). While the average group [1] is not willing to pay to visit a beach with dunes, groups with children [3] are willing to pay the most to have access to dunes. Groups without children [2] are willing to pay for a higher number of parking spaces (0.19 NOK per parking space) and longer beaches (0.04 NOK per meter of length).

5 Policy Implications

With discrete choice models, researchers are capable of estimating the change in compensating variation following a change in quality at the study site (e.g. Lew and Larson, 2005). Compensating variation (CV) is a measure of welfare change given a change in quality. Equation 11 shows how to estimate the CV associated with a change in site quality:

$$E(CV) = \frac{1}{\beta_M}\left\{\ln\left[\sum_{n=1,\dots j\dots,N}^{N} e^{V_{in}\left(M_i - C_{in}{}^1, q_n{}^1\right)}\right] - \ln\left[\sum_{n=1,\dots j\dots,N}^{N} e^{V_{in}\left(M_i - C_{in}{}^0, q_n{}^0\right)}\right]\right\}, \tag{11}$$

where $C_{in}{}^0$ and $C_{in}{}^1$ denote the travel costs under the initial scenario 0 and the new scenario 1, respectively. Likewise, $q_n{}^0$ and $q_n{}^1$ denote the site quality under the initial and new scenarios.

We use three different scenarios to illustrate the change in CV. First, we consider improvement of parking and toilet facilities. These were concluded in 2018 in *Bore* beach, but during the summer season of 2018, these were not open to the public (Personal Communication, *Jæren Friluftsråd*). Further improvements at *Brusand* beach are expected by 2022 (Schibevaag, 2016). The first scenario involves the estimation of the benefits from the improvement of facilities, consisting of 154 additional parking spaces in *Bore* and 20 additional parking spaces in *Brusand* beach, as well as adding an extra toilet in both *Bore* and *Brusand* beaches (Schibevaag, 2016). We expect a slight welfare gain in this scenario.

Second, the *Jæren* area is under several threats, including the wear-and-tear of beach dunes. This threat is especially relevant, not only for visitors but for the coastal environment. In six of the 20 beach sites (*Sola, Vigdel, Hellestø, Bore, Refnes,* and *Brusand*), it is recommended to avoid walking on dunes since these are damaged (Fylkesmannen i Rogaland, 2018). The second scenario simulates the change in CV in case these six sites were to lose their dunes. We expect a substantial welfare loss.

Third, available public transportation to and from the *Jæren* beaches is of poor quality. One coastal manager (*Fylkesmannen i Rogaland)* is currently considering the creation of a free bus route during the summer from the two main cities (Stavanger and Sandnes). We simulate the welfare change from such a bus route to the five most visited beaches. We assume that visitors change from their elicited mode of transportation to this new bus route only if their travel cost decreases.[1] Hence, this change is through decreased travel costs for some of the visitors. We expect a welfare gain from this scenario.

Table 6 presents the CV given the three scenarios for different group compositions.

Table 6 - Compensating Variation in NOK (per group and per visit) for three policy scenarios (standard deviation in parenthesis)

Mean CV in NOK (per group and per visit)	**Mean Group Size (3.12) and children (0.68)**	**Two adults**	**Two adults and two children**
Scenario 1: Increase in number of facilities (i.e., toilets and parking spaces) in two beaches (*Bore* and *Brusand*)	5.33 (0.11)	5.95 (0.13)	3.23 (0.24)
Scenario 2: Loss of dunes in six beaches where dunes are currently damaged	-13.64 (0.57)	-5.64 (0.08)	-36.89 (0.86)

[1] One referee pointed out that groups might have strong preferences towards the mode of transportation. For example, we expect that larger groups with more children would still not opt for using a free bus due to the convenience of traveling by car even if their travel costs are reduced. Hence, the assumption of groups changing their mode of transportation may not hold for some specific groups. In such a case, the number of people that would change mode of transportation would be overstated and the resulting welfare estimates of introduction of a free bus would be biased upwards. However, we do find that smaller groups with less children would use the free bus using the travel cost reduction assumption. We find that the average group size is smaller for the groups that take the free bus (2.5 people), rather than the groups that do not take it (3.3. people). Likewise, the groups that change for the free bus have on average less children (0.34) than the groups that do not take the free bus (0.74 children). Therefore, we recognize the potential bias in the estimated welfare gain, but the resulting group composition gives credibility to the robustness of the assumption.

Scenario 3: New bus route from main nearby cities to the five most popular beaches	4.26 (0.42)	4.55 (0.76)	1.17 (0.33)

As expected, Scenarios 1 and 3 yield welfare gains for visitors of 5.33 NOK and 4.26 NOK per group and per visit, respectively.[2] On the other hand, the loss of beach dunes in Scenario 2 generates a significant welfare loss. In this scenario, visitors are willing to pay 13.64 NOK per group per visit to avoid the loss of beach dunes at the six beaches.

When considering different group compositions, the results suggest that larger groups with children have more modest welfare gains than smaller groups without children. For scenarios 1 and 3, the welfare gain for groups with children (fourth column of Table 6) is smaller than for groups without children. As for Scenario 2 (loss of dunes), the welfare loss of groups with children (-36.89 NOK) is six times higher than the welfare loss for the group without children (-5.64 NOK).

With the estimate from changes in recreational benefits and an estimate of costs (Schibevaag, 2016), it is possible to conduct a benefit-cost analysis for Scenario 1.

We assume a lower bound number of annual visitors to Jæren of 600 000 (Sveen, 2018) and the mean group size from our sample of 3.12 (see Table 6). This results in an estimate of 192 307 groups of visitors per year in the region. Hence, the estimate of the aggregate recreational benefits of Scenario 1 are 1 024 996 NOK per year (number of groups per year multiplied by the mean CV of Scenario 1 in Table 6). We assume that preferences for beach attributes do not change over time, hence recreational benefits are incurred in perpetuity. We use three discount

[2] The number of groups that would change from their elicited mode of transportation to the new bus route is simulated to be 144 out of the 657 responses. For these 144 groups, the travel cost variable decreases, hence the welfare gain in this scenario. While we would also expect that the number of total visits would increase given a new bus route, this model only predicts changes across visitation sites and is not able to predict changes in the number of visits. To this end, a repeated site choice model or a count model would be more appropriate.

rates to calculate the present value of the flow of aggregate benefits: 4% for the first 40 years, 3% for the subsequent 35 years, and 2% thereafter (DFØ, 2018). The present value of benefits associated with Scenario 1 is 38 645 143 NOK (see Table 7).

Table 7 - Benefit-Cost Analysis of Scenario 1 (in Norwegian kroners)

	Mean	***Lower Bound (95% Confidence Interval)***	***Upper Bound (95% Confidence Interval)***
Mean Compensating Variation (kr/group-visit)	5.33	5.11	5.55
Aggregate benefits (M kr)	38.65	37.08	40.21
Benefit-Cost Ratio	7.73	7.42	8.04

According to Schibevaag (2016), the improvement in facilities simulated in Scenario 1 is estimated to cost 5 million kroner. Our estimate of 38.7 million in aggregate benefits exceeds by seven the estimated costs. Even using the lower bound of the 95% confidence interval of the CV, the benefit-cost ratio is always higher than seven. We thus conclude that the proposed facilities improvement in *Bore* and *Brusand* beaches is economically efficient.

6 Conclusions

The quality of coastal areas may change over time, namely due to human intervention. Coastal managers may intervene by improving facilities or restricting access to sites. These interventions change the recreationist's probability of visiting each site, and it is useful for coastal managers to know how recreational values may change when introducing new measures. The application of a site choice model allows us to estimate welfare changes in the face of different scenarios.

When using RP data to estimate a site choice model, identification of relevant attributes is challenging. That is due to multicollinearity and lack of variation in attribute data. Yet, we show how to ensure identification using RP data alone. To the best of our knowledge, this is the first paper to explicitly tackle identification with RP data with limited choice sets. By first simulating recreationists' choices with the attribute matrix, we identify a subset of attributes for which identification is possible in a conditional logit model.

We apply our model to recreational choices in cold-water beaches on the southwestern coast of Norway. Our study is the first site choice model applied to Norway, and the third beach study-site in Europe wherein a site choice model is applied.

The travel cost variable is negative and statistically significant, thus exhibiting negative price sensitivity (Bishop and Boyle, 2019). Like previous studies (Bestard, 2014; Lew and Larson, 2008; Parsons and Stefanova, 2009), we find that parking (i.e., number of parking spaces) and toilets (i.e., number of toilets) are considered amenities and hence these increase the probability of visitation. The opposite is true for the presence of rocks: our results conform to previous findings by Lew and Larson (2005) in San Diego beaches that the presence of rocks decreases the probability of visitation. Whether beach length and width are amenities or not seems to be context-specific: these are found to be disamenities in South African beaches (e.g. Du Preez, 2011) and amenities in the Mid-Atlantic region of the US (e.g. Parsons et al., 1999).

In our application, beach length is a desirable attribute, while beach width is not. Finally, in our case, the presence of dunes is viewed as an amenity, while Bestard (2014) found it is an undesirable attribute for the case of Mallorca beaches.

However, different groups have distinct preferences. We find that groups without children prefer parking spaces and longer beaches. On the other hand, groups with children prefer beaches with dunes, more toilet facilities, and narrower beaches.

We analyze three scenarios involving changes in beach quality. Scenario 1 simulates the improvement in parking and toilet facilities in two beaches, where some of the improvements have recently been concluded. Scenario 2 estimates the welfare loss due to dune deterioration. Scenario 3 simulates the impact of a new bus route, which coastal managers are currently considering.

Scenarios 1 and 3 involve an improvement in beach quality and a decrease in travel costs, respectively, and thus are welfare-enhancing. Scenario 2, on the other hand, results in a loss in welfare, highlighting the critical role of dunes for the experience of visitors. Groups with children appear to be more affected in Scenario 2 compared with groups without children, while the opposite is true in Scenarios 1 and 3. We conduct a benefit-cost analysis to Scenario 1 and conclude that even with conservative estimates on the number of yearly visits, this scenario is economically efficient relative to no changes.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. J. Environ. Econ. Manag. 26, 271–292. https://doi.org/10.1006/jeem.1994.1017

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., Williams, M., 1997. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J. Environ. Econ. Manag. 32, 65–84.

Ben-Akiva, M., McFadden, D., Morikawa, T., 2002. Discrete choice models incorporating revealed preferences and psychometric data, in: Advances in Econometrics, Advances in Econometrics. Emerald Group Publishing Limited, pp. 29–55. https://doi.org/10.1016/S0731-9053(02)16003-8

Bestard, A.B., 2014. Substitution patterns across alternatives as a source of preference heterogeneity in recreation demand models. J. Environ. Manage. 144, 212–217.

Bin, O., Dumas, C., Poulter, B., Whitehead, J.C., 2007. Measuring the impacts of climate change on North Carolina coastal resources, Final Rep. for National Commission on Energy Policy.

Bishop, R.C., Boyle, K.J., 2019. Reliability and Validity in Nonmarket Valuation. Environ. Resour. Econ. 72, 559–582. https://doi.org/10.1007/s10640-017-0215-7

Boxall, P.C., Hauer, G., Adamowicz, W.L., 2005. Modeling Congestion as a Form of Interdependence in Random Utility Models (No. 24078), Staff Paper Series. University of Alberta, Department of Resource Economics and Environmental Sociology.

Bui, L., Sæland, S., 2017. Verdsetting av et gode uten markedsverdi: Tilfellet rekreasjon på Jærstrendene.

Bujosa, A., Riera, A., Hicks, R.L., McConnell, K.E., 2015. Densities Rather than Shares: Improving the Measurement of Congestion in

Recreation Demand Models. Environ. Resour. Econ. 61, 127–140. https://doi.org/10.1007/s10640-014-9785-9

Chen, M., 2013. Valuation of public Great Lakes beaches in Michigan. Michigan State University.

Chen, M., Lupi, F., 2013. Modeling Long Overnight Trips by Chaining Recreation Sites. Presented at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC.

Cushman, C.A., Field, B.C., Lass, D.A., Stevens, T.H., 2004. External costs from increased island visitation: results from the southern Thai islands. Tour. Econ. 10, 207–219.

DFØ, 2018. Veileder i samfunnsøkonomiske analyser. Direktoratet for økonomistyring, Oslo.

Du Preez, M.L., 2011. The recreational value of beaches in the Nelson Mandela Bay area, South Africa. Stud. Econ. Econom. 35, 85–102.

Earnhart, D., 2002. Combining Revealed and Stated Data to Examine Housing Decisions Using Discrete Choice Analysis. J. Urban Econ. 51, 143–169. https://doi.org/10.1006/juec.2001.2241

English, E., von Haefen, R.H., Herriges, J., Leggett, C., Lupi, F., McConnell, K., Welsh, M., Domanski, A., Meade, N., 2018. Estimating the value of lost recreation days from the Deepwater Horizon oil spill. J. Environ. Econ. Manag. 91, 26–45. https://doi.org/10.1016/j.jeem.2018.06.010

Font, A.R., 2000. Mass tourism and the demand for protected natural areas: A travel cost approach. J. Environ. Econ. Manag. 39, 97–116. https://doi.org/10.1006/jeem.1999.1094

Freeman, A.M., Herriges, J.A., Kling, C.L., 2014. The measurement of environmental and resource values: theory and methods, 3rd ed. ed. Resources for the Future, Washington, DC.

Fylkesmannen i Rogaland, 2018. Besøksstrategi for Jærstrendene [WWW Document]. Rogaland. URL https://www.fylkesmannen.no/nb/Rogaland/Miljo-ogklima/Jarstrendene/Besoksstrategi-for-Jarstrendene/ (accessed 11.1.18).

Gilje, H.L., 2018. Surfonomics: The Value of a Wave.

Haab, T.C., McConnell, K.E., 2002. Valuing environmental and natural resources: the econometrics of non-market valuation. E. Elgar Pub., Cheltenham, U.K.; Northampton, Mass.

Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empir. Econ. 39, 413–426. https://doi.org/10.1007/s00181-009-0310-x

Hicks, R.L., Strand, I.E., 2000. The extent of information: Its relevance for random utility models. Land Econ. 76, 374–385. https://doi.org/10.2307/3147035

Hilger, J.H., 2006. Heterogeneous preferences for water quality: a finite mixture model of beach recreation in Southern California.

Hindsley, P., Landry, C., Bin, O., Vogelsong, H., 2007. Site congestion in recreation choice models: A generated regressors approach to beach site selection.

Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. Ecol. Econ. 69, 1595–1603. https://doi.org/10.1016/j.ecolecon.2010.04.011

Kaoru, Y., 1995. Measuring marine recreation benefits of water quality improvements by the nested random utility model. Resour. Energy Econ. 17, 119–136. https://doi.org/10.1016/0928-7655(94)00021-B

Kipperberg, G., Onozaka, Y., Bui, L.T., Lohaugen, M., Refsdal, G., Sæland, S., 2019. The impact of wind turbines on local recreation: Evidence from two travel cost method – contingent behavior studies. J. Outdoor Recreat. Tour. 25, 66–75. https://doi.org/10.1016/j.jort.2018.11.004

Kleppe, A., Jensen, J., 2018. Valuing the Recreational Benefits og Bore and Hellestø Beaches.

Lancsar, E., Louviere, J., 2008. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. Pharmacoeconomics 26, 661–77.

Leggett, C., Scherer, N., Curry, M., Bailey, R., Haab, T.C., 2014. Assessing the Economic Benefits of Reductions in Marine Debris: a pilot study of beach recreation in Orange County, California. Industrial Economics, Incorporated.

LePlat, M., Ropars-Collet, C., Le Goffe, P., 2018. Regulating congestion with prices and non-prices measures: an application of a repeated random utility model to outdoor recreation. Presented at the 6. World Congress of Environmental and Resource Economists (WCERE), Gothenburg.

Lew, D.K., Larson, D.M., 2008. Valuing a Beach Day with a Repeated Nested Logit Model of Participation, Site Choice, and Stochastic Time Value. Mar. Resour. Econ. 23, 233–252. https://doi.org/10.1086/mre.23.3.42629616

Lew, D.K., Larson, D.M., 2005. Valuing Recreation and Amenities at San Diego County Beaches. Coast. Manag. 33, 71–86. https://doi.org/10.1080/08920750590883079

Lindhjem, H., 2011. Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. Int. Rev. Environ. Resour. Econ. 5, 309–351. https://doi.org/10.1561/101.00000045

Massey, D.M., Parsons, G.R., 2007. State Dependence and Long Term Site Capital in a Random Utility Model of Recreation Demand (NCEE Working Paper Series No. 07–11). National Center for Environmental Economics, US Environmental Protection Agency.

Matthews, Y., Scarpa, R., Marsh, D., 2018. Cumulative attraction and spatial dependence in a destination choice model for beach recreation. Tour. Manag. 66, 318–328. https://doi.org/10.1016/j.tourman.2017.12.009

Murdock, J., 2006. Handling unobserved site characteristics in random utility models of recreation demand. J. Environ. Econ. Manag. 51, 1–25. https://doi.org/10.1016/j.jeem.2005.04.003

Navrud, S., 2001. Economic valuation of inland recreational fisheries: empirical studies and their policy use in Norway. Fish. Manag. Ecol. 8, 14.

Parsons, G.R., 2017. The Travel Cost Model, in: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation, The Economics of Non-Market Goods and Resources. Springer Netherlands, Dordrecht, pp. 269–329. https://doi.org/10.1007/978-94-007-0826-6_9

Parsons, G.R., Massey, D.M., Tomasi, T., 1999. Familiar and Favorite Sites in a Random Utility Model of Beach Recreation. Mar. Resour. Econ. 14, 299–315.

Parsons, G.R., Stefanova, S., 2009. Temporal Stability of Recreation Choices. Presented at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, p. 25.

Pendleton, L.; M., 2012. Size Matters: The Economic Value of Beach Erosion and Nourishment in Southern California. Contemp. Econ. Policy 30, 223–237. https://doi.org/10.1111/j.1465-7287.2011.00257.x

Phaneuf, D.J., Requate, T., 2017. Recreation, in: A Course in Environmental Economics: Theory, Policy, and Practice. Cambridge University Press, Cambridge, United Kingdon ; New York, NY, USA.

Schibevaag, T.A., 2016. Ruster opp Borestranda for fem millioner kroner. Stavanger Aftenbladet.

Sveen, G., 2018. Jærstrendene er nå av de mest besøkte naturattraksjonene i Norge. Stavanger Aftenbladet 3.

