



Substitution Effects in Spatial Discrete Choice Experiments

Marije Schaafsma¹ · Roy Brouwer^{2,3}

Accepted: 13 August 2019 / Published online: 24 August 2019
© The Author(s) 2019

Abstract

This paper explores spatial substitution patterns using a choice experiment to estimate the non-market benefits of environmental quality improvements at different sites presented as labelled alternatives. We develop a novel modelling approach to estimate possible disproportional substitution patterns among these alternatives by including cross-effects in site-specific utility functions, combining mixed and universal logit models. The latter model allows for more flexibility in substitution patterns than random parameters and error-components in mixed logit models. The model is relevant to any discrete choice study that compares multiple sites that vary in their comparability and that may be perceived as (imperfect) substitutes. Applying the model in an empirical case study shows that accounting for cross-effects results in a better model fit. We discuss the validity of welfare estimates based on the inclusion of cross-effects. The results demonstrate the importance of accounting for substitution effects in spatial choice models with the aim to inform policy and decision-making.

Keywords Substitution · Spatial choice experiments · Cross-effects · Universal logit model

JEL Classification Q51 · Q25 · H41

1 Introduction

Many environmental valuation studies involve spatial choices in which subjects (respondents in a survey) and objects (sites, ecosystem services) are non-uniformly distributed over space (Bateman et al. 2003). The distance between objects and between subjects and objects are often among the key factors that determine the prices and substitutability of environmental goods and services (Pellegrini and Fotheringham 2002). Substitution effects are expected to be especially relevant when estimating the non-market benefits of policies

✉ Marije Schaafsma
M.Schaafsma@soton.ac.uk

Roy Brouwer
rbrouwer@uwaterloo.ca

¹ School of Geography and Environmental Science, University of Southampton, Southampton, UK

² Department of Economics, University of Waterloo, Waterloo, Canada

³ Department of Environmental Economics, Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

aiming to improve environmental quality across different sites that provide similar *types* of environmental use and non-use values in a confined geographical area, but vary in the magnitude of service provision and substitutability as perceived by respondents (e.g., Hanley et al. 2006). Based on theory (e.g. Bateman et al. 2002) and empirical findings going back as far as to the 1980s (e.g. Sutherland and Walsh 1985), WTP is expected to decline, in particular for use values, as respondents live further away from the site under valuation, which is known as distance-decay. For many environmental goods, the number of available substitutes increases with distance from the site, thereby contributing to distance-decay. The distance from respondents to the locations providing ecosystem services hence affects the substitutability of these services.

Choice behaviour across multiple sites is often modelled using random utility models (RUM), because of their suitability to account for site-substitution (Parsons 2003). Travel-cost studies use RUM to model recreation demand as a function of site access and site-characteristics, including, for example, water quality levels (Kaoru 1995; Needelman and Kealy 1995; Parsons and Massey 2003). More sophisticated studies include a wide range of spatial variables, such as the distribution of substitutes through the model error structure (e.g., Termanen et al. 2008). These spatial choice studies typically use revealed preferences data and therefore provide no information about non-use values. Whilst stated preference (SP) studies allow estimation of non-use values, there are surprisingly few SP studies that account for substitution effects on willingness to pay (WTP) for environmental benefits that may arise when multiple changes take place simultaneously (Carson et al. 2001). Contingent valuation (CV) studies usually focus on a single site, and at most remind respondents of other expenditure options in the survey text following NOAA-panel guidelines (Arrow et al. 1993). These SP studies suffer from the same limitations as single-site travel cost studies: they do not account for changes in the availability and characteristics of relevant alternatives. A small number of CV studies include multi-programme scenarios in which different goods are valued simultaneously to test for substitution and complementarity (e.g., Hoehn and Loomis 1993; Hailu et al. 2000). However, these CV studies too are usually limited to the estimation of the effect of the availability (absence or presence) of alternative policy scenarios at different locations. They do not estimate the extent to which the utility of one alternative is affected by changes in the different characteristics, including the price, of another alternative.