Timmins, C., Murdock, J., 2007. A revealed preference approach to the measurement of congestion in travel cost models. J. Environ. Econ. Manag. 53, 230–249. https://doi.org/10.1016/j.jeem.2006.08.002

von Haefen, R.H., Phaneuf, D.J., 2008. Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference approach. J. Environ. Econ. Manag. 56, 19–32. https://doi.org/10.1016/j.jeem.2008.01.002

Whitehead, J.C., Dumas, C.F., Herstine, J., Hill, J., Buerger, B., 2008a. Valuing Beach Access and Width with Revealed and Stated Preference Data. Mar. Resour. Econ. 23, 119–135. https://doi.org/10.1086/mre.23.2.42629607

Whitehead, J.C., Pattanayak, S.K., Houtven, G.L.V., Gelso, B.R., 2008b. Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science. J. Econ. Surv. 22, 872–908. https://doi.org/10.1111/j.1467-6419.2008.00552.x

Yeh, C.-Y., Haab, T.C., Sohngen, B.L., 2006. Modeling Multiple-Objective Recreation Trips with Choices Over Trip Duration and Alternative Sites. Environ. Resour. Econ. 34, 189–209. https://doi.org/10.1007/s10640-005-6205-1

A peer reviewed version of this article has been published in *Environmental and Resource Economics :*
Lopes, A.F., Kipperberg, G. Diagnosing Insensitivity to Scope in Contingent Valuation. *Environ Resource Econ* (2020).
https://doi.org/10.1007/s10640-020-00470-9

Diagnosing insensitivity to scope in contingent valuation

Ana Faria Lopes[a] and Gorm Kipperberg[b]

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 between the 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.

[a] UiS Business School, University of Stavanger, ana.f.lopes@uis.no.
[b] UiS Business School, University of Stavanger, gorm.kipperberg@uis.no.

This research was funded by the Norwegian Research Council project #255777 (COAST-BENEFIT: Ecosystem Service Valuation for Coastal Zone Management: From Promise to Practice). We are grateful for insightful feedback from Henrik Lindhjem, John Whitehead, and two anonymous reviewers. We also appreciate comments from participants of the 2018 WCERE and internal seminars at the University of Stavanger and Appalachian State University.

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, page 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).

[1]This paper focus on the concept of *scope sensitivity* rather than *embedding*. Embedding implies comparing welfare measures for a good on its own with its value when evaluated as part of a package of different goods (Kahneman and Knetsch, 1992).

[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).

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; Desvougues 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, page 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; Arrow et al., 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, page 123).

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

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

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. (2017).[4]

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; Desvougues 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

[4] 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. (2017) 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.

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., 2018), or not included a scope test (e.g., Loureiro et al., 2009). To our knowledge, only Desvougues 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) 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*.

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., 2017)
	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)

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 while utility functions that are not directionally bounded (such as the Cobb-Douglas) always yield scope sensitivity, directionally bounded utility functions, which also satisfy all the preference axioms, can in turn 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 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 in terms of individual characteristics 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.[1] 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 explanation 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).

[1] 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.

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.

At the estimation stage, the **statistical distribution assumption** may affect scope sensitivity findings (Whitehead et al., 1998; Berrens et al., 2000; Borzykowski et al., 2017). Borzykowski et al. (2017) and Whitehead et al. (1998) find that parametric estimates based on a single-bounded dichotomous choice question is more likely to fail a statistical scope test than non-parametric estimates, spike models, or estimates based on a double-bounded dichotomous choice question.

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 study 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) and a damage table (Appendix 2). 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.[1]

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.[2] 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.[3]

[1] 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.

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

[3] 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.

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 minutes) 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.[4] 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.[5]

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).

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.

[4] 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 and survey inconsequentiality (see Table 2 for details), 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.*

[5] We have tested whether recoding uncertain yes-respondents impacts the results. The scope findings summarized in sections 4.2 and 4.3 remain unchanged.

4 Lofoten Scope Analysis

We use the CV survey data described above to conduct an *ex post* exploration of ten of the scope insensitivity explanations 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 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 explanation 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 Figures 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 (Figure 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 (Figure 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.

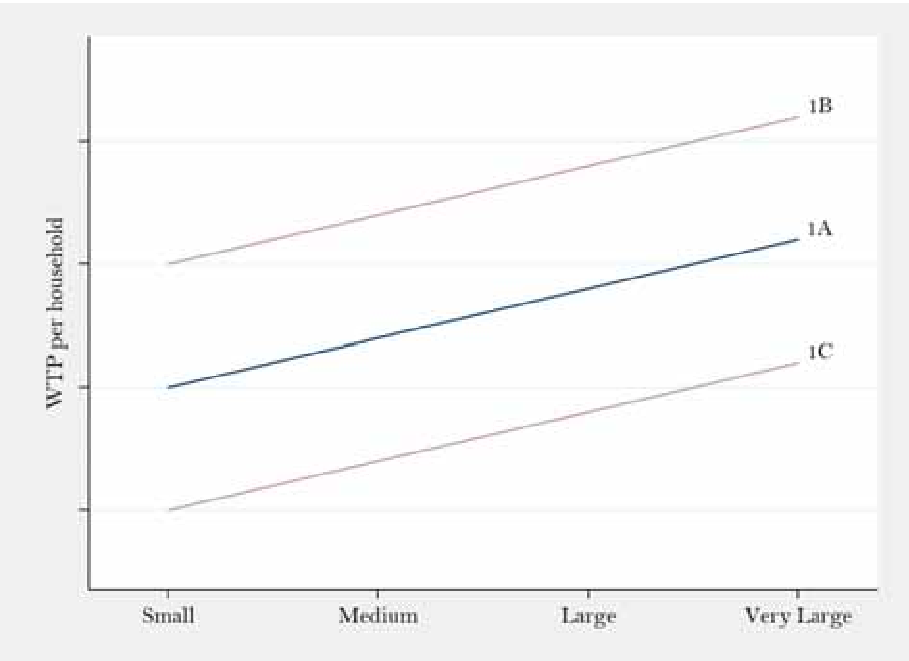


Figure 1 - Conceptualization of no impact on sensitivity to scope

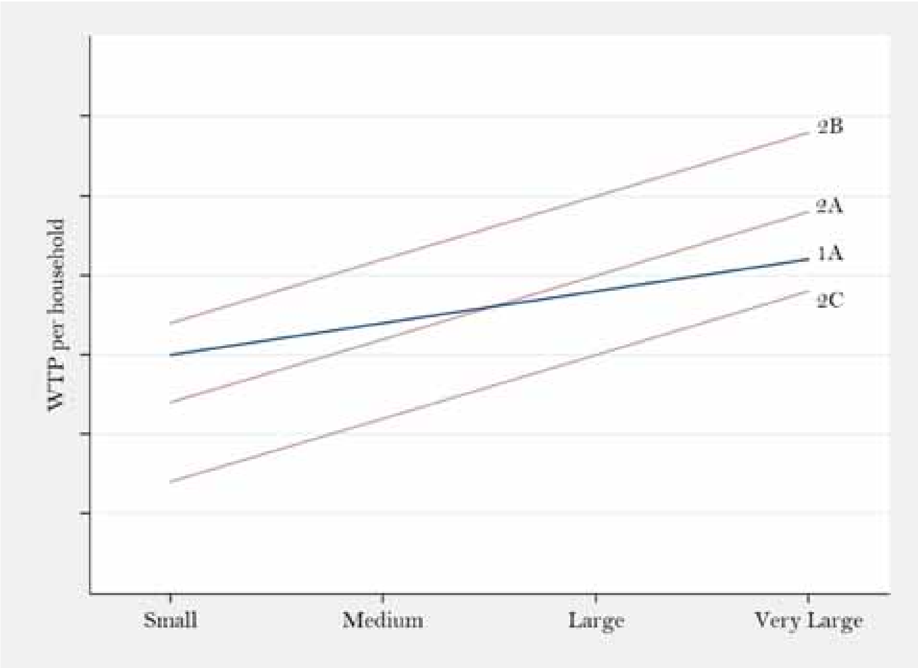


Figure 2 - Conceptualization of impact on sensitivity to scope

In our analysis below, we assess the presence or absence of sensitivity to scope in the following three-fold 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 convolution* and summarized with P-values under the null hypothesis of insensitivity to scope.[1] The total scope convolution test is constructed as

[1] 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, 5,000 replications are generated for each WTP (Jeanty, 2007).

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) as a proxy variable.[2]

4.2 Baseline results

Parametric and non-parametric baseline results are summarized in Figures 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) and utilized 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 Figure 3. Regression outputs are provided in Appendix 4.

[2] The scope elasticity results presented in Section 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).

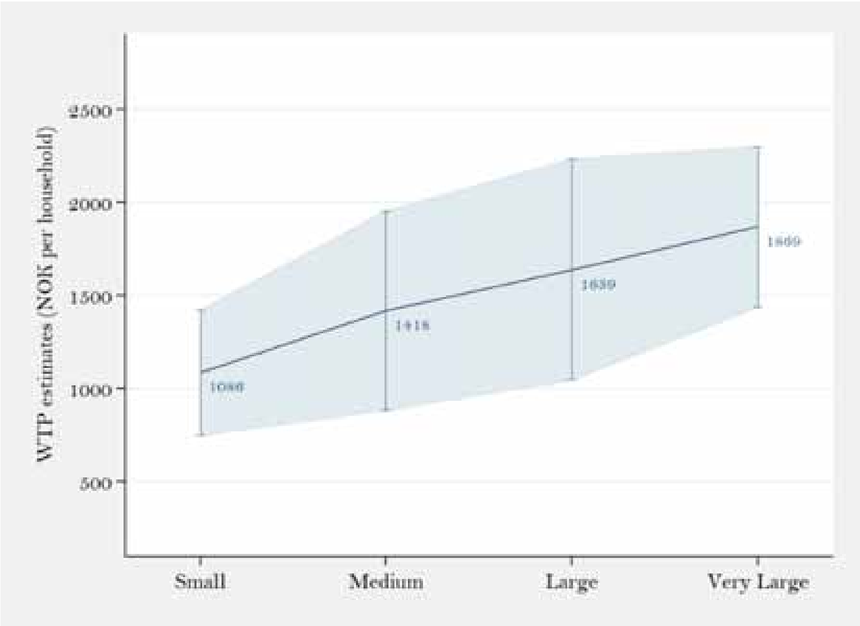


Figure 3 - Baseline parametric WTP welfare estimates (mean estimates and 95% confidence intervals)

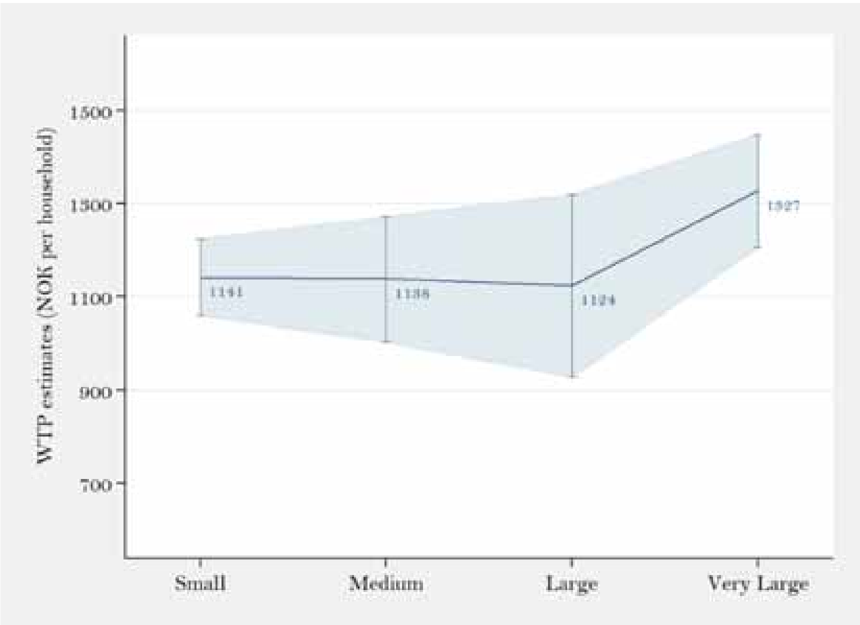


Figure 4 - Non-parametric welfare estimates (mean estimates and 95% confidence intervals)

The estimated annual household WTPs are NOK 1086, 1418, 1639, and 1869 to prevent a small, medium, large, and very large oil spill, respectively.[3] 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 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 three-fold 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 Figure 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 1141 to prevent the smallest oil spill and NOK 1327 to prevent the largest oil spill. As seen in Figure 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

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

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 Section 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 improves scope inference.

Third, we perform *ex post* power calculations to inform the extent to which the quadruple split-sample CV experiment had sufficient power to find full scope sensitivity in the piecewise linear specification we employ.[4] This relates to the issue of **sample size** in scope sensitivity testing. In general, the power of an experiment is "the probability that, for a given size and a given statistical significance level, we will be able to reject the hypothesis of zero effect" (Duflo et al., 2007). We find that only the small versus very large oil spill comparison has sufficient power (87%) under a conventional power threshold of 80% (Duflo et al., 2007). However, these results do not undermine our exploration of scope insensitivity explanations below. Rather, it can be argued that low power strengthens the analysis for two reasons: 1) Modest sub-sample sizes make it more difficult to pass statistical scope tests as pointed out from the literature review. Hence, it becomes less likely to uncover explanations that influence the scope inference. 2) With knowledge of the power of the small versus large oil spill comparisons, the partial

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

scope tests and partial elasticity estimates become more important for the overall scope-impact inferences below.

4.3 Sensitivity to scope analysis

Results from the scope diagnostics are presented in Figures 5-12. 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 & 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 describes how the control variable was constructed for each sensitivity analysis and reports corresponding convolution tests and scope elasticity estimates.

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 to 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 & 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 of my household tend to give 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

In summary, accounting for amenity misspecification (Figure 5), and imposing diminishing marginal utility (Figure 6) have positive impacts on the scope inference. In contrast, accounting for preference heterogeneity (Figure 7), consequentiality (Figure 8), experience, familiarity, knowledge and/or use (Figures 9 and 10), controlling for incomplete multi-stage budgeting (Figure 11), and warm glow (Figure 12) do not appear to affect statistical scope inference in our case. Interestingly, several of the explorations imply differences in mean welfare estimates. For example, controlling for perceived consequentiality (Figure 8) and prior recreational use of the Lofoten Archipelago (Figure 9) lead to higher WTP estimates. Next we discuss each exploration in further details. The underlying regression results are provided in Appendix 4.

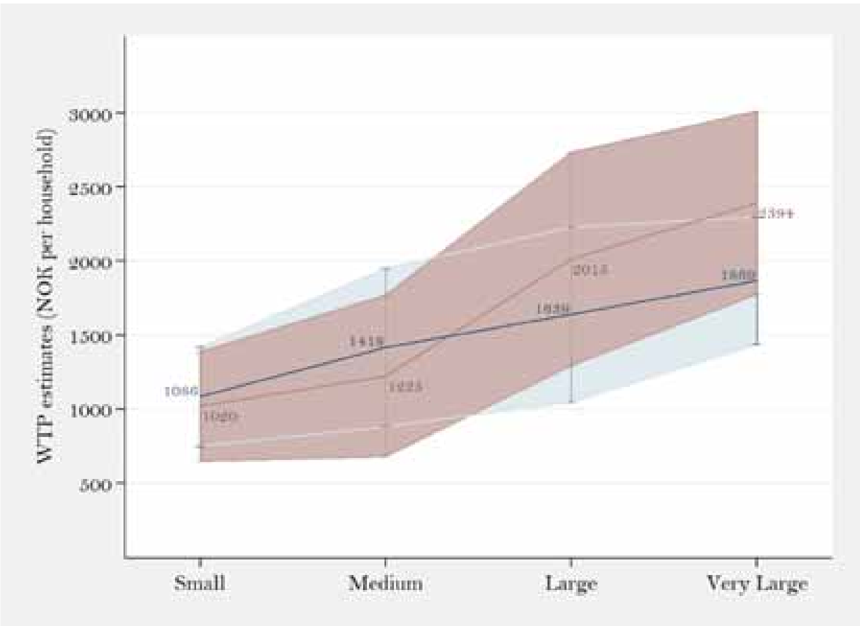


Figure 5 - Controlling for subjective oil spill probabilities (amenity misspecification)

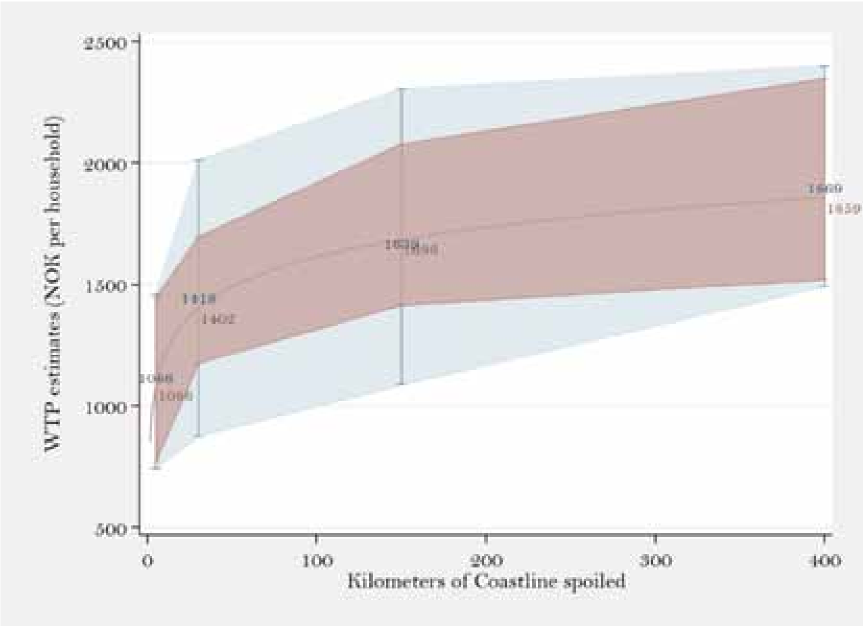


Figure 6 - Controlling for diminishing marginal utility

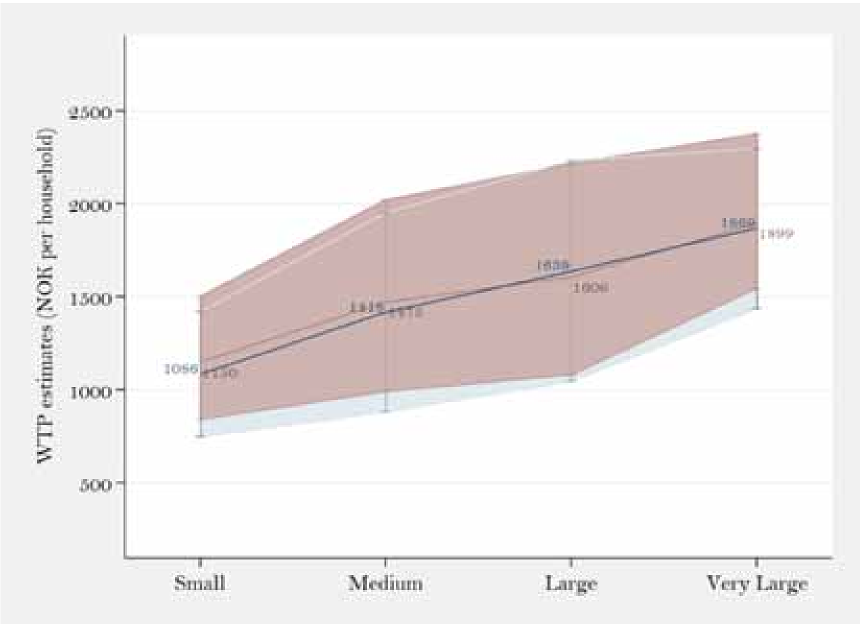


Figure 7 - Controlling for subjective importance of preventing oil spills (preference heterogeneity)

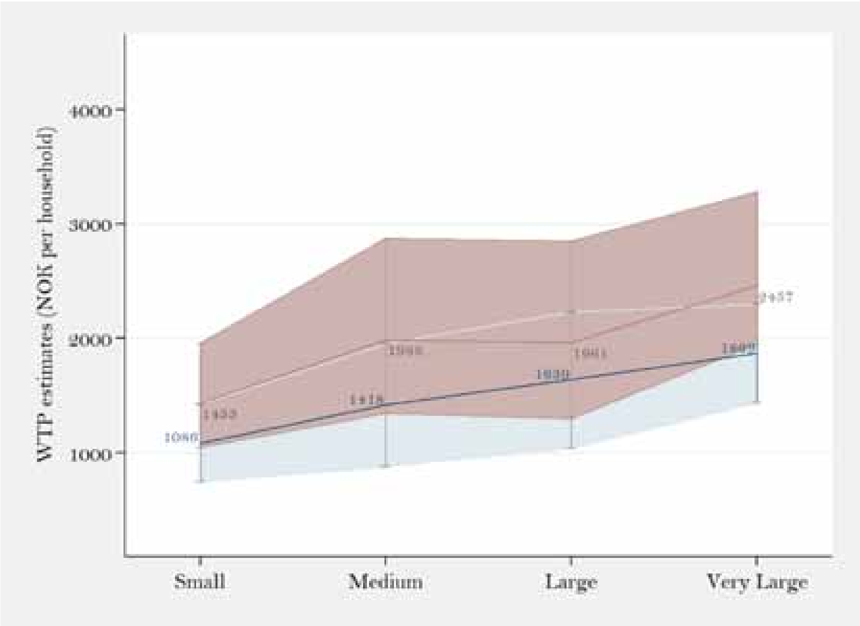


Figure 8 - Controlling for consequentiality (survey design)

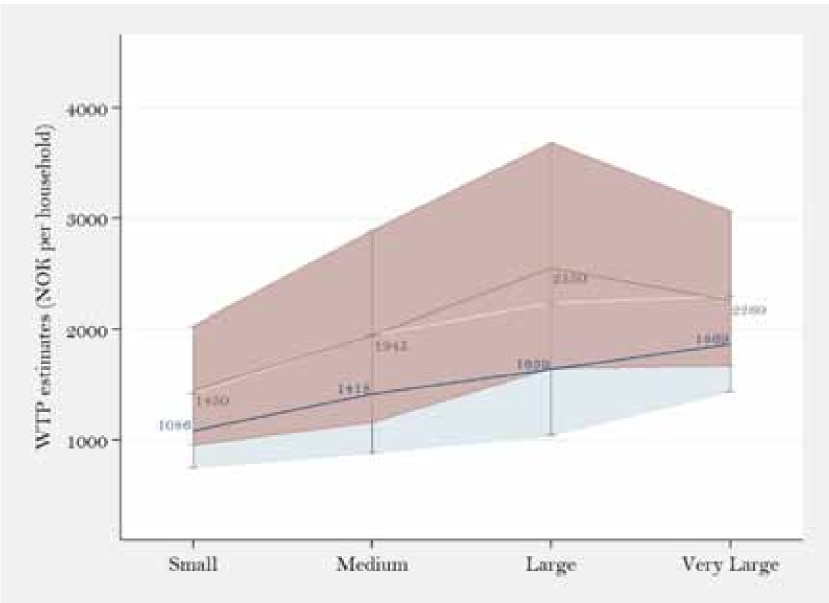


Figure 9 - Controlling for prior use of the Lofoten Archipelago (experience, familiarity, knowledge and/or use)

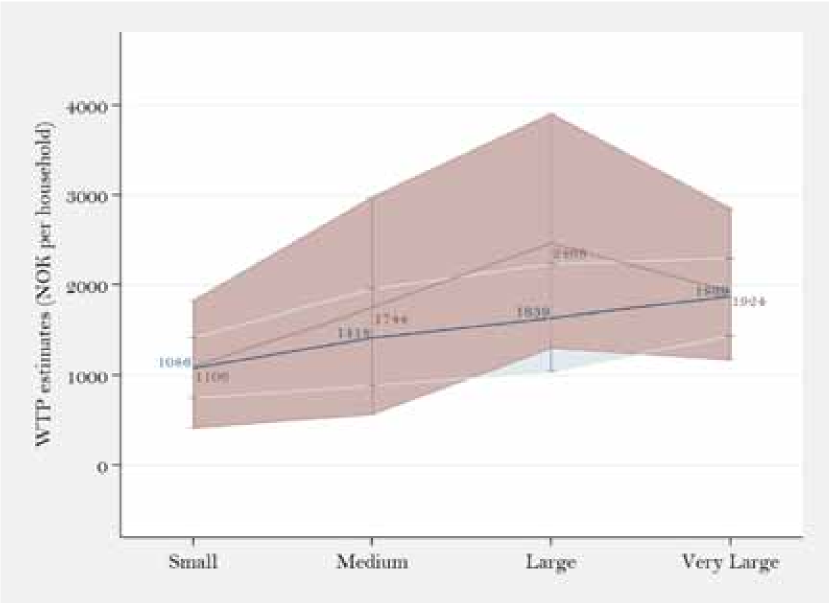


Figure 10 - Controlling for previous experience with oil spills (experience, familiarity, knowledge and/or use)

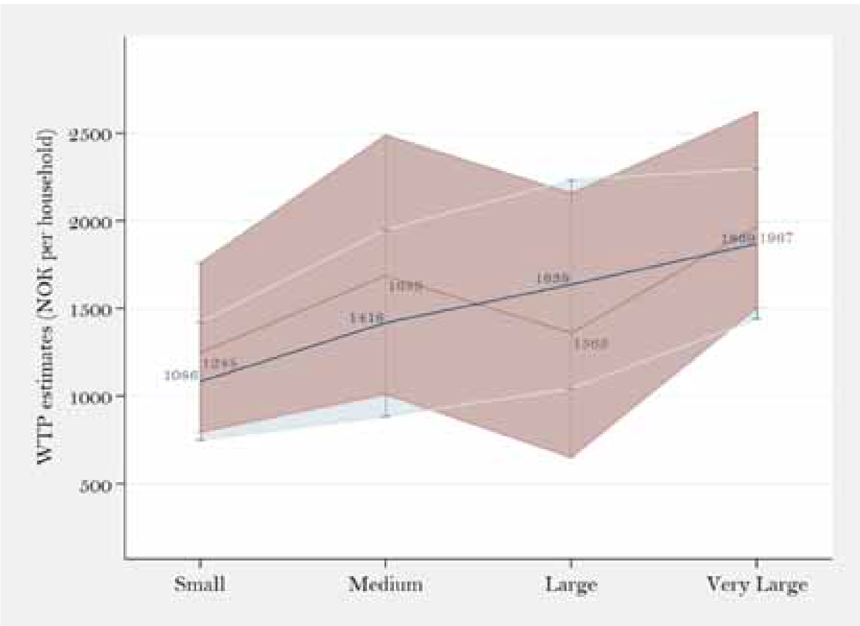


Figure 11 - Controlling for budget constraints

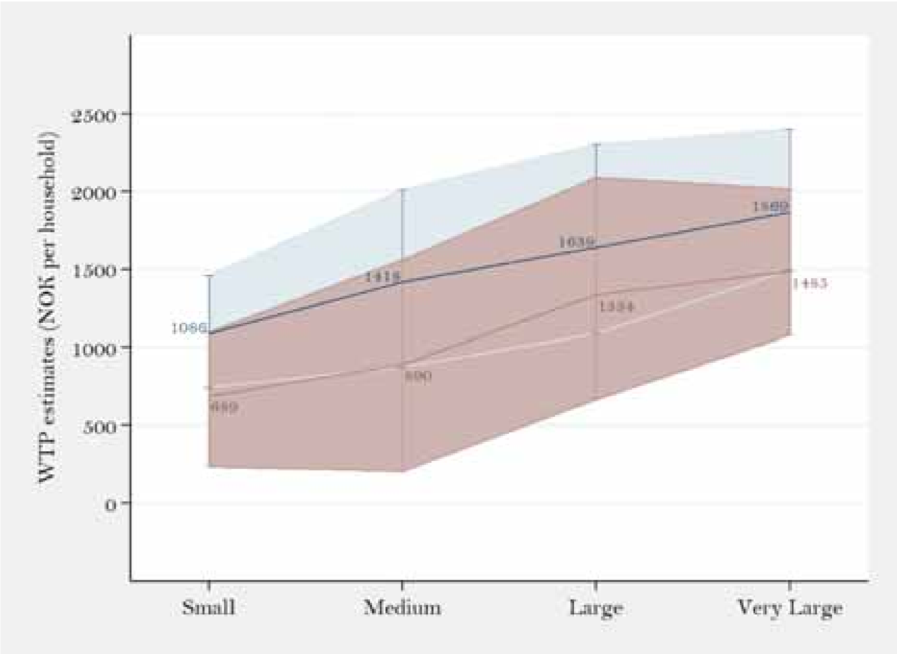


Figure 12 - Controlling for altruistic motivations (warm glow)

Controlling for perceived oil spill probabilities (Figure 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 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. As seen in Figure 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.

Controlling for diminishing marginal utility (Figure 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 piece-wise linear specification. As seen in Appendix 2, the small oil spill implies 5 kilometers of coastline soiled, while the very large oil spill implies 400 kilometers. The empirical scope line in red in Figure 7 is produced from the specification with the logarithmic size variable as reported in Appendix 4. 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 (*versus* 0.1446 for the baseline).

Controlling for preference heterogeneity (Figure 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.[1] 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).

Controlling for consequentiality (Figure 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 Figure 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).[2]

Controlling for prior use of the Lofoten Archipelago (Figure 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. (1998) as a factor likely to influence scope sensitivity. We explore this hypothesis by interacting the size dummies with an indicator for use

[1] 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.

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

of the Lofoten Archipelago, defined as having visited or residing there. As can been seen in Figure 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.

Controlling for previous experience with oil spill (Figure 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 indicated by the test statistics in Table 2, this does not improve the scope inference or lead to statistically different WTP estimates.

Controlling for incomplete multi-stage budgeting (Figure 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 25^{th} percentile of the income distribution). We hypothesize that respondents less constrained by income would be more sensitive to scope. However, as seen in Figure 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).

Controlling for warm glow (Figure 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 Figure 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; Desvougues 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 1086 and NOK 1869, 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 improves 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 explanations 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; Desvougues et al., 2012; Whitehead, 2016: Johnson 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.

References

Alevy, J., List, J. and Adamowicz, W. (2011). How Can Behavioral Economics Inform Nonmarket Valuation? An Example from the Preference Reversal Literature. Land Economics, 87(4).

Amiran, E. and Hagen, D. (2010). The scope trials: Variation in sensitivity to scope and WTP with directionally bounded utility functions. Journal of Environmental Economics and Management, 59(3), pp.293-301.

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(10), 4601-4614.

Arrow, K., Leamer, E., Schuman, H., Solow, R. (1994). Appendix D in "Comments on Proposed NOAA/DOI Regulations on Natural Resource Damage Assessment." U.S. Environmental Protection Agency (October).

Banerjee, S., & Murphy, J. H. (2005). The scope test revisited. Applied Economics Letters, 12(10), 613-617.

Barton, D., Navrud, S., Siripala, N., Gunawardane, P., Perera, S. & Perera, H. D. B. S. (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. INSTCOM-Report 05-2003 (NIVA Report 5396-2007).

Bateman, I. J., Cole, M., Cooper, P., Georgiou, S., Hadley, D., & Poe, G. L. (2004). On visible choice sets and scope sensitivity. Journal of environmental economics and management, 47(1), 71-93.

Berrens, R. P., Bohara, A. K., Silva, C. L., Brookshire, D., & Mckee, M. (2000). Contingent values for New Mexico instream flows: With tests of scope, group-size reminder and temporal reliability. Journal of Environmental Management, 58(1), 73-90.

Bishop, R.C., Boyle, K.J., Carson, R.T., Chapman, D., Hanemann, W.M., Kanninen, B., Kopp, R.J., Krosnick, J.A., List, J., Meade, N.

& Paterson, R. (2017). Putting a value on injuries to natural assets: The BP oil spill. Science, 356(6335), 253-254.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). Introduction to meta-analysis. John Wiley & Sons.

Borzykowski, N., Baranzini, A., & Maradan, D. (2018). Scope Effects in Contingent Valuation: Does the Assumed Statistical Distribution of WTP Matter?. Ecological Economics, 144, 319-329.

Boyle, K. J., Desvousges, W. H., Johnson, F. R., Dunford, R. W., & Hudson, S. P. (1994). An investigation of part-whole biases in contingent-valuation studies. Journal of Environmental Economics and Management, 27(1), 64-83.

Burrows, J., Newman, R., Genser, J., & Plewes, J. (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 Contingent Valuation of Environmental Goods. Edward Elgar Publishing.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. Journal of environmental economics and management, 15(3), 355-379.

Carson, R. T. (1997). Contingent valuation surveys and tests of insensitivity to scope. In Determining the value of non-marketed goods (pp. 127-163). Springer, Dordrecht.

Carson, R. T. (2012). Contingent valuation: a comprehensive bibliography and history. Edward Elgar Publishing.

Carson, R. T., Conaway, M. B., Hanemann, W. M., Krosnick, J. A., Mitchell, R. C., & Presser, S. (2013). Valuing oil spill prevention: a case study of California's central coast (Vol. 5). Springer Science & Business Media.

Carson, R. T., Flores, N. E., & Meade, N. F. (2001). Contingent valuation: controversies and evidence. Environmental and resource economics, 19(2), 173-210.

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

Carson, R. T., & Mitchell, R. C. (1995). Sequencing and nesting in contingent valuation surveys. Journal of environmental economics and Management, 28(2), 155-173.

Carson, R. T., Mitchell, R. C., Hanemann, M., Kopp, R. J., Presser, S., & Ruud, P. A. (2003). Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. Environmental and resource economics, 25(3), 257-286.

Casey, J. F., Kahn, J. R., & Rivas, A. A. (2008). Willingness to accept compensation for the environmental risks of oil transport on the Amazon: A choice modeling experiment. Ecological Economics, 67(4), 552-559.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. Agricultural and Resource Economics Review, 38(2), 166-180.

Czajkowski, M., & Hanley, N. (2009). Using labels to investigate scope effects in stated preference methods. Environmental and Resource Economics, 44(4), 521.

Desvousges, W. H. (1992). Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy (Vol. 1992). Research Triangle Park, NC: Research Triangle Institute.

Desvousges, W., Mathews, K., & Train, K. (2012). Adequate responsiveness to scope in contingent valuation. Ecological Economics, 84, 121-128.

Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. Journal of Environmental Economics and Management, 30(3), 337-347.

Diamond, P. A., & Hausman, J. A. (1994). Contingent valuation: is some number better than no number?. Journal of economic perspectives, 8(4), 45-64.

Duflo, E., Glennerster, R., Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. Discussion Paper Series No. 6059.

Frederick, S. & Fischhoff, B. (1998). Scope (in) sensitivity in elicited valuations. Risk Decision and Policy, 3(2), 109-123.

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

Giguere, C., Moore, C., & Whitehead, J. C. (2020). Valuing Hemlock Woolly Adelgid Control in Public Forests: Scope Effects with Attribute Nonattendance. Land Economics, 96(1), 25-42.

Haab, T. C., Interis, M. G., Petrolia, D. R., & Whitehead, J. C. (2013). From hopeless to curious? Thoughts on Hausman's "dubious to hopeless" critique of contingent valuation. Applied Economic Perspectives and Policy, 35(4), 593-612.

Hausman, J. (2012). Contingent valuation: from dubious to hopeless. Journal of Economic Perspectives, 26(4), 43-56.

Heberlein, T. A., Wilson, M. A., Bishop, R. C., & Schaeffer, N. C. (2005). Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management, 50(1), 1-22.

Jeanty, P. W. (2007). Constructing krinsky and robb confidence intervals for mean and median willingness to pay (wtp) using stata. In Sixth North American Stata Users' Group Meeting, Boston, August (pp. 13-14).

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

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: the purchase of moral satisfaction. Journal of environmental economics and management, 22(1), 57-70.

Kaltenborn, B. P., & Linnell, J. D. (2019). Who are legitimate stakeholders? National and local perceptions of environmental change in the Lofoten islands, Norway. Polar Geography, 42(4), 236-252.

Kling, C. L., Phaneuf, D. J., & Zhao, J. (2012). From Exxon to BP: Has some number become better than no number?. Journal of Economic Perspectives, 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 Economics, 79(4), 483-499.

Krinsky, I & Robb, A. L. (1986). "On Approximating the Statistical Properties of Elasticities." Review of Economic and Statistics, 68, 715-719.

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

Loomis, J. B. (2014). 2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys. Journal of Agricultural and Resource Economics, 34-46.

Lopez-Feldman, A. (2013). SINGLEB: Stata module to compute Contingent Valuation using Single-Bounded Dichotomous Choice. Statistical Software Components.

Loureiro, M. L., Loomis, J. B., & Vázquez, M. X. (2009). Economic valuation of environmental damages due to the Prestige oil spill in Spain. Environmental and Resource Economics, 44(4), 537-553.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). Microeconomic theory (Vol. 1). New York: Oxford university press.

Mitchell, R. C., & Carson, R. T. (1989). Using surveys to value public goods: the contingent valuation method. Resources for the Future.

Navrud, S., Lindhjem, H., & Magnussen, K. (2017). Valuing Marine Ecosystem Services Loss from Oil Spills for Use in Cost-Benefit Analysis of Preventive Measures. Handbook on the economics and management of sustainable oceans, 124-137.

Ojea, E., & Loureiro, M. L. (2011). Identifying the scope effect on a meta-analysis of biodiversity valuation studies. Resource and Energy Economics, 33(3), 706-724.

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

Poe, G. L. (2016). Behavioral Anomalies in Contingent Values and Actual Choices. Agricultural and Resource Economics Review, 45(2), 246-269.

Randall, A., & Hoehn, J. P. (1996). Embedding in market demand systems. Journal of Environmental Economics and Management, 30(3), 369-380.

Rollins, K., & Lyke, A. (1998). The case for diminishing marginal existence values. Journal of Environmental Economics and Management, 36(3), 324-344.