Unlike the CV method, discrete choice experiments (DCEs) can be used to present multiple sites or destinations as alternatives or attributes in choice sets (Rolfe et al. 2002; Brouwer et al. 2010). Some recent site choice experiments studies in the SP literature include distance and substitute characteristics as explanatory variables in the indirect utility function. For example, Nielsen et al. (2016) demonstrate how local forest availability and quality affect preferences in a DCE focusing on a national wildlife protection programme. De Valck et al. (2017) account for the distance to substitute sites in unlabelled DCEs related to site-changes. Applying labelled site DCEs, Schaafsma et al. (2012, 2013) demonstrate how distance decay functions vary across sites, attributed to the presence of substitutes. Similarly, Lizin et al. (2016) test the effect of spatial heterogeneity on welfare estimates and transfer errors for river restoration in labelled river-specific utility functions, accounting for key variables such as site visitation, distance-decay and spatial clustering. Schaafsma and Brouwer (2013) find that expanding the number of sites, presented as labelled alternatives, does not affect WTP values.

These approaches provide more explicit indications of substitution effects related to options outside the choice set. As substitution effects create correlation between alternatives, different discrete choice models have been developed that allow for flexible substitution patterns in the stochastic component of the RUM (Train 2002). The inclusion of

random parameters and error-components not only captures respondent heterogeneity, but also accommodates correlation between alternatives. However, such models provide little insight into different disproportional shifts in substitution patterns that may arise when some alternatives are more similar than others, an effect that has been addressed in marketing and retail studies on product assortment decisions (e.g. Tversky 1972; Roederkerk et al. 2011; Lipovetsky and Conklin 2014).

In this paper, we develop a novel modelling approach to reveal how mean WTP estimates change as characteristics of substitute sites in the choice set change, by adjusting the systematic (rather than stochastic) part of the model. We present a labelled site-selection DCE, designed to capture the effect of changes in ecosystem services provision at one site on the WTP for substitute sites. The selected case study sites vary in their ecological functionality and are therefore imperfect substitutes, leading to the expectation underlying our central hypothesis that disproportional substitution patterns play a role. To identify and model disproportional substitution patterns, we combine the advantages of mixed and universal logit models (McFadden 1975). We show that accounting for disproportional substitution may lead to different WTP estimates compared to standard mixed logit models, but emphasise that no consensus exists in the literature as to whether these WTP estimates have the same conventional welfare interpretation. We find significant cross-effects, reflecting a correction of the utility of one site predicted by the effect of changes in the attributes of other alternatives. We show that these cross-effects cannot be captured by random parameters and error-components alone to control for taste heterogeneity and correlation between choice tasks, alternatives and attributes. These cross-effects imply that the composition of the choice set, i.e. the inclusion of substitutes to which the site of interest is compared, affects the mean WTP estimates. This approach has, as far as we know, not been used before to capture substitution across sites in environmental valuation and offers more flexibility than the commonly used mixed logit models. The main contribution of our paper to the DCE literature is that we show that spatial substitution patterns may be more complex than commonly applied mixed logit models can account for and that such complex substitution patterns may potentially lead to different WTP estimates and probability shares.

The next section discusses different possibilities to estimate substitution effects using DCE models. Section 3 presents our modelling approach and Sect. 4 the design of our experiment. The results of the estimated choice models and WTP are presented in Sect. 5. Finally, Sect. 6 concludes and discusses the implications and limitations of the study.