Rowe, *R.D.*, Shaw, W.D. and Schulze, W. (*1992*), *Nestucca oil spill*, in K.M. Ward and J.W. Duffield (eds.), *Natural Resource Damages*: *Law and Economics*, Wiley, New York, pp. *527–554*.

Siikamäki, J., & Larson, D. M. (2015). Finding sensitivity to scope in nonmarket valuation. Journal of Applied Econometrics, 30(2), 333-349.

Smith, V. K., & Osborne, L. L. (1996). Do contingent valuation estimates pass a “scope” test? A meta-analysis. Journal of Environmental Economics and Management, 31(3), 287-301.

Spencer-Cotton, A., Kragt, M. E., & Burton, M. (2018). Spatial and scope effects: valuations of coastal management practices. Journal of agricultural economics, 69(3), 833-851.

Søgaard, R., Lindholt, J., & Gyrd-Hansen, D. (2012). Insensitivity to Scope in Contingent Valuation Studies. Applied health economics and health policy, 10(6), 397-405.

Urminsky, O., & Kivetz, R. (2011). Scope insensitivity and the “mere token” effect. Journal of Marketing Research, 48(2), 282-295.

Van Biervliet, K., Le Roy, D., & Nunes, P. A. (2006). An accidental oil spill along the Belgian Coast: Results from a CV Study.

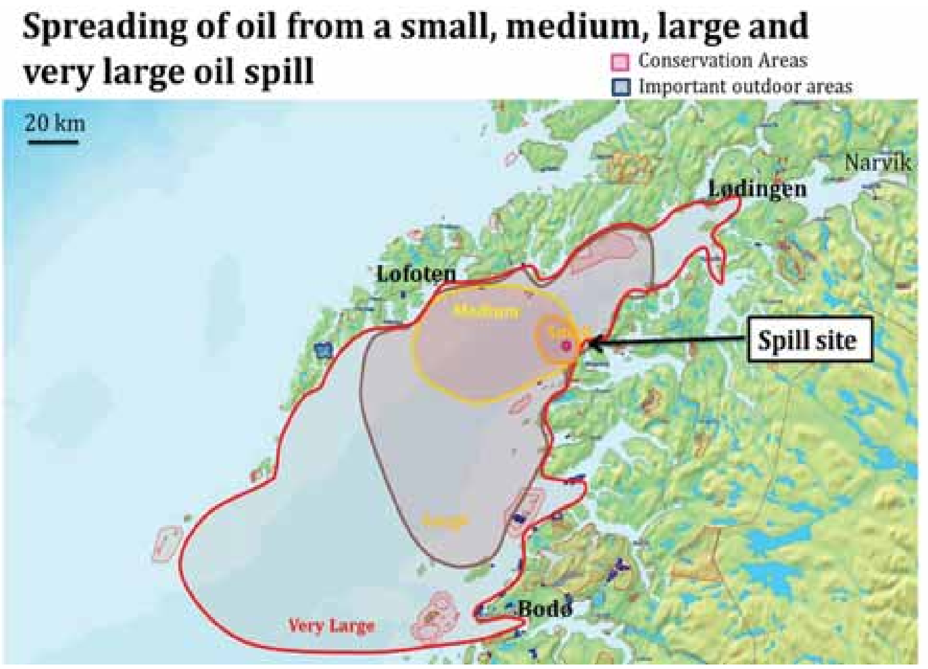
Varian, H. R. (2014). Intermediate Microeconomics: A Modern Approach: Ninth International Student Edition. WW Norton & Company.

Whitehead, J. C. (2016). Plausible responsiveness to scope in contingent valuation. Ecological Economics, 128, 17-22.

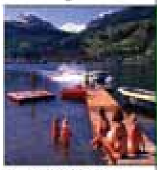
Whitehead, J. C., Haab, T. C., & Huang, J. C. (1998). Part-whole bias in contingent valuation: Will scope effects be detected with inexpensive survey methods?. Southern Economic Journal, 160-168.

Appendices

Appendix 1 – Oil Spill Dispersion Map



Appendix 2 – Damage Table

	Present Conditions	Small Oil Spill	Medium Oil Spill	Large Oil Spill	Very Large Oil Spill
Damage to birds					
	Area as important nesting, migration and wintering area for seabirds. There was a decline in seabird population last year. But stocks remain in good condition	Bird population mainly in good condition In total 1 500 dead birds	Guillemots locally endangered Other stocks return to normal after 1 year In total 15 000 dead birds	Population of kittiwakes and guillemots locally endangered Other stocks return to normal after 2 years In total 50 000 dead birds	Population of kittiwakes and guillemots endangered in Norway Other stocks return to normal after four years In total 120 000 dead birds
Damage to seals					
	Area very important for seals Seal population in good condition	Seal population in good condition In total 30 dead seals	Seal population in good condition In total 100 dead seals	Population of arctic seals locally endangered In total 500 dead seals	Population of arctic seals endangered in Norway Other species back to normal after four years In total 1 000 dead seals
Damage to coast					
	Rich ravines and deepwater corals Area important for recreation and outdoor activities for both residents and visitors	5 km of shoreline (cliffs and beaches) soiled by oil Affected land-based and water-based outdoor recreation Affected areas operate normally after 6 months	30 km of shoreline (cliffs and beaches) soiled by oil Affected land-based and water-based outdoor recreation Affected areas operate normally after 1 year	150 km of shoreline (cliffs and beaches) soiled by oil Affected land-based and water-based outdoor recreation Affected areas operate normally after 3 years	400 km of shoreline (cliffs, beaches and fishing villages) contaminated with oil Affecting land-based and water-based outdoor recreation Affected areas operate normally after 5 years
Damage to other marine life					
	The area is important spawning and internationally important nursery and feeding grounds for several fish species.	Can be harvested as before Safe for human consumption Spawning and nursery areas not affected	Can be harvested as before Safe for human consumption Spawning and nursery areas return to normal after one year	Fish, shellfish, shells and seaweed not safe within three years after the spill Spawning and nursery areas return to normal after three years	Fish, shellfish, shells and seaweed not safe within 5 years after the spill Spawning and nursery areas return to normal after 5 years

Appendix 3 – Sample Representativeness

	Mean (Population Statistics)[1]	Mean (Full Sample N=1401)	Mean (Reduced sample; N=863)	Small Oil Spill (N=314)	Medium Oil Spill (N=128)	Large Oil Spill (N=116)	Very Large Oil Spill (N=295)
Age	39.31	47.02	47.11	46.28	49.13	49.5	46.19
Gender (1=male)	0.50	0.49	0.49	0.51	0.52	0.38	0.50
Household Income (NOK)	730 800	718 712	710 375	718 949	730 468	676 724	705 762
Number of people per Household	2.21	2.46	2.40	2.48	2.28	2.34	2.40
Education level (%)							
Basic school level	0.28	0.06	0.03	0.03	0.03	0	0.03
Upper secondary education	0.42	0.35	0.32	0.33	0.26	0.31	0.34
Tertiary education (up to 4 years)	0.22	0.41	0.42	0.43	0.49	0.37	0.42
Tertiary education (5 years or more)	0.08	0.18	0.18	0.18	0.17	0.23	0.17

[1] Year = 2013. Data from Statistics Norway (www.ssb.no)

Appendix 4 – Regression results for baseline estimation and sensitivity analyzes (in WTP space)

	Baseline estimation	**Diminishing marginal utility**	**Incomplete multi-stage budgeting**	**Experience, familiarity, knowledge and/or use**		**Preference Hetero-geneity**	**Amenity Misspeci-fication**	**Survey Design**	**Warm Glow**
				Experience	*Visitor*			*Consequentiality*	
Medium	332.50		114.98	245.95	-43.06	764.23	-194.14	-136.98	829.71
Large	553.78		1327.33*	281.49	631.28	1114.14	-6.42	1623.63	536.95
Very Large	783.52**		872.86	769.97**	688.30	1369.14	875.89*	624.01	985.84
Interaction term (Small)			459.03	29.59	528.93	1256.83***	1997.11**	3819.31***	-1932.96**
Interaction term (Medium)			784.88	422.11	1069.35	1158.65***	3011.56**	4510.90**	-2561.60**
Interaction term (Large)			-753.73	1110.63	986.13	1110.49**	4546.72**	2723.33*	-1825.12*
Interaction term (Very Large)			305.02	77.31	650.66	1118.87***	3265.36**	4218.87***	-2122.87***
Log of Coastline affected		176.35**							
Gender	87.79	90.86	59.74	98.86	-83.63	396.79*	343.03	158.65	181.19
Age	18.93**	18.88**	19.56**	18.77**	11.85	16.59**	19.44**	22.99**	19.86**
Member_EnvOrg	1488.71**	1491.28**	1489.90**	1456.53**	1429.49**	1143.73**	1097.77*	1304.77*	1317.98*
Income	0.0003	0.0003	0.00003	0.0003	0.0003	0.0003	0.0003	0.0003	.0002
Constant	-150.46	-434.57	-274.72	-140.89	123.01	-5777.66***	-1124.52	-3842.98***	1362.74
# of Observations	853	853	853	853	794	844	708	853	853
Log likelihood Value	-531.55	-531.57	-529.16	-529.90	-484.26	-474.37	-407.59	-477.27	-503.83
AIC	1081.10	1077.13	1084.32	1085.79	994.5227	974.74	841.18	980.54	1033.66
BIC	1123.84	1110.37	1146.06	1147.53	1055.325	1036.34	900.49	1042.27	1095.39

Notes: All estimations were performed in STATA 14 using the singleb command (Lopez-Feldman, A., 2011). *** denotes statistical significance at 0.001; ** at 0.01; * at 0.05.

Estimating the Ex-ante Recreational Loss of an Oil Spill using Revealed Preference Site Selection and Multinomial Stated Preference Data

Ana Faria Lopes[a] and John C. Whitehead[b]

Abstract: This paper combines revealed preference and stated preference data to estimate the recreational impact of four hypothetical oil spills on the Jæren beaches in Norway. Our application is the first to consider two simultaneous changes due to an oil spill: we hypothesize that an oil spill not only reduces the choice set available to recreationists but also reduces the perceived beach site quality. Our application highlights the gains in combining stated and revealed preferences. Including variation in perceived site quality due to an oil spill and allowing for substitution to other recreational sites is only possible due to the inclusion of stated preference data. We conclude that the separate estimation of either SP or RP models results in misspecification due to the inability to estimate all parameters that drive site choice. When combining data sources, we estimate a welfare loss ranging from 122 to 288 NOK across the four oil spill scenarios. We show that omitting perceived site quality when using revealed preference data leads to low welfare losses, while naively omitting alternative specific constants when using stated preference data leads to high welfare losses.

[a] UiS Business School, University of Stavanger; ana.f.lopes@uis.no
[b] Department of Economics, Appalachian State University; whiteheadjc@appstate.edu

This research was funded by the Norwegian Research Council in the context of the COAST-BENEFIT project (project number 255777). We thank participants at a Brown Bag Seminar at UNC Greensboro and the poster session at the 2019 AERE Summer Conference. The corresponding author would like to thank Gorm Kipperberg and Maria Loureiro for their insightful comments, as well as John Rolfe for his input at the survey design stage.

1 Introduction

Coastal areas, especially those near heavy oil tanker traffic or oilrigs, are under increasing pressure due to economic activity offshore. One of these threats is increased risk of oil spill accidents, whose consequences to marine and coastal habitats are numerous, implying both use and non-use value losses. Numerous studies estimate the losses due to oil spill accidents in terms of non-use (e.g., Bishop et al., 2017; Carson et al., 2003) or use values (e.g., Winkler and Gordon, 2013). Changes in use values typically employ revealed preference (RP) methods by applying the travel cost method. This method is especially relevant in settings where the losses or gains in terms of recreational value are likely to be substantial (e.g., Alvarez et al., 2014).

Two different approaches have been used to retrieve estimates of the impact of an oil spill on beach recreation in the context of a multi-site model. The first approach is to calculate the number of lost recreational trips (or days) and multiply them by the value of a beach trip or day (English et al., 2018; Glasgow and Train, 2018). The second approach is to estimate the change in welfare per trip due to the presence of an oil spill and multiply it by the number of total trips (Alvarez et al., 2014; Hausman et al., 1995). Both of these approaches focus on the impact of an oil spill ex post.

Neither of these approaches is adequate if the research contemplates the ex-ante impact of an oil spill. Parsons (2008) is the only study to estimate the ex-ante impact of an oil spill. He assumes that the loss from an oil spill is a result of a reduction in the choice set of recreational sites, as previously available recreational sites are now soiled and closed. However, this approach implicitly assumes that if the beach is not closed, individuals would not change their behavior and thus incur no loss. Yet, even in the event of no beach closure, preferences towards oil spill avoidance might lead individuals to opt-out from engaging in beach recreation all-together or decide instead to engage in a completely different recreational activity. This type of behavior can be elicited using contingent behavior (CB) questions. Contingent behavior involves asking individuals how they would change their behavior given an hypothetical scenario on overall site quality (e.g., Landry et al., 2012), attributes (e.g., Adamowicz et al., 1997), travel cost (e.g., Azevedo et al., 2003), or access to sites (e.g., Grijalva et al., 2002).

Given the dual potential of observed and stated behavior to study preferences towards oil spill avoidance, one could combine the two types of behavior data. Combining revealed preference (RP) and stated preference (SP) data is appealing because RP-SP data are complementary in regards to their weaknesses (Whitehead et al., 2008). RP data is grounded on the individuals' observed choices, whereas SP data is criticized due to its hypothetical nature (Scott, 1965). SP data can be collected with an experimental design to introduce variation in attribute levels while modeling RP data is challenging due to the multicollinearity in the attribute data. In recognition of their duality, the number of applications of joint RP-SP data grew, especially after Adamowicz et al. (1994). The combination of the RP-SP data generally results in a better fit of the models and the possibility to estimate welfare losses from attributes that would not be possible from using either dataset separately. However, there may be differences across RP-SP datasets that should be accounted for, for example if the respondents are more uncertain when facing SP scenarios (Whitehead and Lew, 2019).

In light of the above discussion, this research paper aims at combining RP-SP data to estimate the recreational impact of four hypothetical oil spills in Norway. Our application is the first to consider two simultaneous changes due to an oil spill: a reduction in the available choice set of recreationists, and in the perceived beach site quality. We estimate a welfare loss ranging from 122 to 288 NOK per choice occasion across the four oil spill scenarios. We conclude that separate estimation of either RP or SP data leads to misspecification in our context since not all variables that drive site choice can be included. We also find that combining RP-SP data has several advantages in terms of welfare analysis: omitting perceived site quality when using RP data leads to low welfare losses, while omitting alternative specific constants when using SP data leads to high welfare losses. Combination of RP-SP data also allows for the estimation of additional parameters and accounts for scale heterogeneity.

An additional contribution of this paper is to expand the available choice set in the contingent behavior (CB) question. In previous CB surveys, individuals are given a limited choice set in the CB question: the option to visit one or more recreational sites within the same study area (e.g. Truong et al., 2018), the option to "stay at home" (Yi and Herriges, 2017), or the option of postponing the trip (Parsons and Stefanova 2011). Yet, when faced with an oil spill, an individual might opt for an activity or site that is not available in the CB question (e.g. a park, hike, forest). Stafford (2018) shows the importance of considering appropriate outside or opt-out options to obtain unbiased welfare estimates. We build on existing efforts by expanding the individual's choice set including other recreational sites and refer to this as "multinomial" CB data. This paper is the first to jointly estimate RP site selection data and multinomial CB data in a Random Utility Model framework.

To the best of our knowledge, this is the first RP-SP discrete choice model focusing on oil spill accidents. Previous RP-SP studies have considered the ex-ante impact of wind farms (Landry et al., 2012), water flow (Loomis, 1997), or forest fires (Simões et al., 2013). Despite our focus on oil spill impacts, our discussion applies to any future threat that changes the quality, travel cost or available choice set of goods and services.

The remainder of this paper is structured as follows. Section 2 discusses the overall motivation of the paper and the chosen approach. Section 3 presents the case study and survey data. Section 4 reports the estimation results and welfare estimates. Section 5 discusses the results. Section 6 concludes.

2 Methods

Any losses arising due to an oil spill accident entail both use and non-use value losses. If non-use value losses comprise the majority of the welfare loss at stake, researchers should use SP methods. To this end, both the contingent valuation method (e.g., Carson et al., 2003; Loureiro and Loomis, 2013) and choice experiments (e.g., Casey et al., 2008; Tuhkanen et al., 2016) have been applied to estimate the non-use value losses due to an oil spill.

Alternatively, this paper focuses on estimating recreational (use) value losses from an oil spill. In such a context, RP methods are more appropriate. The travel cost method (e.g., English et al., 2018) and the hedonic pricing method (e.g., Cano-Urbina et al., 2019) have been applied to estimate the use-value losses due to an oil spill. While the former tends to focus on impacts on local recreation, the latter has estimated the losses in market prices in the housing market (Winkler and Gordon, 2013), fish stocks (Domínguez Alvarez and Loureiro, 2013), or wages (Aldy, 2014).

Out of the methods presented, the Travel Cost Method is the most appropriate in our context due to its exclusive focus on recreation. To infer how to estimate the recreational impact of an oil spill on recreation, we identified twelve (12) prior studies that do so in different contexts.[1] These studies follow two different approaches to estimate the recreational loss due to an oil spill.

One approach is to calculate the value of a trip, usually in terms of consumer surplus, and multiply it by the number of lost trips due to the oil spill, thus yielding aggregate losses. Bonnieux and Rainelli (2003), English et al. (2018) and Stratus Consulting Inc. (2010) follow this approach. For example, English et al. (2018) estimate the value of a lost

[1] These twelve studies were identified after a through literature search of both peer and non-peer-reviewed literature. We used the keywords “travel cost method” and “oil spill” to identify possible studies, as well as snowballing as a literature search approach.

user day to be $37.23 and multiply this value by the estimate of lost user days (10 million), which equals $379 million in aggregate losses. This approach, however, assumes that recreationists who visit the same site or choose to recreate at a different recreational site do not incur any welfare loss (Glasgow and Train, 2018).[2]

An alternative approach is to calculate the welfare loss per trip or choice occasion due to an oil spill. The aggregate loss due to the oil spill is estimated by multiplying the loss per trip by the total number of trips or choice occasions (e.g., Alvarez et al., 2014; Hausman et al., 1995; Whitehead et al., 2018). Whitehead et al. (2018) estimate a change in consumer surplus due to the oil spill of $43 and multiply it by 4.82 million households, which yields $207 million in aggregate damages.

However, the two reviewed approaches require some form of historical trip data before and after the oil spill. In the first approach, the historical RP data is used to estimate the number of lost trips, whereas in the second approach, the historical data is used to estimate the welfare loss per trip or choice occasion.

Out of the twelve studies identified, Parsons (2008) is the only study to estimate the ex-ante impact of an oil spill. Estimating the loss of an oil spill before it has occurred may be more useful in the sense of motivating the creation or improvement of prevention measures. Parsons (2008) estimates the loss per trip due to a hypothetical oil spill in South Padre Island (US). In order to do so, Parsons (2008) assumed that an oil spill would decrease welfare exclusively due to closure of soiled beaches. Using RP data, the author estimates the compensating surplus by reducing the choice set faced by recreationists using the discrete choice model.

This site choice model is one way to operationalize the travel cost method. The model explains recreationists' choice of recreational site given observable site attributes, their travel cost, alternative specific constants (ASCs), and other relevant variables. Site choice models

[2] English et al. (2018) accounts for a demand-shift of recreation due to the oil spill by calibrating the alternative specific constants.

estimate welfare measures due to beach closure in a straightforward manner, and allow for site substitution (Parsons, 2017). Site choice models can consider a vast range of external pressures, such as changes in site attributes or quality (e.g., Hicks and Strand, 2000), changes in travel cost (e.g., Leggett et al., 2014), or closure of one or more sites (e.g., Parsons et al., 2009).

If we follow the approach of Parsons (2008) and estimate welfare loss due to a reduction in the available choice set, then it suffices to apply a discrete choice model to RP visitation data. We use the Random Utility Model (RUM). For individual i, let M denote income, ${C_{ij}}^0$ represents travel cost associated with getting to beach j and ${Q_{jt}}^0$ denotes quality of site j. The utility of going to site j at choice occasion t is given by:

$$U_{ijt} = V_{ijt}(M_i - {C_{ij}}^0, {Q_{jt}}^0) + \frac{1}{\sigma}\varepsilon_{ijt} , \qquad (1)$$

where σ is the scale parameter, which is usually normalized to one.

With the above framework in place, we can calculate the compensating surplus (CS) given various scenarios of beach closure. Let J represent the initial choice set (before the oil spill), and J_1 the final choice set. The final choice set size is expected to decrease with oil spill size. The marginal utility of money is denoted by β_1. To estimate the loss due to an oil spill, we calculate the CS per choice occasion as derived by Small and Rosen (1981):[3]

$$E(CV) = \frac{1}{\beta_1}\left\{\ln\left[\sum_{j\epsilon J_1} e^{V_{ijt}(M_i - {C_{ij}}^0, {Q_{jt}}^0)}\right] - \ln\left[\sum_{j\epsilon J} e^{V_{ijt}(M_i - {C_{ij}}^0, {Q_{jt}}^0)}\right]\right\}. \qquad (2)$$

Following Parsons (2008)'s approach, a scenario that does not imply beach closure yields a CS estimate of zero. This approach implicitly assumes that a recreationist would not change their behavior if the beach is not closed.

However, recreationists may have preferences towards oil spill avoidance, which translates into changes in their recreational behavior. Such changes might occur even if their preferred or last visited beach

[3] This estimate of CS assumes the indirect utility to be linear in parameters. See Hanemann (1984).

remains open and not soiled. For example, recreationists might prefer either to opt-out from engaging in beach recreation all-together, or to visit another beach and/or recreational site. In fact, there was evidence in the Deepwater Horizon oil spill that "individual perceptions of or uncertainty about conditions in the Gulf altered the (…) recreation behavior" of recreationists, "even in areas where the oil never actually made it to local beaches" (English et al., 2018). Glasgow and Train (2018) also propose the idea of welfare losses arising because recreationists "anticipated that the sites would or might be degraded." We interpret this as evidence of a reduction in the individuals' perceived quality of the beach sites, even if the beach is not actually soiled.

When considering the ex-ante impact of an oil spill, the researcher cannot capture this type of behavior with RP data alone. Instead, such behavior can be elicited using SP data. Whereas in the RP data there is no variation in the perceived site quality, we assume that in the SP scenario all beach sites near the hypothetical oil spill suffer from a drop in perceived site quality from ${Q_j}^0$ to ${Q_j}^1$.

Hence, welfare losses due to an oil spill are given by: 1) a reduction in the choice set, and 2) a reduction in perceived site quality. The resulting CS per choice occasion due to an oil spill is given by:

$$E(CV) = \frac{1}{\beta_1}\left\{\ln\left[\sum_{j\epsilon J_1} e^{V_{ijt}(M_i - {C_{ij}}^0, {Q_{jt}}^1)}\right] - \ln\left[\sum_{j\epsilon J} e^{V_{ijt}(M_i - {C_{ij}}^0, {Q_{jt}}^0)}\right]\right\}. \quad (3)$$

The inclusion of Q_j is only possible if SP data and RP data are combined. The combination of RP-SP data has several advantages compared with the use of a single data source. One of the advantages is attenuation of hypothetical bias stemming from the SP data source (Whitehead and Lew, 2019). SP data is particularly vulnerable to hypothetical bias due to the unfamiliarity associated with the hypothetical scenarios (Whitehead et al., 2008). Thus SP data may be calibrated to match actual market shares from RP data (e.g., Revelt and Train, 1998).

On the other hand, the hypothetical nature of SP data enables the creation of different policy scenarios to be considered, which is not possible with RP data alone (Whitehead et al., 2008). RP data frequently suffer from lack of variation or high multicollinearity, while SP data introduces greater variation in the levels of the attributes. Other advantages of combining RP-SP data include greater efficiency in estimating the parameters of interest, as well as capturing changes in both use and non-use values.

To the best of our knowledge, sixty-nine (69) studies combine RP-SP data to estimate recreational value changes. However, only nineteen (19) combine discrete RP-SP data, while the overwhelming majority (39) combines RP-SP count data. Given our focus on discrete data, we review two types of discrete SP data that have been combined with RP data: discrete choice experiments (e.g., Whitehead and Lew, 2019) or contingent behavior (e.g., Zimmer et al., 2012).

In discrete choice experiments (DCEs), individuals choose among two or more scenarios, which differ in terms of their attribute levels. Past studies combining RP and DCE data introduce changes in fish catches (Whitehead and Lew, 2019), moose population and hunting conditions (Adamowicz et al., 1997), tree species variation and hiking conditions (Abildtrup et al., 2015), or water quality (Cheng and Lupi, 2016).

While DCEs introduce more generic alternatives but allow for greater variation across attribute levels, CB alternatives usually include actual recreational sites or activities but introduce fewer changes in attribute levels. In CB questions, individuals are presented with a scenario featuring changes in site quality, travel cost and/or choice sets. Individuals are then asked to either anticipate how many trips they expect to make to each of the available sites (Jeon and Herriges, 2010; Truong et al., 2018; Yi and Herriges, 2017; Zimmer et al., 2012), or answer how their behavior would change relative to some past visit (Boxall et al., 2003; Loomis, 1997; Parsons and Stefanova, 2011). Given our interest to mimic actual choices, we opt for combining RP data with CB, wherein we elicit behavior relative to the last recreational visit.

2.1 Econometric Approach

In order to combine RP-SP data, one could assume the utility function as specified in Equation 1 to hold in both datasets. Respondents should exhibit the same preferences regarding the travel cost C_{ij} and environmental quality Q_j across datasets. Then one could stack the data and jointly estimate Equation 1 in a “naïve” way.

We call it “naïve” because this approach ignores potential scale differences across RP-SP datasets. Ideally, recreationists have the same underlying preferences when facing RP or SP scenarios. This should be reflected by the equality of parameters that define the indirect utility function, V_{ijt}, across the two datasets. However, as illustrated Swait and Louviere (1993), SP and RP data may appear to lead to different parameters of the utility function due to scale differences for RP-SP datasets. That is, the scale parameter σ in Equation 1 is dataset-specific, $\sigma^{RP} \neq \sigma^{SP}$. Scale differences across RP-SP datasets may be due to various factors: “random noise” (Hensher and Bradley, 1993), rank order or fatigue effects (Bradley and Daly, 1994), choice uncertainty (Lundhede et al., 2009) or different “effect of unobserved factors (…) between revealed and stated preferences” (Morikawa, 1994). To be able to compare and jointly estimate SP and RP data, differences in scale should first be accounted for (Swait and Louviere, 1993).

Another reason why stacking the data is “naïve” is that it does not account for scale heterogeneity. Scale heterogeneity entails the existence of a scale-adjusting term φ_i, which is individual specific (Hess and Rose, 2012). For example, lack of experience with recreation choices might increase uncertainty for respondents and result in larger variation of the scale-adjusting term (Hensher, 2012).

With these two considerations, we can expand Equation 1 for the case of RP-SP data. We assume that the underlying preferences are the same, that is, the indirect utility functions are the same in both datasets. In RP data, there is no variation in the environmental quality attribute.

The utility of alternative j for individual i at choice occasion t is represented by:

$$U_{ijt}^{RP} = V_{ijt}(M_i - C_{ij}, Q) + \frac{1}{\sigma^{RP}\varphi_i^{RP}}\varepsilon_{ijt}. \quad (4)$$

In our SP data, we only have one choice occasion per individual, but we introduce variation in the environmental quality attribute. Hence, we omit the subscript t and the utility function is expressed as:

$$U_{ij}^{SP} = V_{ij}(M_i - C_{ij}, Q_j) + \frac{1}{\sigma^{SP}\varphi_i^{SP}}\varepsilon_{ij}. \quad (5)$$

Let us illustrate the importance of accounting for both scale parameter and scale heterogeneity in our application. Since utility is latent, the researcher can rewrite Equations 4 and 5 above by multiplying all terms by $\varphi_i\sigma$ as such:

$$\varphi_i^{RP}\sigma^{RP}U_{ijt}^{RP} = \varphi_i^{RP}\sigma^{RP}V_{ijt}(M_i - C_{ij}, Q) + \varepsilon_{ijt}, \text{ and} \quad (6)$$

$$\varphi_i^{SP}\sigma^{SP}U_{ij}^{SP} = \varphi_i^{SP}\sigma^{SP}V_{ij}(M_i - C_{ij}, Q_j) + \varepsilon_{ij}. \quad (7)$$

Let the indirect utility function be linear in parameters as follows:

$$V_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 Q_j, \quad (8)$$

where β_0, β_1 and β_2 are the parameters of interest which are the same in both the RP-SP dataset. If there is in fact scale heterogeneity within the RP and SP datasets (i.e. $\varphi_i^{RP} \neq \varphi_i^{SP}$) and the scale parameter differs in RP and SP ($\sigma^{RP} \neq \sigma^{SP}$), then for β_0:

$$\hat{\beta}_0^{SP} \neq \hat{\beta}_0^{RP} \Leftrightarrow \widehat{\varphi_i^{SP}\sigma^{SP}}\beta_0 \neq \widehat{\varphi_i^{RP}\sigma^{RP}}\beta_0. \quad (9)$$

The estimated parameters $\hat{\beta}_0^{SP}$ and $\hat{\beta}_0^{RP}$ will appear to differ from one another. The researcher is led to erroneously conclude that the underlying preferences in SP and RP data differ, when in fact this difference is driven by distinct scale parameters and scale heterogeneity.

One interesting observation arises. Past combinations of RP-SP data have accounted for both scale differences as well as scale heterogeneity, wherein $\varphi_i^{RP} = \varphi_i^{SP}$. However, scale heterogeneity may also differ in SP and RP datasets, that is $\varphi_i^{RP} \neq \varphi_i^{SP}$. Combinations of RP-SP data seldom explicitly considered this difference, neither theoretically nor empirically (Hensher, 2012). While it may not be possible to disentangle scale heterogeneity from preference heterogeneity (Hess and Train, 2017), ignoring differences in scale heterogeneity may also lead to erroneous conclusions of differing parameters in SP and RP data. While we are aware of this gap in the literature, the nature of our data does not allow us to account for φ_i^{SP}, because our data consists of a single choice occasion per individual in the SP data. Hence, we assume $\varphi_i^{RP} = \varphi_i^{SP}$.

When jointly estimating RP-SP data, we can estimate the relative scale parameter so long as one of the scale parameters is fixed (Adamowicz et al., 1997). The scale parameter for RP data, σ^{RP}, is typically fixed at one (e.g. Adamowicz et al., 1997; Cheng and Lupi, 2016). The estimated scale parameter for SP data varies across studies: between 0.05 and 0.22 in Jeon and Herriges (2017), 0.13 in Adamowicz et al. (1997), between 0.12 and 0.45 in von Haefen and Phaneuf (2008), 0.59 in Truong et al. (2018), 0.62 in Cheng and Lupi (2016), between 0.61 and 0.70 in Haener et al. (2001), and 0.76 in Whitehead and Lew (2019). In some of these studies, the relative scale parameter is lower and statistically different from one, which implies greater variance in the SP data (Whitehead et al., 2008).

Several statistical models allow the estimation of the scale parameter when jointly estimating RP-SP data. One approach is to use the "nested logit trick" (Hensher et al., 2008; Hensher and Bradley, 1993). In such models, the RP-SP data are in different branches, and we retrieve the scale parameter through the dissimilarity parameter across branches. More recently other approaches have been applied, such as the conditional logit model with scale (Cheng and Lupi, 2016; Haener et al., 2001; Truong et al., 2018), the latent class model (Jeon and Herriges, 2017), the generalized mixed logit model (Cha and Melstrom, 2018), or the Error Component Mixed Logit model (Abildtrup et al., 2015). However, the choice of statistical model has modest to negligible effects on welfare estimates (Whitehead and Lew, 2019).

This is the first RP-SP application to account for the scale parameter in a standard mixed logit model.[4] The mixed logit model is a generalization of the conditional logit by allowing coefficients to be random variables (Revelt and Train, 1998). Mixed logit models also allow for unrestricted substitution patterns across alternatives and correlation in unobserved factors (Train, 2003). The motivation for choosing the mixed logit model for the combined data is that it allows for efficient estimation when data is comprised of repeated choices per individual rather than single choices (Revelt and Train, 1998). Our approach is to account for any differences in scale when estimating the parameters of interest. By fixing the RP scale parameter to be one but allowing the scale to differ in the SP data, the resulting attribute parameters are estimated relative to the RP data.

Mixed logit models also allow for scale heterogeneity. Hess and Train (2017) point out that if all utility coefficients are assumed to be randomly distributed, scale heterogeneity is captured and entangled in the standard errors of the estimated parameters from a mixed logit model (Hess and Train, 2017). Mariel and Meyerhoff (2018) suggest specifying correlated random parameters to capture scale heterogeneity.

The estimated mixed logit model in our paper accounts for preference heterogeneity, but uncorrelated random parameters. Following up on the utility function as defined in Equation 1, the indirect utility function to be estimated is as follows:

$$V_{ijt} = ASC_{ij} + \beta_{1i} * C_{ij} + \beta_{2i} * Q_{jt}. \tag{10}$$

We allow all coefficients to be randomly distributed. This means that individuals are allowed to have different preferences regarding the travel cost, the environmental quality, each of the available beaches, as well as the opt-out alternatives.

We assume that the travel cost (TC) coefficient follows a log-normal distribution. By assuming a lognormal distribution, we ensure that the distribution of the change in welfare estimate does not have infinite

[4] Hensher et al. (2008) proposed a mixed logit error components model to jointly model RP-SP data.

moments (Daly et al., 2012).[5] The coefficient associated with travel cost is expressed as follows:

$$\beta_{1i} = -\exp\,(\bar{\beta}_1 + \sigma_1.\,\varepsilon_{1i}), \qquad (11)$$

where $\bar{\beta}_1$ is the mean of the log-normal distribution, σ_1 is the standard deviation of β_1, and ε_{1i} follows a normal distribution across individuals. The parameters to be estimated are $\bar{\beta}_1$ and σ_1.

We assume that the parameters associated with alternative-specific environmental quality (Q) and the ASCs follow normal distributions. The random coefficient associated with environmental quality (Q) is expressed as follows:

$$\beta_{2i} = \bar{\beta}_2 + \sigma_2.\,\varepsilon_{2i}, \qquad (12)$$

where ε_{2i} is individual-specific and follows a normal distribution. The parameters to be estimated are $\bar{\beta}_2$ and σ_2.

The ASCs are random coefficients, as follows:

$$ASC_{ij} = \overline{ASC}_j + \sigma_{ASCj}.\,\varepsilon_{ASCji}, \qquad (13)$$

where the parameters to be estimated are $\overline{ASC}_j$ and σ_{ASCj}.

[5] The coefficient associated with travel cost is in the denominator of the site closure loss estimate. If the distribution of the travel cost variable allows values marginally close to zero, the resulting estimate in welfare loss converges to infinite. For an explanation, see Daly et al. (2012).

3 Data

The study site concerns the *Jæren* beaches on the south-western coast of Norway (illustrated in Figure 1). Annual visitation to the *Jæren* beaches is estimated to be at least 600.000 visits (Sveen, 2018). In the study area oil spills are a relevant threat due to heavy marine traffic along the coast. Oil tankers navigate as close as three kilometers from the coast.[1] Since 2011, the Norwegian Maritime Authority has recorded a total of 132 cargo ship accidents in the jurisdiction of the Rogaland county. The most recent ship accident occurred in February 2017 when the ship "Tide Carrier" was grounded just two kilometers away from the *Jæren* beaches but no oil was spilled. Nonetheless, the environmental damage would have been considerable if the oil aboard the ship had spilled (around 600 m^3 of heavy oil and around 300 m^3 of diesel oil).[2]

To collect data on general visitation patterns and elicit people's behavior in case of a hypothetical oil spill, we conducted an original survey during October-November 2018. A national survey company (*Norstat*) sampled residents in the Rogaland county of Norway from their national web-panel. The response rate was 25.9%. A detailed overview of survey design process, survey implementation, and data description is available in Lopes and Mariel (2020).

The survey is organized into four sections. In the first two sections of the survey, we collect RP data, wherein individuals reported all their beach visits during the summer season (previous four months) across all twenty-six beaches and identified their last visited beach. In the third section we elicit the individuals' preferences towards oil spill aversion using a CB question. In the final section, we ask individuals to report their household and individual characteristics.

[1] Norwegian marine traffic data for 2015 is available at https://kystinfo.no/.

[2] This ship accident was mentioned in the survey to ensure consequentiality of the survey instrument, and 41% of the respondents indicated they had heard of this accident beforehand. More information about this ship grounding is available in Norwegian at https://www.kystverket.no/.

3.1 Revealed Preference Data

The visitation (RP) data comes from the number of reported visits to each beach during the summer season. The summer season consists of 123 days, comprising the months of May, June, July, and August. In total, the 647 individuals in our sample visited *Jæren* beaches 5985 times, which constitutes the number of observations in the RP data set. Similarly to previous studies (e.g., Haener et al., 2001), the number of choice occasions is equal to the number of beach visits during the summer season (i.e., individual-specific). The number of beach visits during the summer season ranges from 1 visit (96 individuals) to 123 visits (3 individuals). The mean number of beach visits is 9.25 (median is 5 visits).

The choice set is comprised of 26 beach alternatives. When modeling the RP outcomes, we assume individuals face the full choice set.