2 Modelling Substitution Effects

Discrete choice models are rooted in random utility maximization theory (McFadden 1974). The general indirect utility function U_j , representing the utility that an individual derives from an alternative j , decomposes into two parts: a deterministic (V_j) and stochastic part (ε_j) (Eq. 1). V is usually specified as a linear function, additive in utility, where X is a vector of k attributes associated with alternative j , β is a coefficient vector, and ε represents unobservable influences on individual choice:

$$U_j = V_j + \varepsilon_j = \beta_k X_{kj} + \varepsilon_j \quad (1)$$

Substitutability plays a central role: consumers are assumed to choose the alternative that provides the highest utility among goods that are substitutable, at least in the margin. That is, giving up a unit of one good to obtain an additional unit of another good results

in the same level of utility. The marginal rate of substitution (MRS) between two goods reflects the trade-off that people are willing to make between these goods and is determined by personal preferences for the characteristics of the goods. In DCEs, the MRS indicates how much attribute k_1 of an alternative will have to change for a unit change in another attribute k_2 so that the probability of choosing that alternative remains unchanged (Hensher et al. 2005a). A distinction can be made between direct marginal effects and cross-marginal effects. Direct marginal effects are the changes in the probability of choosing an alternative in the choice set given a unit change in an attribute of the alternative itself, whereas cross-marginal effects reflect the probability changes given a unit change in a competing alternative. The latter thus capture the degree of substitutability between alternatives.

The basic discrete choice model, the multinomial logit (MNL) model (McFadden 1974), assumes that the random components of the utility of the alternatives are independently and identically (Gumbel) distributed (*i.i.d.*) with a type I extreme value (EV) distribution. In the MNL model, substitution patterns are defined by the Independence of Irrelevant Alternatives (IIA) restriction. The IIA assumption states that the relative probabilities of two alternatives are unaffected by the existence and attributes of other alternatives (Kaninen 2007). The IIA property follows directly from the *i.i.d.* EV error terms. It imposes proportional substitution rates on alternatives instead of allowing individual substitution patterns to vary depending on the similarity of alternatives. This property is also known as the Invariant Proportion of Substitution (IPS) (Steenburgh 2008), and implies that no matter which attribute of a good changes, it will draw choices away from competing alternatives proportional to the initial probability shares. This assumption is likely to be too restrictive for spatial choice studies as many studies often involve more than two alternatives of which some may be perceived as more similar than others.

Alternative modelling approaches have been developed to relax the IIA assumption. Generalised extreme value (GEV) models, including nested and cross-nested models, have a closed-form formula and allow for correlation in the error term over alternatives. Random parameter logit (RPL) and error-component (EC) models account for respondent differences (taste heterogeneity), correlation across alternatives and repeated choices (Train 2002). By specifying random attribute parameters that are common across alternatives or permitted to be correlated, it is possible to allow for correlation among alternatives (Hensher et al. 2005a). EC models accommodate correlation between the utilities of alternatives and can be applied when substitution patterns are disproportional (Brownstone and Train 1999). Here, correlation between alternatives is accounted for by including a common random parameter with zero mean in the utility function specification of those alternatives that are likely to be correlated (Herriges and Phaneuf 2002). Scarpa et al. (2005) recommend applying EC models when comparing less familiar (hypothetical) alternatives with better known (existing) ones. These models still assume that IIA holds at the individual respondent level (Hahn et al. 2007). As Steenburgh and Ainslie (2008) point out, although accommodating preference heterogeneity, it is not clear how the mixed logit model allows for individual choice behaviour to address different types of substitution patterns, for example when facing choices between perfect substitutes. This implies that the mixed logit model is less flexible than previous research has suggested.

Brownstone and Train (1999) and McFadden and Train (2000) show that mixed logit models can approximate any random utility model arbitrarily well, provided the researcher uses the correct mixing distribution. Hess and Train (2017) explore this potential by specifying the most general form of a mixed logit model in which all utility coefficients are randomly distributed and estimate a full covariance matrix among them. However, one of the main drawbacks of estimating RPL and EC models is that

the correct specification of the random parameters and their distribution may involve a long trial-and-error process. The same holds for the EC structure in the absence of a priori expectations regarding substitution patterns (Walker et al. 2007). Hess et al. (2017) operationalise the general mixed logit model proposed by Hess and Train (2017) using a sample of 5000 respondents facing 15 choices each. Such sample size requirements can be prohibitive in many applied studies. A further warning is that the correlation of random parameters is introduced through Cholesky factorization, which is sensitive to the order of the variables (Hensher et al. 2005a).

Given these limitations, alternatives to the mixed logit model seem relevant. Another possibility to relax the IIA and IPS assumptions and allow for disproportional substitution rates among alternatives is using the universal logit (UL) model (McFadden 1975; Louviere et al. 2000; Kuhfeld 2010). UL models offer the possibility to estimate cross-effects by including the attributes of other alternatives in the specification of the utility function of each alternative (Oppewal and Timmermans 1991). The cross-effects give corrections of the probability distribution across alternatives as predicted by the main effects of attribute changes, so they are not the same as cross-marginal effects. In an MNL model, the main effect of an attribute determines its direct marginal effect on the choice probability of its own alternative and its cross-marginal effect on the competing alternatives, where the cross-marginal effects lead to a proportional substitution pattern under the IIA restriction. The cross-effects of the UL model provide a correction of the proportional probability distribution. For example, a positive cross-effect indicates that the utility of an alternative is underestimated by an IIA model (Timmermans and Molin 2009). Significant cross-effects in the UL model imply that the effect of a marginal change of the attribute of alternative A on the choice probability of alternative B is not the same as the effect on the choice probability of alternative C. Thereby, the UL model is not restricted by the IIA assumption and permits disproportional substitutability, which may arise if some alternatives are perceived as closer substitutes than others (Steenburgh 2008). Together with the main effects of alternative-specific attributes, cross-effects determine the degree of substitutability of alternatives (Timmermans et al. 1991), where the disproportional substitution pattern is explained by changes in attributes of the alternatives. Since cross-effects reveal information about the effect of the presence or characteristics of alternatives on the alternative of interest, significant cross-effects can be interpreted as choice set composition or context-dependency effects. The UL has been applied in research on consumer shopping behaviour (Timmermans et al. 1991; Oppewal and Holyoake 2004), telecommunications (Agarwal 2002), transport economics (Bos and Molin 2006), food choices (Chowdhury et al. 2011), and tourism (Crouch et al. 2007).

UL models have two main disadvantages. First, UL models are closed-ended and therefore unable to account for the panel structure of datasets and random taste heterogeneity. Cross-effects can arise if population segments with different preferences are present in the sample (Crouch et al. 2007). Thus, if IIA violations are caused by preference heterogeneity and the panel structure of the data, specifying a RPL model may be sufficient to reflect the substitution pattern (Swait and Louviere 1993). However, as stated before, a mixed logit model is insufficient when IIA does not hold at individual level. Second, Ben-Akiva (1974) argues that the cross-effects can lead to “counter-intuitive” cross-elasticities, for instance when the sign of the cross-elasticity of the price variable switches from positive to negative. This may occur if a cross-effect of the price variable is negative and has a larger effect than the main effect of price changes. A practical remedy would be to exclude such cross-effects from the model, much like the suggestion to restrict the distribution of the price

parameter in mixed logit models in line with theoretical expectations (Hole and Kolstad 2012).