We expect that the probability of visiting a given beach given the choice set is negatively related to the travel cost. We construct the travel cost C_{ij} for individual i associated with beach j as follows:

$$C_{ij} = 2 * \left(0.79 * d_{ij} + w_i t_{ij}\right). \tag{14}$$

We measure the one-way driving distances (d_{ij}) and times (t_{ij}) from the reported postal codes of each individual to each of the 26 beaches using the Google Maps API tool. Cost per kilometer of driving a diesel car is assumed to be 0.79 NOK per kilometer. We assume the opportunity cost of time to be 33% of the hourly wage rate (w_i). We input annual mean net income (376,897 NOK) for individuals with missing income data. Travel cost is calculated per person and per visit. Mean travel cost for the sample is 175 NOK (median=144).[3]

[3] As of 14/02/2019: 1 Euro = NOK 9.7652; 1 USD = NOK 8.6442 (Source: Bloomberg)

3.2 Stated Preference Data

Individuals are randomly assigned to one oil spill size: either a small, medium, large or very large oil spill. The hypothetical oil spill would occur due to a ship grounding south of the *Jæren* beaches. The company DNV-GL simulated four oil spill dispersion scenarios given the quantity of oil spilled in each scenario, local ocean currents and the origin of the oil spill, which was chosen given current marine traffic data in the study area. Figure 1 illustrates the four oil spill sizes as well as the oil dispersion in each case.

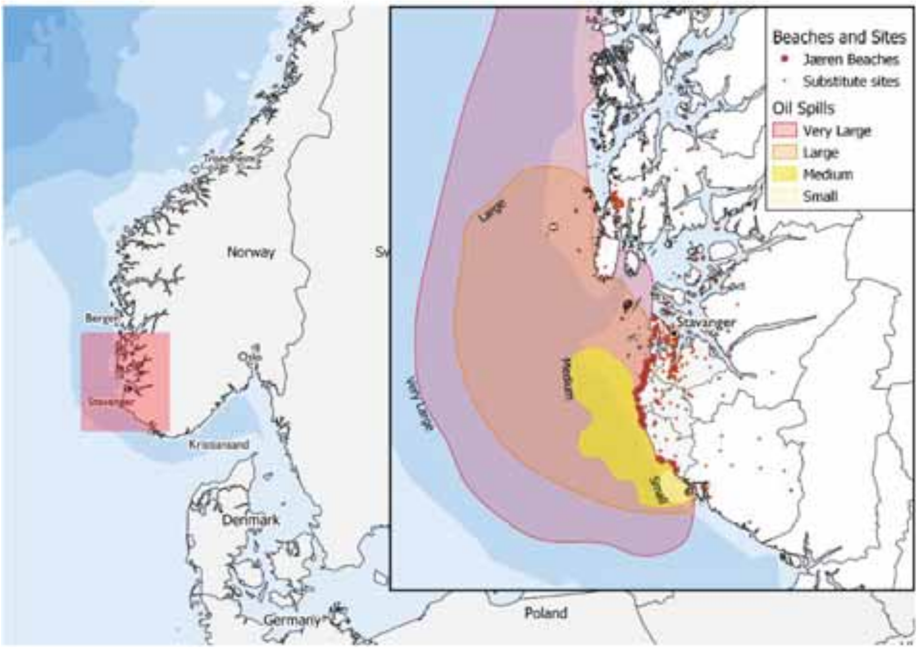


Figure 1 – Study site (left) and Oil Spill Illustration (right) for the four sizes considered

The design of the CB question builds upon previous CB surveys, such as Parsons and Stefanova (2011), Boxall et al. (2003) and Loomis (1997). We designed this unique CB question to mimic a real-life recreational choice, by giving individuals not only the options of visiting another beach in the study area or stay at home, but also the options to engage in other recreational activities, i.e. a beach outside the study area, or a different recreational site. Previous studies included limited opt-out options: either not to engage in the recreational activity (e.g., Adamowicz et al., 1997) or to stay at home (e.g., Adamowicz et al., 1994).

An example of the CB question for a large oil spill is in Figure 2. The elicited behavior is relative to the respondent's last beach visit, hence visiting the same site is not always an available option if the previously visited beach is now closed.

Suppose that a large oil spill would occur. (…)

You have indicated earlier in the survey that your last visited beach along the *Jæren* coast in the summer of 2018 was *[BEACH ID]*.

This oil spill would imply closing almost all the *Jæren* beaches (…). A coastline of <u>50 kilometers</u> would be affected, and it would take around <u>three years for it to recover</u> to the same state as before the oil spill.

Closure of the beaches is expected to take <u>several weeks</u>. You can still visit *[BEACH ID]* as previously.

Think about the last trip you took. What would you have done if the described oil spill had happened?

- ☐ Visit *[BEACH ID]* as previously
- ☐ Go to another *Jæren* beach
- ☐ Go to another beach outside *Jæren*
- ☐ Visit another recreational site (not a beach)
- ☐ Stay at home / Do something else

Figure 2 – CB Question for large oil spill treatment

Each oil spill was described in terms of the number of beaches available, the kilometers of coastline soiled, time of beaches closure, and time required for the ecosystem to recover from the oil spill. Recreational impacts increase with oil spill size, as described in Table 1.

Table 1 – Oil Spill Attributes for each oil spill scenario (CB question)

Oil Spill Size	Number of beaches available	Kilometers of coastline soiled	Time of beaches closure	Time required to recover from oil spill
Small	All (26)	5 km	0	6 months
Medium	16	20 km	1 to 2 weeks	1 year
Large	5	50 km	Some weeks	3 years
Very Large	2	250 km	Several weeks	5 years

The CB question was posed if the individual had had one *Jæren* beach visit during the summer season of 2018. This comprises the single choice occasion per individual in SP data.

In the case of a small oil spill, the choice set includes 29 alternatives: the initial 26 beach sites and 3 opt-outs (i.e., Go to another beach outside *Jæren*, Visit another recreational site, or Stay at home / Do something else). For the remaining scenarios, we assume individuals face a reduced choice set for beach alternatives. The choice set size decreases for the case of a medium (19 alternatives = 16 beaches + 3 opt-outs), large (8 = 5 beaches + 3 opt-outs), and very large oil spills (5 = 2 beaches + 3 opt-outs).

Table 2 summarizes the number of responses in the RP and CB data. Following up on the discussion in Section 2, if the loss of recreational values is exclusively due to beach closure, in the case of a Small oil spill, we would expect all respondents to visit the same beach, which is not the case. As expected, the number of individuals who choose to visit the *Jæren* beaches (i.e., Visit the same beach, or Go to another *Jæren* beach) decreases with increases in the oil spill size. In parallel, the number of people choosing to stay at home increases with the oil spill size.

Table 2 – Number of responses to CB question for each oil spill scenario

	Oil Spill Size	Visit the same beach	Go to another *Jæren* beach	Go to another beach outside *Jæren*	Visit another recreational site	Stay at home / Do something else
RP Data	*None*	5985		0	0	0
SP Data	*Small*	93	32	6	24	12
	Medium	79	24	10	36	22
	Large	9	23	17	76	26
	Very Large	2	12	16	98	40
	Total	183	91	49	234	100

Almost half of the respondents (43%) chose one of our two novel opt-outs (i.e., Go to another beach outside *Jæren*, or Visit another recreational site). For these respondents, we then asked them in an open-ended question to indicate a specific recreational site where they would go. Respondents identified 58 different recreational sites which include a wide variety of recreational sites, such as mountains, forests and lakes (which jointly serve as alternative 29), or other local beaches (which serve as alternative 28). These alternative recreational sites are illustrated in the "Substitute Sites" dots in Figure 1, and comprise our multinomial SP data.

We calculate individual and site-specific travel costs using Equation 14. The respondents who indicated that they would still visit a *Jæren* beach were not asked to state other recreational sites they would visit. We assume that the substitute sites that these respondents face (another beach and another recreational site) are the sites with the lowest travel cost.[4] The average travel costs to visit other beaches or recreational sites

[4] For every respondent with missing alternative recreational sites, we calculated their travel costs for the five most popular beaches (*Godalen, Vaulen, Skadbergsanden,*

are slightly lower than the travel cost to visit the *Jæren* beaches: the average TC to visit another beach outside *Jæren* (alternative 28) is 92 NOK, and to visit another recreational site (alternative 29) is 114 NOK.

The high proportion of respondents (43%) choosing to visit other recreational sites suggests that local lakes, mountains and hikes are relevant recreational substitutes to beach recreation and their omission is problematic as welfare estimates will be biased.

When contemplating how to measure environmental quality, a more direct approach would be to include oil spill size dummies for a small, medium, large and very large oil spill. However, it is not possible to identify the oil spill dummies due to them being confounded by the reduction in choice set. Instead, to account for people's averting behavior towards the oil spill, we introduce a variable (Q_j) that represents the proximity of each beach to the oil spill. We calculated the Euclidian distance from the oil spill to each of the beaches or recreational sites given the oil spill scenario.[5] In the case of inland recreational sites or when selecting the stay-at-home option, proximity was set equal to the maximum distance calculated to any coastal recreational site (125 kilometers). This strategy was motivated to obtain conservative welfare estimates of recreational losses. The average proximity from each beach site to the oil spill is 33 kilometers (median=22).

Table 3 reports descriptive statistics of the proximity variable for RP data and each oil spill scenario (SP data).

Mollebukta, Sandvesanden) and the five most popular recreational sites (*Dalsnuten, Stokkavatnet, Melsheia, Sørmarka, Preikestolen*). We then identify for each respondent the beach and the recreational site with the lowest travel cost.

[5] We also calculated the distance along the coast as an alternative measure of proximity to oil spill. However, the resulting model has the same fit (LL=-13347.08).

Table 3 - Descriptive statistics of Proximity Variable (in kilometers)

	Oil Spill Size	**Mean**	**Median**	**Min**	**Max**
RP data	*None*	125.11	125.11	125.11	125.11
SP data	*Small*	37.83	34.21	0.69	125.11
	Medium	13.55	4.30	0	125.11
	Large	9.11	0	0	125.11
	Very Large	8.89	0	0	125.11

The inclusion of the proximity variable is only possible due to the availability of SP data. Since no historical data on oil spill averting behavior is available in our study area, we cannot infer on people's preferences from the RP data alone.

4 Results

Our sample is comprised of 6632 choice occasions among 647 individuals. The former includes 5985 beach visits (90.24% of the sample) during the summer season (RP data) and 647 responses to the CB question (SP data). We apply a repeated site choice logit model, wherein the number of choice occasions is individual-specific (i.e., the number of beach visits in the study area, plus one answer to the CB question).

We use the full set of ASCs, travel cost and proximity to explain site choice as specified in Equation 10. The coefficients associated with the ASCs capture the average utility of observable and unobservable characteristics of each site. We chose to omit observed site attributes to explain site choice (e.g., parking) since the ASCs implicitly capture the value of site attributes. Moreover, our focus is to calculate site closure welfare estimates, for which we only need ASCs. Including the full set of ASCs in detriment of observed site attribute also avoids endogeneity concerns due to unobserved site attributes (Murdock, 2006), such as congestion. The 26 *Jæren* beaches are coded from 1 to 26 (ASC1 is fixed at zero). The utility of staying at home relative to ASC1 (Bore beach) is captured by ASC27, the utility of going to another beach is captured by ASC28 and the utility of going to another recreational site is captured by ASC29.

The novelty of our paper is through the inclusion of the Proximity variable and ASC28 and ASC29, which is possible through the inclusion of SP data. Moreover, the stay at home option (ASC27) is also seldom included in discrete data choice models applied to recreation. Out of the seven studies reviewed that combine discrete RP and CB data, only Yi and Herriges (2017), Loomis (1997) and Parsons and Stefanova (2011) added the stay at home alternative in their discrete choice model.[1]

[1] These three papers are discrete data applications. Count data applications of RP-SP include by default the option of not to recreate (i.e., count equal to zero).

Preferences for proximity are expected to exhibit diminishing marginal utility. That is, a marginal increase in proximity (i.e. by one kilometer) is expected to generate greater utility when the individual is considering a beach near the oil spill, in contrast to a beach farther away. In other words, we expect the proximity variable to enter the indirect utility function in a non-linear fashion. The proximity variable in Table 4 is the squared root of proximity that allows for diminishing marginal utility.[2]

We apply the conditional logit model when using the SP data and the mixed logit model when using the RP or the combined RP-SP data. The results for the SP, RP and RP-SP data are summarized in Table 4 (full output with ASC and standard deviations is reported in Appendix 1). The estimated parameters are in utility space.

Table 4 – Selection of Regression Results

	SP Data		**RP Data**		**RP & SP Data**	
	Conditional Logit Model		Mixed Logit Model		Mixed Logit Model	
	Estimate	**Robust s. e.**	**Estimate**	**Robust s. e.**	**Estimate**	**Robust s. e.**
TC	**-0.002**	0.001	**-3.653**	0.057	**-3.901**	0.057
SQRT(PROX)	**0.250**	0.049			**0.428**	0.143
ASC27	-0.737	0.448			**-9.739**	3.835
ASC28	-0.021	0.192			**-3.520**	1.696
ASC29	-0.461	0.440			**-5.480**	3.149
Standard deviations						
Sigma TC			**0.8364**	0.0442	**0.847**	0.095
Sigma SQRT(PROX)					0.163	0.246

[2] We also estimated specifications including proximity as linear, logarithmic, quadratic and as its inverse. The resulting Log-Likelihood values in the RP-SP mixed logit model are -13358, -13347, -13365 and -13362, respectively.

Sigma ASC27					**7.131**	3.198
Sigma ASC28					1.566	2.604
Sigma ASC29					-5.159	6.276
Scale parameters						
RP (fixed)					1.000	
SP					**0.983**	0.229
Parameters		5		53		62
Individuals		647		647		647
Observations		647		5985		6632
Log-likelihood value		-1414.38		-12178.12		-13348.71

Notes: Coefficients in bold represent statistical significance at the 5% level. We used 500 Halton draws when running the mixed logit models. Models were estimated using the Apollo package in R (Hess and Palma, 2019). We estimate a full set of ASC and standard deviations for the distributions of ASC1 through 26 in both the RP and RP-SP models.

The coefficient associated with the travel cost variable is negative and statistically significant across all models, thus as expected respondents exhibit negative price sensitivity to beach visits. Individuals exhibit preference heterogeneity regarding the travel cost variable, since the standard deviation (σ_1) is statistically significant. As specified in Equation 12, the coefficient associated with the travel cost variable when using RP or RP-SP data is assumed to be negative log-normally distributed in the mixed logit models, hence the travel cost variable may be transformed as $-\exp(-3.901) = -0.02$. The resulting parameter is approximately ten times higher than the coefficient obtained when using the SP data ($\beta_1 = -0.002$).

As discussed in Section 2, if the scale differs across the two datasets, the estimated coefficients from the SP data are not directly comparable with the coefficients from the RP or RP-SP data. We find that the estimated scale parameter (0.983) is not statistically different from one (t-statistic of $(0.983 - 1)/0.229 = -0.07$). Likewise, other studies (e.g., Adamowicz et al., 1997; Parsons and Stefanova, 2011) estimate a scale parameter not statistically different from one.

The coefficients associated with staying at home (ASC27), visiting another beach outside *Jæren* (ASC28) or going to another recreational site (ASC29) are negative and statistically significant ($\overline{ASC_j} < 0, j = \{27,28,29\}$). This implies that staying at home or visiting a different recreational site bring disutility to recreationists relative to visiting Bore beach. Staying at home brings the highest disutility ($\overline{ASC}_{27} = -9.74$), but respondents exhibit preference heterogeneity regarding staying at home ($\sigma_{ASC27} = 7.13$). Respondents do not exhibit preference heterogeneity regarding the options of going to another beach or another recreational site (i.e., σ_{ASC28} and σ_{ASC29} are statistically insignificant).

Accounting for proximity of recreational sites to the oil spill (in kilometers) improves the statistical fit of the model (AIC of 26821, compared with AIC of 26860 when omitting distance). The coefficient associated with proximity is positive and statistically significant ($\beta_2 = 0.428$). However, the standard deviation of the proximity distribution is statistically insignificant, hence respondents do not exhibit preference heterogeneity regarding the proximity variable.

Because the chosen specification includes the root square of proximity as the variable of interest, the marginal WTP per kilometer is given by $-\frac{\beta_{2i}}{2\sqrt{PROX}} \Big/ -e^{\beta_{1i}}$. The mean marginal WTP is 28 NOK per person to get 1 kilometer farther away from the oil spill (median=18.5). Figure 3 illustrates marginal WTP per visit given proximity.

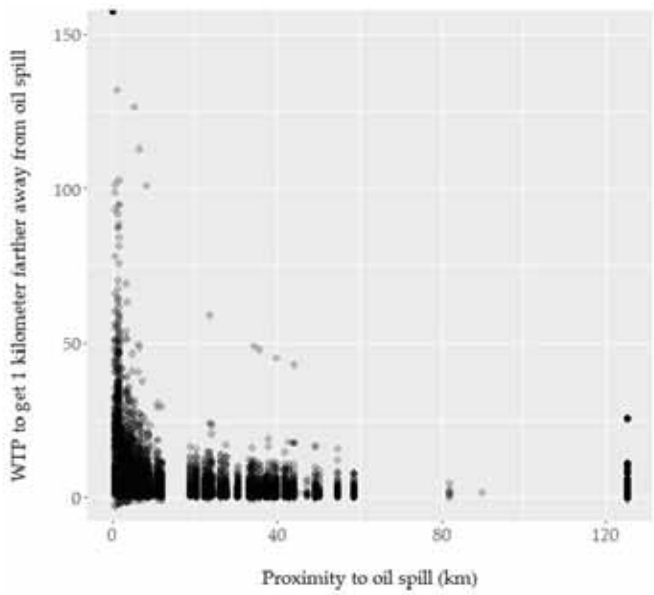


Figure 3 – Marginal WTP for proximity attribute given proximity level (in kilometers). **Note:** The estimated marginal WTP is based on 1000 draws from the distributions of travel cost and proximity variables.

We find that WTP to be farther away from the oil spill may be zero or even negative. When proximity is close to zero, marginal WTP per kilometer ranges between -4.4, and 316.6 NOK. When considering recreational sites farther away from the spill (80 to 120 km away), marginal WTP ranges from 0 to 4 NOK. The negative WTP may be explained by the phenomenon of solidarity visits. Loureiro et al. (2006) reported a "solidarity effect" in the case of the Prestige oil spill, as some visitors specifically choose to visit the site soiled by the oil spill given their interest and desire to experience the Prestige oil spill by themselves.

In Table 5, we report the CS estimates per choice occasion and per person when using SP data, RP data and RP-SP data. When using RP-SP estimates, we obtain a CS of -122.94 NOK for small, -188.30 NOK for medium, -261.62 NOK for large, and -288.64 NOK to avoid a very large oil spill.