A key argument against the use of UL models is that “there is no easy way to tell whether a mother logit model was consistent with RUM” (McFadden 2000, p. 15). Cross-effects may lead to a violation of the regularity assumption (Oppewal and Timmermans 1991). The weak regularity assumption (Block and Marschak 1960) imposes a consistency restriction on choice behavior and dictates that the probability of choosing an alternative from any subset of a choice set of alternatives cannot rise if the choice set is expanded. For example, the probability of choosing alternative A in a comparison of A and B cannot be lower than the probability of choosing alternative A in a comparison of A, B and C. It is important to note here that the IIA is the strict version of the regularity assumption. Under the IIA assumption, the ratio of choice probabilities is not allowed to change if another alternative is added to the choice set; the weak regularity assumption allows for probabilities to decrease if the choice set is expanded. Brownstone and Train (1999, p. 113) seem to build on the strong regularity assumption when they argue that “when cross-alternative attributes are entered, the logit model is no longer a random utility model (i.e. is not consistent with utility maximizing behavior and cannot be used for welfare analysis) since the utility of one alternative depends on the attributes of other alternatives”. Hess et al. (2017, 2018) follow a similar logic and argue that the mother logit model may be problematic when the cross-substitution effects are captured through the observed component of utility as it “open(s) up the possibility of preference reversals” (Hess et al. 2018, p. 195). This might mean, according to the latter authors, that the mother logit model is not consistent with RUM, but they do not provide any proof why a mother logit model would (by definition) result in violations of transitivity assumptions. Carson et al. (1994) argue that “particular restrictions” must be imposed to ensure that the UL model is consistent with utility maximization, but do not describe these restrictions. Experimental studies show that the regularity assumption is sometimes violated and anomalies in choice behaviour are found when the choice set is expanded with an alternative that is similar (Tversky 1972) or dominant to one of the existing alternatives (Huber et al. 1982).

The existing discrete choice literature provides no theoretical or formal proof for the inconsistency of the universal logit model with RUT. However, when using UL models in cases where utility maximisation conditions are important, such as for societal welfare estimation, it is important to test if the regularity assumption (Timmermans and Molin 2009) and transitivity assumption (Hess et al. 2017) are violated. In cases where the regularity or transitivity assumptions are violated, the interpretation of the MRS of price and non-price attributes is different from RUM based models for welfare analysis and cannot be used in the same way as WTP estimates derived from a RUM-consistent model. Therefore, the WTP estimates based on the UL model presented in this study have to be interpreted with caution.

3 Modelling Framework

In our modelling approach, we develop a mixed logit model extended with cross-effects, combining the advantages of mixed and universal logit models to allow for more flexible substitution across the alternatives in the choice set. We control for correlation across observations of the same respondent, taste heterogeneity and additional disproportional substitution patterns. The main objective of our modelling approach is to reveal how

mean choice probabilities and WTP estimates change depending on which substitutes are included in the choice set, by adjusting the deterministic rather than the stochastic part of the model.

We present two models that will be compared with each other: one mixed logit model with random parameters and error-components (Model I), and one extended mixed logit model with cross-effects (Model II). The first model presented in Eq. (2) includes an alternative specific constant (ASC) α for each alternative location i to control for location characteristics not covered by the attributes X . The alternative-specific attribute parameters β_{ki} account for possible differences in the values of the same ecosystem services as per the attribute labels across sites (direct effects). We specify random parameters with normal distributions reflecting random taste heterogeneity for the attributes in X_{ki} , and test whether these are site-specific or common across alternatives. Random parameters that are common across alternatives allow for correlation between these alternatives. μ_{kin} is a vector of random coefficients of the attributes X_{ki} for individual n . The price attribute is included as a fixed parameter. An error-component is added to allow for heteroscedasticity between the hypothetical alternatives and the opt-out, following Scarpa et al. (2005). To this end, a dummy variable d_i taking the value 1 for each hypothetical alternative and 0 for the opt-out is included in the utility function of alternative i . Finally, x additional error-components c are added, mimicking a nested logit model. These error-components c take the value 1 for alternatives that are expected to be correlated and the value 0 otherwise. The λ parameters of these individual-specific random error-components c are assumed to have a zero-mean normal distribution $N[0, \sigma^2]$.

$$\text{Model I: } U_i = \alpha_i + \beta_{ki}X_{ki} + \mu_{kin}X_{ki} + \lambda_{in}d_i + \lambda_{in}^{1 \dots x} c_i^{1 \dots x} + \epsilon_{in} \tag{2}$$

$$\text{Model II: } U_i = \alpha_i + \beta_{ki}X_{ki} + \mu_{kin}X_{ki} + \beta_{kij}X_{kj} + \lambda_{in}d_i + \lambda_{in}^{1 \dots x} c_i^{1 \dots x} + \epsilon_{in} \tag{3}$$

We estimate Model II to assess if despite the flexibility provided through the random parameters and additional error-components in Model I, further disproportional correlation is present by including cross-effects, similar to a universal logit model. Hence, Model II aims to identify significant cross-effects, whilst accounting for site-specific values, random parameters, error-components and the panel structure of the choice data. The utility specification includes not only the alternative's own attributes X_{ki} , but also the attributes of alternative sites $j \neq i$ in X_{kj} . This model is used to capture possible disproportional substitution effects, reflected by the site-specific parameters β_{kij} representing the cross-effects of X_{kj} on the utility of alternative i . Significant cross-effects imply that disproportional substitution patterns between alternatives remain present despite the random parameters in the deterministic part and the error components in the stochastic part of the indirect utility function. When any two alternatives A and B are more similar compared to a third alternative C, a change in the characteristics of A would be expected to result in stronger substitution with B than with C, in which case cross-effects may be significant. Hence, we test whether it is sufficient to model substitution effects through a mixed logit specification or if it is necessary to extend the mixed logit specification with cross-effects to capture additional disproportional substitution patterns to better reflect choice behaviour. We test whether including cross-effects results in a better model fit and therefore explains choice behaviour better statistically.

We also discuss the WTP estimates that result from the inclusion of cross-effects, and evaluate whether these effects have a plausible explanation in the empirical case study application, but note that it is not clear whether these are consistent with RUT. We estimate the welfare measures of an environmental change at one or more sites, rather than at all

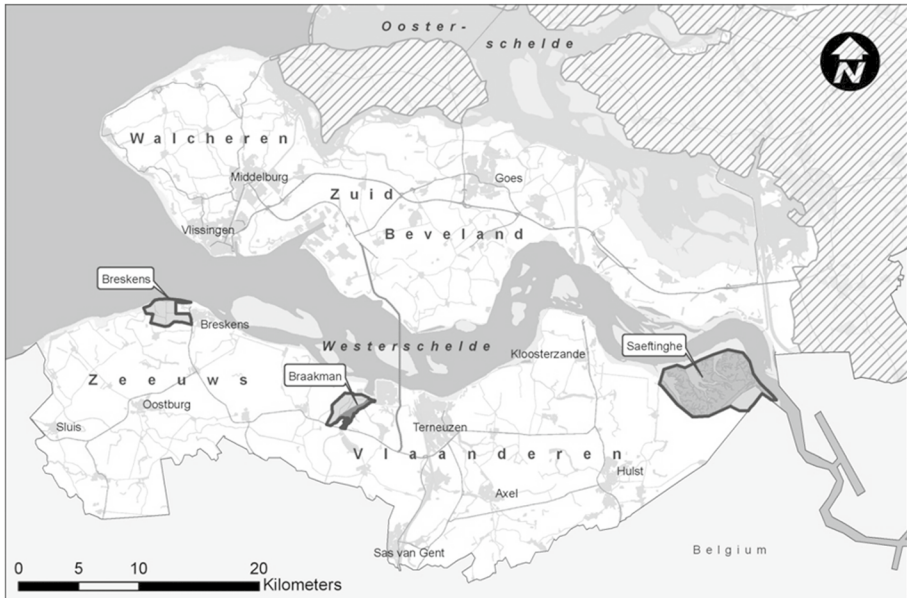


Fig. 1 Map of the study area and the three study sites Breskens, Braakman and Saeftinghe. *Note:* Belgium and the dashed areas were not included in the sample

sites at the same time, resulting from the same water quality improvement policy, considering relevant substitution effects assuming that the same calculation procedure can be applied as for standard mixed logit models (Bockstael and McConnell 2007). Since there is limited to no guidance in the literature on how to formally assess welfare estimates for models with cross-effects, the welfare estimates based on Model II presented in this paper should, as mentioned, be interpreted and used with caution. To test empirically if the cross-effects lead to violations of regularity conditions, we also simulate probability shares to check whether the resulting probabilities increase across alternatives if relevant attributes change. We first present the design of the choice experiment in the next section before turning to the estimation results in Sect. 5.