Table 5 – Mean Compensating surplus for oil spill of varying sizes

CS Estimates	**SP Data**	**RP Data**	**RP-SP Data**
Small	-772.27 [-961.00 ; -583.54]	0.00	-122.94 [-271.27 ; +17.15]
Medium	-1267.82 [-1551.28 ; -984.36]	-3.66 [-15.21 ; +7.89]	-188.30 [-361.24 ; -35.27]
Large	-1621.11 [-2071.31 ; -1170.92]	-83.14 [-99.81 ; -66.47]	-261.62 [-486.94 ; -61.19]
Very Large	-1706.82 [-2224.19 ; -1189.45]	-157.21 [-178.49 ; -135.93]	-288.64 [-538.98 ; -57.06]

Notes: Means of CS are shown. Values are in Norwegian kroner (NOK) per person (1 NOK = 0.11 US dollars). CS estimates for RP and RP-SP data are based on 10000 draws from the parameter distributions. The 95% confidence intervals are reported in brackets.

Our welfare estimates are similar to previous consumer surplus estimates for this study area. The CS from a small oil spill of 122 NOK per person per visit is in the ball-park of the consumer surplus per visit to either *Sola* or *Orre* beach of 135 NOK per person (Kipperberg et al., 2019). Bui and Sæland (2017) found a consumer surplus of 91 per visit to *Orre* beach and 150 NOK per visit to *Sola* beach. Kleppe and Jensen (2018) estimated a consumer surplus per person per trip of 124 NOK for both *Bore* and *Hellestø* beaches.

Throughout Section 2, we argue that the combination of RP-SP data allows for the estimation of the attribute relative to perceived quality of beaches. We hypothesize that excluding this proxy for site quality decreases welfare losses. As seen in Table 5, when using RP data and thus excluding perceived site quality, welfare losses are lower than those using RP-SP data. These losses range from 0 to 225 NOK. When including proximity these range from 122 to 288 NOK per person and per visit. The magnitude of the difference is greater for smaller oil spill sizes (small and medium).

In Section 2, we also argue that grounding SP outcomes on actual choices is one of the advantages of combining RP-SP data. Indeed, estimated losses using the SP data alone are around 6 times higher (between 728 and 1684 NOK) than estimated losses using the combined RP-SP data (between 123 and 289 NOK). Adamowicz et al. (1994) also finds that estimated welfare measures are higher by several orders of magnitude using SP data compared with RP or RP-SP data.

One would expect there would be statistical scope in the welfare estimates, that is, the welfare loss due to a small oil spill would be smaller and statistically different from the loss due to a medium, large or very large oil spill. The welfare losses due to an oil spill accident are not statistically different, but they are increasing with the oil spill size. Note that the welfare estimates across all oil spill scenarios are expressed in per choice occasion terms. The welfare estimates per choice occasion do not necessarily need to exhibit scope for validity.

5 Discussion

While the scale parameter was found not to be statistically different from one, the only variable common in both RP and SP data (i.e., travel cost variable) was substantially different across datasets. The welfare estimates in SP, RP and RP-SP models also differ to a large extent. This different is not explained by differences in scale, since the estimated scale parameters in RP and SP data do not differ statistically. This disparity may lead the researcher to suspect of underlying preferences differing across datasets. Instead, we argue that model misspecification when modeling RP and SP drives this disparity.

When we estimate SP data with a single choice occasion, we are unable to estimate a full set of ASCs for beach sites, the standard deviations of all random coefficients or scale heterogeneity. The omission of these variables when modelling SP data is problematic and biases welfare estimates upwards. When we estimate RP data, we are unable to estimate the proximity variable as well as ASCs for staying at home (ASC27) or going to another recreational site (ASC28 and ASC29). The omission of the proximity variable when modelling RP data is problematic and biases welfare estimates downwards. When we model RP-SP data jointly, all parameters of interest can be estimated, as well as a scale parameter.[1] In line with von Haefen and Phaneuf (2008), "jointly estimating preferences with RP and SP data (…) permits identification of all structural parameters" rather than separately estimate RP or SP data. Jointly estimating RP-SP data also allows for preference heterogeneity and scale heterogeneity.

Given our concern with ensuring a correctly specified model, we investigate whether our combined RP-SP model is misspecified. That does not seem to be the case. To jointly model RP-SP choices, we

[1] Because of the additional parameters that we estimate in the RP-SP model (e.g., standard deviation of ASC27, ASC28, ASC29 and Proximity variable), this model outperforms the separate estimation of RP and SP models. Hence, we do not conduct a Likelihood Ratio Test as proposed by Swait and Louviere (1993).

estimated a mixed logit model. McFadden and Train (2000) argued that "any discrete choice model derived from random utility maximization has choice probabilities that can be approximated as closely as one pleases by a [mixed multinomial logit] model." However, the random coefficients of the mixed logit model need to follow flexible distributions. Given the concern for misspecification of our mixed logit model, we perform several sensitivity analyzes to our baseline model. These relate to preference heterogeneity and scale heterogeneity assumptions. The results from various specifications are reported in Appendix 2.

First, we limit the number of coefficients that are random by: 1) not allowing for preference heterogeneity; and 2) allowing for preference heterogeneity only for the proximity and travel cost variables. We argued above that the omission of standard deviations when specifying the SP data led to misspecification. Our RP-SP models that omit standard deviations corroborates our argument. We run a conditional logit model (Appendix 2, column 1), and a mixed logit model with ASCs as fixed parameters (Appendix 2, column 2). Both models result in worse fit (AIC of 31915 and 29832) compared with the baseline RP-SP model that allows all coefficients to be random parameters (AIC of 26821). We also conclude that assumptions regarding random coefficient distributions have important impacts in the estimated scale parameter. The estimated scale parameter is 0.9827 in the baseline model, which is statistically different from zero, but not statistically different than one. The scale parameter, however, differs from unity in the models with restrictive patterns of preference heterogeneity. Welfare measures computed with the estimates from the conditional logit model are in the same order of magnitude as the baseline mixed logit model, but specifying only travel cost and proximity as random parameters results in implausibly high welfare losses. In conclusion, allowing for preference heterogeneity is important to avoid misspecification.

Given the sensitivity of the mixed logit model to distributions of the random parameters, we use different distributions for the travel cost variable. We estimate models with the travel cost parameter as fixed, normal or triangularly distributed. When restricting the travel cost coefficient to be fixed ($\beta_1 = -0.0144$), the fit deteriorates considerably (AIC of 27292). Moreover, the scale parameter differs from unity. The results are reported in Appendix 2, column 3. When the travel cost parameter follows a normal or triangular distribution, the fit is similar to the baseline model (AIC between 26802 and 26882). The results are reported in Appendix 2, columns 4 and 5. However, the resulting distributions of the travel cost variable have masses at zero. This implies infinite moments for the distribution of WTP (Daly et al., 2012).

In Section 2.1, we argue about the importance of allowing for scale heterogeneity to ensure we can perform correct inference about (in)equality of parameters. Allowing for scale heterogeneity is an advantage of pooling the RP-SP datasets. For example, respondents more familiar with oil spills or more knowledgeable about the *Jæren* beaches in study area are more likely to be consistent in their choices. That would translate into the scale parameter being smaller for familiar or knowledgeable respondents rather than unfamiliar or unknowledgeable respondents. To allow for other patterns of scale heterogeneity, Fiebig et al. (2010) propose the generalized multinomial logit (GMNL) model. Hensher (2012) illustrate how the GMNL model can accommodate for differences scale heterogeneity differences across SP and RP datasets. We combine the RP-SP data using the generalized MNL model.[2] The scale parameter is a distribution which depends on τ and on the coefficient associated with the SP dummy (η), as follows: $\sigma_i = \exp\left[-\frac{(\tau+\eta SP)^2}{2} + (\tau + \eta SP)w_i\right]$, where w_i is standard normally distributed. Given the estimates reported in Appendix 2, column 6, the mean scale for the RP data is 0.9798 ($sd = 0.3637$) and the mean scale for the SP data is 0.7573 ($sd = 1.0019$). These estimates are in line with the models estimated with mixed logit model. The fit of the generalized

[2] Given the limitations of the NLOGIT software, we estimated this model with limited preference heterogeneity and the travel cost variable normally distributed.

MNL model is superior to the baseline mixed logit model (AIC of $23979 < 26821$). The resulting welfare estimates are virtually the same as the baseline mixed logit model.

A second approach to capture scale heterogeneity is to specify random parameters as correlated. Mariel and Meyerhoff (2018) recommend doing so to prevent scale heterogeneity from being picked by the parameters of interest. We estimate a mixed logit with correlated random parameters. Resulting estimates are reported in Appendix 2, column 9. The model outperforms the restricted version of the mixed logit model if we consider the AIC (AIC of 25826 < 26821), but the BIC of the model with correlated parameters is higher than uncorrelated (BIC of 29410 > 27243) due to the additional 527 parameters to be estimated in the correlated parameters model. Estimated welfare losses are slightly higher but not statistically different from those reported using the baseline model. Given the improvements in fit of these two models that allow for more flexible patterns of scale heterogeneity (GMNL and mixed logit with random parameters), we conclude that accounting for scale heterogeneity is an important advantage of pooling RP and SP datasets.

If underlying preferences differ across datasets, then the two datasets should not be pooled. To account for this possibility, we follow two different approaches: 1) we calibrate the scale to be the ratio of the travel cost parameters from the RP-SP models when computing welfare estimates as suggested by von Haefen and Phaneuf (2008); and 2) we add interaction of the parameters with the SP dummy to allow parameters to differ in RP and SP data

von Haefen and Phaneuf (2008) suggest using RP parameters and filling in missing parameters from SP data to estimate welfare measures. The authors transfer SP parameters to RP-space by fixing the scale parameter as the ratio of the travel cost parameters in SP and RP data. Accordingly, the scale parameter is $\bar{\sigma}_{SP} = -\frac{-0.002}{exp(-3.65)} = 0.07$. This fixed scale is fourteen times lower than the freely estimated scale parameter

($\hat{\sigma}_{SP} = 0.98$). The estimated welfare losses, however, are substantially higher than those from the baseline model. These are reported in Appendix 3. Alternatively, von Haefen and Phaneuf (2008) suggest filling-in missing parameters in SP data with scale-adjusted RP parameters. Again, the estimated welfare losses are higher than those estimated with the parameters from the jointly estimated data. These are reported in Appendix 3. The fact that estimated welfare losses differ greatly from those resulting from jointly estimating RP-SP data further corroborates our conclusion that estimating RP or SP data separately leads to misspecification. In conclusion, the proposed strategy by von Haefen and Phaneuf (2008) does not provided unbiased welfare estimates because the separately estimated RP and SP models were misspecified in the first place.

Second, we allow the parameters to enter the log likelihood function differently in the RP-SP data sets by adding the interaction of all parameters and the dummy for the SP scenario. We also consider an alternative specification wherein we add the interaction between the travel cost parameter and the SP dummy to allow only for the travel cost to differ in RP and SP data. The estimated models are in Appendix 2, columns 7 and 8. If we consider the AIC as the measure of fit, the model with the best fit is the one with all interactions with the SP dummy. This conclusion differs when we consider the BIC as the measure of fit, wherein the model with the SP dummy interacted with travel cost results in the best fit. If we consider the BIC, the baseline mixed logit model without interactions outperforms the model with all the interactions. The coefficient of the interaction term with travel cost parameter is positive and statistically significant ($\hat{\beta}_{TC*SP} = 0.0115$) in both specifications. Nonetheless, the impact on welfare estimates of these alternative specifications is modest. The estimated welfare losses are slightly lower but in the same order of magnitude as those from the baseline model (see Appendix 3). Hence, we conclude that the travel cost variable does differ across the SP and RP datasets besides what would be already allowed by the scale parameter, but the impact on welfare analysis is modest.

The second contribution of our work is to provide an alternative design of a contingent behavior question that expands the choice set in the SP scenario (see Section 3.2 for details). Our motivation was to mimic real-life recreational choices by including not only the option to stay at home, but also to recreate at other sites, such as forests, mountains or lakes. We find that omitting extra alternatives in our RP-SP model does not allow the mixed logit model to converge. In our context, omitting these alternatives results in misspecification of the RP-SP model. We also hypothesize that excluding the possibilities to recreate at another site or stay at home increases welfare losses, since it forces visitors to choose another beach. Welfare losses are slightly lower when we omit the alternatives of engaging in another recreational activity or beach (ASC28 and ASC29). These welfare estimates are reported in Appendix 3.

6 Conclusions

In this paper we combine RP and SP data to estimate the recreational impact of four hypothetical oil spills on the Jæren beaches in Norway. We make two contributions to the state-of-the-art of RP-SP data combinations. First, we argue that joint estimation of RP-SP models rather than separate estimation of either dataset avoids misspecification and results in unbiased welfare estimates. Second, we propose the design of a contingent behavior question that expands the recreationist's choice set and more faithfully mimics real life recreational choices. This paper is the first to jointly estimate RP site selection data and multinomial CB data.

We present separately estimated SP and RP data models and jointly estimated RP-SP models. We find that the scale parameter is not different between data sources. We conclude that the data is compatible for joint estimation, so long as allowances are made for the travel cost variable to enter the utility functions in SP and RP data differently. We find hypothetical bias in the SP data with a travel cost coefficient that is ten orders of magnitude different than the travel cost coefficient from the RP data. This leads to welfare estimates in SP data to be six times higher than the those estimated using RP data. Jointly estimating the RP-SP data calibrates the hypothetical bias. The additional flexibility generated by the SP scenario allows for an unbiased estimate of the welfare change from the hypothetical scenarios. We find that the welfare loss from an oil spill is 123 to 289 NOK per person per beach visit over the small to very large oil spill range. When considering 600 000 annual visits, the recreational value lost due to an oil spill ranges from 368 million NOK (Small) to 718 million NOK (Very Large).

Our application is the first to consider two simultaneous changes due to an oil spill: we hypothesize that an oil spill not only reduces the available choice set available to recreationists but also reduces the perceived beach site quality. The inclusion of the site quality attribute through the variation present in the SP data leads to calculation of welfare loss estimates (RP-SP data) which are higher than the estimates when excluding the proximity variable (RP data). Hence, we argue that estimating RP data without variation in site quality will lead to misspecification. The same is true for SP data, wherein ASCs or preference heterogeneity cannot be included to explain site choice.

Given the concerns of misspecification, we further show that our baseline RP-SP model and resulting welfare estimates are robust to preference heterogeneity assumptions. We also test whether underlying preferences differ across SP and RP datasets. While there is evidence that the travel cost parameter enters the utility function differently in SP and RP, welfare estimates do not differ substantially when allowing for the TC or any other parameters to differ across datasets.

Common practice in contingent behavior recreational choice modeling is to include a general opt-out option (i.e., not to engage in the recreational activity or to stay at home). We add visiting other beach sites and other non-beach sites outside the study areas as alternatives. We jointly estimate these two different types of demand in the same RUM framework. When modeling recreational choices, excluding the alternatives to engage in other recreational sites is in fact a reduction in the available choice set. We show that excluding viable substitutes to beach recreation may lead to a slight underestimation of welfare loss.

Our study has limitations. We exploit a simple CB question which requires some assumptions when combining the revealed and stated preference data. Our SP data is comprised of a single choice task, which complicates identification of relevant parameters. While the evidence suggests that the revealed and stated preference data are compatible, future combinations of RP-SP data should include sensitivity analysis around these decisions.

References

Abildtrup, J., Olsen, S.B., Stenger, A., 2015. Combining RP and SP data while accounting for large choice sets and travel mode – an application to forest recreation. J. Environ. Econ. Policy 4, 177–201. https://doi.org/10.1080/21606544.2014.986210

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. J. Environ. Econ. Manag. 26, 271–292. https://doi.org/10.1006/jeem.1994.1017

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., Williams, M., 1997. Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. J. Environ. Econ. Manag. 32, 65–84.

Aldy, J.E., 2014. The labor market impacts of the 2010 Deepwater Horizon oil spill and offshore oil drilling moratorium. National Bureau of Economic Research.

Alvarez, S., Larkin, S.L., Whitehead, J.C., Haab, T., 2014. A revealed preference approach to valuing non-market recreational fishing losses from the Deepwater Horizon oil spill. J. Environ. Manage. 145, 199–209. https://doi.org/10.1016/j.jenvman.2014.06.031

Azevedo, C.D., Herriges, J.A., Kling, C.L., 2003. Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations. Am. J. Agric. Econ. 85, 525–537.

Bonnieux, F., Rainelli, P., 2003. Lost recreation and amenities: the Erika spill perspectives. Rennes, INRA.

Boxall, P.C., Englin, J., Adamowicz, W.L., 2003. Valuing aboriginal artifacts: a combined revealed-stated preference approach. J. Environ. Econ. Manag. 45, 213–230. https://doi.org/10.1016/S0095-0696(02)00063-3

Bradley, M., Daly, A., 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21, 167–184. https://doi.org/10.1007/BF01098791

Bui, L., Sæland, S., 2017. Verdsetting av et gode uten markedsverdi: Tilfellet rekreasjon på Jærstrendene.

Cano-Urbina, J., Clapp, C.M., Willardsen, K., 2019. The effects of the BP Deepwater Horizon oil spill on housing markets. J. Hous. Econ. 43, 131–156.

Carson, R.T., Mitchell, R.C., Hanemann, M., Kopp, R.J., Presser, S., Ruud, P.A., 2003. Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. Environ. Resour. Econ. 25, 257–286.

Casey, J.F., Kahn, J.R., Rivas, A.A.F., 2008. Willingness to accept compensation for the environmental risks of oil transport on the Amazon: A choice modeling experiment. Ecol. Econ. 67, 552–559. https://doi.org/10.1016/j.ecolecon.2008.01.006

Cha, W., Melstrom, R.T., 2018. Catch-and-release regulations and paddlefish angler preferences. J. Environ. Manage. 214, 1–8. https://doi.org/10.1016/j.jenvman.2018.02.054

Cheng, L., Lupi, F., 2016. Combining Revealed and Stated Preference Methods for Valuing Water Quality Changes to Great Lakes Beaches 50.

Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. Transportation 39, 19–31. https://doi.org/10.1007/s11116-011-9331-3

Domínguez Alvarez, R., Loureiro, M.L., 2013. Environmental accidents and stigmatized fish prices: Evidence from the Prestige oil spill in Galicia. Econ. Agrar. Recur. Nat. 13, 103–126.

English, E., von Haefen, R.H., Herriges, J., Leggett, C., Lupi, F., McConnell, K., Welsh, M., Domanski, A., Meade, N., 2018. Estimating the value of lost recreation days from the Deepwater Horizon oil spill. J. Environ. Econ. Manag. 91, 26–45. https://doi.org/10.1016/j.jeem.2018.06.010

Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2010. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Mark. Sci. 29, 393–421. https://doi.org/10.1287/mksc.1090.0508

Glasgow, G., Train, K., 2018. Lost Use-Value from Environmental Injury When Visitation Drops at Undamaged Sites. Land Econ. 94, 87–96. https://doi.org/10.3368/le.94.1.87

Grijalva, T.C., Berrens, R.P., Bohara, A.K., Shaw, W.D., 2002. Testing the Validity of Contingent Behavior Trip Responses. Am. J. Agric. Econ. 84, 401–414.