4 The Choice Experiment

To analyse the presence of disproportionate substitution patterns, a site-selection DCE was developed with alternative specific attribute descriptions and labelled alternatives based on location names. The popular tourist destination Zeeuws-Vlaanderen in the south-western part of the Netherlands within the Scheldt catchment was chosen as case study area. Choice data were collected here before as part of a European valuation study, testing the impact of distance-decay on recreational choice behaviour applying the spatial expansion method to account for directional differences in distance-decay (Schaafsma et al. 2013). Three recreational sites were selected in this area: Breskens, Braakman, and Saeftinghe. Figure 1 presents a map of the study area and the study sites. These three alternatives were included in the choice set after thorough pre-testing. Three alternatives are sufficient to analyse

disproportional substitution patterns without imposing too much of a cognitive burden on survey participants by including too many sites in the choice task.

The case study sites represent the most important water body types in the catchment, provide a broad range of nature-experience and water-recreation possibilities, and are well-known among residents. Because they vary in their functionality and the characteristics of their recreational and nature amenities, the sites are particularly suitable for our analysis of disproportional substitution patterns. Breskens is a popular beach site, attracting local, national, and even international visitors. Braakman is located at the mouth of a small river, has brackish water, and attracts especially families for water-based recreation. Saeftinghe does not provide any bathing opportunities, but is a unique, ecologically valuable tidal mudflat, which provides a habitat for various protected species. Some parts can only be visited when accompanied by a guide. For all three sites area-expansion plans exist, which will increase the potential for recreational walking. During the tourist season, Saeftinghe attracts mostly nature enthusiasts, while Breskens and Braakman attract more visitors who are interested in bathing. The latter sites are therefore expected to be closer substitutes compared to Saeftinghe.

The three sites are subject to environmental quality improvements under the European Water Framework Directive (WFD). The WFD imposes standards for all European water bodies and specifies water quality objectives in terms of ‘good ecological status’ (GES). Quality levels at the three sites did not meet the WFD standards at the time of the survey. Achieving the WFD objectives is expected to increase ecosystem services provision levels and generate substantial public use and non-use values, justifying the application of a SP method (Brouwer 2008). The use of a DCE enables assessing the trade-offs of the amenities subject to possible changes under the WFD and resident’s willingness to pay for these changes. The site-specific characteristics were translated into three easily understandable attributes: walking, bathing and nature quality. Table 1 gives an overview of the attributes and their levels used in the design of the DCE.

Since reaching the WFD objectives is expected to result in different bio-physical outcomes for the three selected locations, the foreseen quality improvements, reflected in the different levels of the attributes X_{ki} , are explained by site-specific descriptions and different photographs for walking and nature improvements. For bathing water quality, the same attribute description and pictures were used. The DCE design was developed in direct collaboration with ecologists and regional water managers. Important to note is that achieving the proposed quality improvements at one site will not improve conditions at other sites, because they function biophysically speaking completely separately from each other.

Two restrictions were imposed in the DCE design to increase the realism and credibility of the quality improvement scenarios. First, the current ecological quality was given a relatively poor rating at all three sites by the consulted ecologists, and therefore the status quo levels of the attributes are ‘poor’, except for current bathing water quality at Breskens, which is considered ‘moderate’. Second, the bathing water quality attribute was excluded at Saeftinghe, because this site does not (and is not envisioned to) provide any bathing possibilities. Site characteristics that are not affected by the WFD are not included as an attribute in the DCE, but their potential effect on choices is expected to be captured by including the site names as labels. The fourth and final attribute in the DCE was the monetary attribute, i.e. a proposed increase in annual water agency taxes paid by all local households. The six levels of this attribute ranged from 5 to 80 Euros per year.