Haener, M.K., Boxall, P.C., Adamowicz, W.L., 2001. Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior? Am. J. Agric. Econ. 83, 629–642. https://doi.org/10.1111/0002-9092.00183

Hanemann, W.M., 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. Am. J. Agric. Econ. 66, 332–341. https://doi.org/10.2307/1240800

Hausman, J., Leonard, G.K., McFadden, D., 1995. A utility-consistent, combined discrete choice and count data model Assessing recreational use losses due to natural resource damage. J. Public Econ. 56, 1–30.

Hensher, D.A., 2012. Accounting for scale heterogeneity within and between pooled data sources. Transp. Res. Part Policy Pract. 46, 480–486. https://doi.org/10.1016/j.tra.2011.11.007

Hensher, D.A., Bradley, M., 1993. Using stated response choice data to enrich revealed preference discrete choice models. Mark. Lett. 4, 139–151. https://doi.org/10.1007/BF00994072

Hensher, D.A., Rose, J.M., Greene, W.H., 2008. Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. J. Transp. Geogr. 16, 126–133. https://doi.org/10.1016/j.jtrangeo.2007.07.001

Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application [WWW Document]. Choice Model. Cent. R Package Version 007 WwwApolloChoiceModellingcom. URL (accessed 5.16.19).

Hess, S., Rose, J.M., 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39, 1225–1239. https://doi.org/10.1007/s11116-012-9394-9

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. J. Choice Model. 23, 1–8. https://doi.org/10.1016/j.jocm.2017.03.001

Hicks, R.L., Strand, I.E., 2000. The extent of information: its relevance for random utility models. Land Econ. 76, 374–385.

Jeon, H., Herriges, J.A., 2017. Combining Revealed Preference Data with Stated Preference Data: A Latent Class Approach. Environ. Resour. Econ. 68, 1053–1086. https://doi.org/10.1007/s10640-016-0060-0

Jeon, Y., Herriges, J.A., 2010. Convergent Validity of Contingent Behavior Responses in Models of Recreation Demand. Environ. Resour. Econ. 45, 223–250. https://doi.org/10.1007/s10640-009-9313-5

Kipperberg, G., Onozaka, Y., Bui, L.T., Lohaugen, M., Refsdal, G., Sæland, S., 2019. The impact of wind turbines on local recreation: Evidence from two travel cost method – contingent behavior studies. J. Outdoor Recreat. Tour. 25, 66–75. https://doi.org/10.1016/j.jort.2018.11.004

Kleppe, A., Jensen, J., 2018. Valuing the Recreational Benefits og Bore and Hellestø Beaches.

Landry, C.E., Allen, T., Cherry, T., Whitehead, J.C., 2012. Wind turbines and coastal recreation demand. Resour. Energy Econ. 34, 93–111. https://doi.org/10.1016/j.reseneeco.2011.10.001

Leggett, C., Scherer, N., Curry, M., Bailey, R., Haab, T.C., 2014. Assessing the Economic Benefits of Reductions in Marine Debris: a pilot study of beach recreation in Orange County, California. Industrial Economics, Incorporated.

Loomis, J.B., 1997. Panel Estimators to Combine Revealed and Stated Preference Dichotomous Choice Data. J. Agric. Resour. Econ. 22, 1–13.

Lopes, A.F., Mariel, P., 2020. Estimating the economic benefits of coastal quality change: An Application to Beach Recreation in Norway. Manuscript.

Loureiro, M.L., Loomis, J.B., 2013. International public preferences and provision of public goods: assessment of passive use values in large oil spills. Environ. Resour. Econ. 56, 521–534.

Loureiro, M.L., Ribas, A., López, E., Ojea, E., 2006. Estimated costs and admissible claims linked to the Prestige oil spill. Ecol. Econ. 59, 48–63. https://doi.org/10.1016/j.ecolecon.2005.10.001

Lundhede, T.H., Olsen, S.B., Jacobsen, J.B., Thorsen, B.J., 2009. Handling respondent uncertainty in Choice Experiments: Evaluating recoding approaches against explicit modelling of uncertainty. J. Choice Model. 2, 118–147. https://doi.org/10.1016/S1755-5345(13)70007-1

Mariel, P., Meyerhoff, J., 2018. A More Flexible Model or Simply More Effort? On the Use of Correlated Random Parameters in Applied Choice Studies. Ecol. Econ. 154, 419–429. https://doi.org/10.1016/j.ecolecon.2018.08.020

McFadden, D., Train, K., 2000. Mixed MNL Models for Discrete Response. J. Appl. Econom. 15, 447–470.

Morikawa, T., 1994. Correcting state dependence and serial correlation in the RP/SP combined estimation method. Transportation 21, 153–165. https://doi.org/10.1007/BF01098790

Murdock, J., 2006. Handling unobserved site characteristics in random utility models of recreation demand. J. Environ. Econ. Manag. 51, 1–25. https://doi.org/10.1016/j.jeem.2005.04.003

Parsons, G., K. Kang, A., Leggett, C., Boyle, K., 2009. Valuing Beach Closures on the Padre Island National Seashore. Mar. Resour. Econ. 24, 213–235. https://doi.org/10.5950/0738-1360-24.3.213

Parsons, G.R., 2017. The Travel Cost Model, in: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation, The Economics of Non-Market Goods and Resources. Springer

Netherlands, Dordrecht, pp. 269–329. https://doi.org/10.1007/978-94-007-0826-6_9

Parsons, G.R., 2008. Monetary Values and Restoration Equivalents for Lost Recreational Services on the Gulf Coast of Texas Due to Oil Spills and Other Environmental Disruptions (Final Report Submitted to The Coastal Response Research Center).

Parsons, G.R., Stefanova, S., 2011. Gauging the Value of Short-Term Site Closures in a Travel-Cost RUM Model of Recreation Demand With a Little Help from Stated Preference Data, in: Preference Data for Environmental Valuation: Combing Revealed and Stated Preference Approaches, Routledge Explorations in Environmental Economics. Routledge, Taylor & Frances Group.

Revelt, D., Train, K., 1998. Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level. Rev. Econ. Stat. 80, 647–657.

Scott, A., 1965. The valuation of game resources: some theoretical aspects. Can. Fish. Rep. 4, 27–47.

Simões, P., Barata, E., Cruz, L., 2013. Joint estimation using revealed and stated preference data: An application using a national forest. J. For. Econ. 19, 249–266. https://doi.org/10.1016/j.jfe.2013.03.001

Small, K.A., Rosen, H.S., 1981. Applied Welfare Economics with Discrete Choice Models. Econometrica 49, 105–130. https://doi.org/10.2307/1911129

Stafford, T.M., 2018. Accounting for outside options in discrete choice models: An application to commercial fishing effort. J. Environ. Econ. Manag. 88, 159–179. https://doi.org/10.1016/j.jeem.2017.10.006

Stratus Consulting Inc., 2010. Appendix J. Damage Estimate for Shoreline Recreation. Prepared for: Cosco Busan Natural Resource Damage Assessment, Washington, DC.

Sveen, G., 2018. Jærstrendene er nå av de mest besøkte naturattraksjonene i Norge. Stavanger Aftenbladet 3.

Swait, J., Louviere, J., 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. J. Mark. Res. 30, 305–314. https://doi.org/10.2307/3172883

Train, K., 2003. Discrete choice methods with simulation. Cambridge University Press, New York.

Truong, T., Adamowicz, W., Boxall, P.C., 2018. Modelling the Effect of Chronic Wasting Disease on Recreational Hunting Site Choice Preferences and Choice Set Formation over Time. Environ. Resour. Econ. 70, 271–295. https://doi.org/10.1007/s10640-017-0120-0

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. Sci. Total Environ. 551–552, 367–375. https://doi.org/10.1016/j.scitotenv.2016.02.011

von Haefen, R.H., Phaneuf, D.J., 2008. Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference approach. J. Environ. Econ. Manag. 56, 19–32. https://doi.org/10.1016/j.jeem.2008.01.002

Whitehead, J., Lew, D.K., 2019. Estimating Recreation Benefits through Joint Estimation of revealed and Stated Preference Discrete Choice Data. Empir. Econ.

Whitehead, J.C., Haab, T., Larkin, S.L., Loomis, J.B., Alvarez, S., Ropicki, A., 2018. Estimating Lost Recreational Use Values of Visitors to Northwest Florida due to the Deepwater Horizon Oil Spill Using Cancelled Trip Data. Mar. Resour. Econ. 33, 119–132. https://doi.org/10.1086/697560

Whitehead, J.C., Pattanayak, S.K., Houtven, G.L.V., Gelso, B.R., 2008. Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science. J. Econ. Surv. 22, 872–908. https://doi.org/10.1111/j.1467-6419.2008.00552.x

Winkler, D.T., Gordon, B.L., 2013. The Effect of the BP Oil Spill on Volume and Selling Prices of Oceanfront Condominiums. Land Econ. 89, 614–631. https://doi.org/10.3368/le.89.4.614

Yi, D., Herriges, J.A., 2017. Convergent Validity and the Time Consistency of Preferences: Evidence from the Iowa Lakes Recreation Demand Project. Land Econ. 93, 269–291. https://doi.org/10.3368/le.93.2.269

Zimmer, N.M.P., Boxall, P.C., Adamowicz, W.L., 2012. The Impacts of Chronic Wasting Disease and its Management on Recreational Hunters. Can. J. Agric. Econ. Can. Agroeconomie 60, 71–92. https://doi.org/10.1111/j.1744-7976.2011.01232.x

Appendices

Appendix 1 – Full Regression Results

	SP Data		RP Data		RP & SP Data	
	Conditional Logit Model		Mixed Logit Model		Mixed Logit Model	
	Estimate	**Robust s. e.**	**Estimate**	**Robust s. e.**	**Estimate**	**Robust s. e.**
TC	**-0.002**	0.001	**-3.653**	0.057	**-3.901**	0.057
SQRT(PROX)	**0.250**	0.049			**0.428**	0.143
ASC1 (fixed)			0.000		0.000	
ASC2			**0.438**	0.125	**0.284**	0.251
ASC3			-0.561	0.298	**-0.514**	0.256
ASC4			**-1.939**	0.320	**-3.099**	0.888
ASC5			**-2.106**	0.215	**-1.675**	0.431
ASC6			**0.242**	0.115	**-0.228**	0.258
ASC7			**-1.588**	0.201	**-1.694**	0.450
ASC8			**-0.505**	0.184	**-0.784**	0.234
ASC9			**-3.046**	0.428	**-2.570**	0.777
ASC10			**-1.549**	0.478	**-2.144**	0.758
ASC11			**-1.816**	0.344	**-1.440**	0.995
ASC12			**0.624**	0.213	**0.581**	0.290
ASC13			**0.943**	0.227	**1.036**	0.500
ASC14			**-3.529**	0.464	**-4.242**	1.764
ASC15			0.223	0.244	**0.027**	0.260
ASC16			**-1.736**	0.444	**-2.237**	1.442
ASC17			-0.852	0.580	**-0.718**	0.861
ASC18			**-1.204**	0.252	**-1.538**	0.964
ASC19			**-2.921**	0.344	**-2.997**	0.822
ASC20			**-2.914**	0.428	**-2.253**	2.781
ASC21			**-1.637**	0.329	**-1.069**	0.308
ASC22			**-3.955**	0.780	**-3.812**	1.086
ASC23			**-7.127**	0.795	**-6.238**	5.314
ASC24			**-5.419**	0.541	**-4.341**	2.938
ASC25			**-2.663**	0.803	**-4.554**	2.992
ASC26			**-4.003**	0.677	**-3.203**	1.509
ASC27	-0.737	0.448			**-9.739**	3.835
ASC28	-0.021	0.192			**-3.520**	1.696
ASC29	-0.461	0.440			**-5.480**	3.149
Standard deviations						
Sigma TC			0.836	0.044	**0.847**	0.095
Sigma SQRT(PROX)					0.163	0.246

Sigma ASC1		**1.382**	0.115	**1.069**	0.140
Sigma ASC2		**-1.142**	0.152	**1.038**	0.194
Sigma ASC3		**-1.755**	0.224	**1.939**	0.205
Sigma ASC4		**-0.980**	0.363	**-1.858**	0.899
Sigma ASC5		**1.554**	0.130	**1.394**	0.276
Sigma ASC6		**0.710**	0.159	**1.424**	0.285
Sigma ASC7		**1.044**	0.177	**-1.265**	0.379
Sigma ASC8		**-1.434**	0.126	**1.462**	0.188
Sigma ASC9		**-1.882**	0.348	1.303	0.896
Sigma ASC10		0.849	0.448	**1.454**	0.531
Sigma ASC11		**1.178**	0.252	-0.037	0.826
Sigma ASC12		**0.921**	0.154	0.433	0.359
Sigma ASC13		**-1.570**	0.229	**1.074**	0.255
Sigma ASC14		0.419	0.466	-1.099	3.545
Sigma ASC15		**1.279**	0.259	**1.160**	0.190
Sigma ASC16		-0.354	0.607	0.568	1.435
Sigma ASC17		**-1.181**	0.494	-0.893	0.672
Sigma ASC18		**0.698**	0.321	0.706	1.631
Sigma ASC19		**1.274**	0.280	1.301	0.983
Sigma ASC20		**-1.638**	0.246	1.121	3.290
Sigma ASC21		**1.797**	0.269	0.934	0.598
Sigma ASC22		**2.302**	0.297	**-2.129**	0.604
Sigma ASC23		**1.286**	0.275	0.410	10.477
Sigma ASC24		**-1.809**	0.164	0.637	1.180
Sigma ASC25		**-0.666**	0.085	1.843	2.062
Sigma ASC26		**-1.483**	0.501	0.442	0.864
Sigma ASC27				**7.131**	3.198
Sigma ASC28				1.566	2.604
Sigma ASC29				-5.159	6.276
Scale parameters					
RP (fixed)				1.000	
SP				**0.983**	0.229
# of Parameters	5		53		62
# of Individuals	647		647		647
# of Observations	647		5985		6632
Log-likelihood value	-1414.38		-12178.12		-13348.71
AIC	2839		24462		26821
BIC	2861		24817		27243

Notes: Coefficients in bold represent statistical significance at the 5% level. We used 500 Halton draws when running the mixed logit models. Models were estimated using the Apollo package in R (Hess and Palma, 2019).

Appendix 2 – Coefficient Estimates for different models using the RP-SP data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Conditional Logit Model	**Mixed Logit Model**	**Mixed Logit Model**	**Mixed Logit Model**	**Mixed Logit Model**	**Generalized Mixed Logit Model**	**Mixed Logit Model**	**Mixed Logit Model**	**Mixed Logit Model**
Travel Cost									
TC normal				**-0.02**		**-0.0224**			
TC Log-normal		**-3.8501**					**-3.8740**	**-3.9431**	**-3.9621**
TC triangular					**-0.05**				
TC Fixed	**-0.0139**		**-0.0144**						
Proximity SQRT(PROX)	**0.4993**	1.0134	**0.4927**	0.3972	**0.3577**	**0.6086**	0.4333	0.5107	0.7020
Alternative Specific Constants									
ASC27	**-3.9034**	-9.4492	-6.1317	-5.8889	**-6.0762**	**-7.0123**	**-8.1734**	**-6.2185**	-13.44
ASC28	**-2.0438**	-3.2375	-2.3201	-2.2174	**-2.9847**	**-4.6493**	-1.9553	-1.8889	-8.2166
ASC29	**-2.1938**	-6.614	**-5.6464**	-3.409	**-2.9844**	**-4.6374**	-3.8819	**-3.6106**	-9.8982
Standard deviations									
Sigma TC		**0.9576**		**0.0119**	**0.032**	**0.0161**	**0.8097**	**0.8137**	**0.7452**
Sigma SQRT(PROX)		0.685	-0.0298	0.0499	-0.1071	**0.2257**	0.0086	0.0587	0.1230
Sigma ASC27			3.6688	3.1416	**3.2752**	**1.1449**	**5.9942**	**5.0007**	6.6904
Sigma ASC28			0.5595	-0.2556	1.5536	**2.1794**	0.5876	0.6745	0.8640
Sigma ASC29			-6.3541	-2.8644	-1.2254	**0.9692**	**-2.1727**	-3.9057	-4.8539
SP Interactions									

TC							**0.0115**	**0.0093**	
SQRT(PROX)								0.1832	
Scale parameters									
RP (fixed)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SP	**0.4251**	**0.419**	**0.7392**	**0.7744**	**0.7936**		**1.1288**	**0.3777**	**1.4822**
Tau (τ)						**0.4184**			
SP (η)						**0.9892**			
# of Parameters	31	33	61	62	62	52	63	90	527
# of Individuals	647	647	647	647	647	647	647	647	647
# of Observations	6632	6632	6632	6632	6632	6632	6632	6632	6632
Log-likelihood value	-15927	-14883	-13585	-13339	-13379	-11938	-13306	-13239	-12386
AIC	31915	29832	27292	26802	26882	23979	26738	26658	25826
BIC	32126	30056	27707	27223	27304	23418	27166	27270	29410

Notes: Coefficients in bold represent statistical significance at the 5% level. We used 500 Halton draws when running all the mixed logit models. Models were estimated using the Apollo package in R (Hess and Palma, 2019). The Generalized Mixed Logit Model was estimating using NLOGIT. We estimate a full set of ASC (1 through 29) and standard deviations for the distributions of ASC1 through 29 in both the RP and RPSP models but ASC1-26 are omitted for the sake of brevity. Model 7 also included a full set of interactions of the SP dummy with ASCs. Model 9 estimated a full covariance matrix, but parameters are omitted for the sake of brevity.

Appendix 3 – Welfare estimates (CS) using the RP-SP models reported in Appendix 2 and referred to in the Discussion

		Small	***Medium***	***Large***	***Very Large***
	Mixed Logit Model (Baseline)	-122.94	-188.30	-261.62	-288.64
(1)	Conditional Logit Model	-157.11	-222.36	-244.22	-244.21
(2)	Mixed Logit Model	-481.72	-664.26	-731.84	-734.52
(3)	Mixed Logit Model	-137.60	-214.26	-296.16	-308.49
(4)	Mixed Logit Model	-83.02	-130.60	-177.47	-186.09
(5)	Mixed Logit Model	-119.52	-194.25	-267.39	-287.50
(6)	Generalized Mixed Logit Model	-169.11	-242.53	-276.85	-280.51
(7)	Mixed Logit Model	-95.76	-152.43	-214.29	-227.04
(8)	Mixed Logit Model	-106.99	-166.51	-210.71	-219.08
(9)	Mixed Logit Model	-186.86	-302.70	-422.07	-464.87
	SP Data (with missing RP-parameters adjusted)	-284.7	-396.6	-391.4	-384
	RP Data (with missing SP-parameters adjusted)	-723.7	-1206.4	-1551.7	-1630.3
	Without alternatives 28 & 29	-82.18	-138.83	-213.05	-237.80