The final design consisted of 24 different versions (blocks) of choice sets including five choice tasks each, where each block was used as a minimum 15 times. The order in which the sites appeared was kept the same in each choice task. The choice sets are

Table 1 Overview of attributes and levels

WFD objective	Attribute	Levels	Description
Hydro-morphology	Walking	Good	Area expansion and natural banks/dunes
		Moderate	No area expansion, natural banks/dunes
		Poor	Status quo: stone (artificial) banks/dunes
Chemical water quality	Bathing ^a	Good	High clarity, no algal blooms or odour
		Moderate	Moderate clarity, occasional algal blooms and odour
		Poor	Low clarity, frequent algal blooms and odour problems ^b
Ecology	Nature amenities	Good	High number and diversity of fish and birds, including rare species
		Moderate	Moderate number and diversity of fish and birds, few rare species
		Poor	Status quo: low number and diversity, hardly any rare species

^aThe bathing attribute is only specified for the sites Breskens and Braakman, as Saeftinghe does not provide bathing amenities

^bFor Breskens, the moderate level represents the status quo. No deterioration to a poor level was offered in the future scenarios. The poor level represents the status quo for Braakman

based on an experimental design that minimised the correlation among all attributes and alternatives and enabled the estimation of cross-effects. The efficiency design was based on (naïve) zero-priors (Walker et al. 2018) and approximated a full-factorial orthogonal design generated with Sawtooth Software (2008). The correlations of the design are minimal so we have no reason to expect that any cross-effects would be caused by correlations in the experimental design. Besides the practical advantage of limiting the number of choice sets for face-to-face interviews, the use of different versions of choice sets prevents confounding differences in individual preferences and error variability with differences in design, which could result if individualised designs would have been used.

The survey was pre-tested in four rounds over a three-month period at different locations in the study area. Based on the pre-test results, residents appeared to be sufficiently knowledgeable and familiar with the selected sites to be able to make well-informed choices. The final version of the questionnaire consisted of 45 questions, divided into five main parts, covering (1) general water recreation activities, (2) water recreation at the three selected sites and possible substitute sites, (3) the choice experiment, (4) socio-demographic characteristics, and (5) debriefing questions about the questionnaire itself. The second part included questions about visitation frequency, recreational activities carried out at each site, perception of water quality and other site characteristics. Apart from possible substitution between the three sites in the DCE, respondents were asked to list a maximum of three other recreational sites they visit frequently in the study area, how often they visit these sites and which activities they undertake at these locations. By including detailed questions about other sites people visited, the study went beyond a standard substitute reminder, the effectiveness of which is questionable (e.g., Loomis et al. 1994; Kotchen and Reiling 1999; Whitehead and Blomquist 1999). Instead, respondents were stimulated to think actively about the relevant alternatives in their choice set. They were also asked what they would do if their currently most preferred site in the DCE would face a decrease in quality, so that it could no longer offer the same recreational opportunities.

The DCE began with an explanation of the choice task, including an overview of all attributes and levels, the current situation at each of the sites, and an example choice task (see the example choice card in Fig. 2). Next, five different choice cards were presented to every respondent. Each card presented four alternatives: the three sites, which all improved in at least one attribute against a certain payment, and the 'opt-out' defined in terms of the current quality levels for all attributes at all sites at zero price. Respondents were asked to choose the site they preferred to be improved. It was emphasized that only one site could be chosen at the given price levels and that choosing the preferred improvement implied that their money would only be spent on the chosen site and other sites would remain at current levels. Each time respondents chose the opt-out, they were asked to motivate their answer in a follow-up question to identify possible protest responses.

In addition to the site-attributes, another important determinant of choice behaviour was the distance between the sites and the respondents. A geographical sampling strategy was followed to ensure sufficient variation exists in distance between respondents and the locations used in the DCE to be able to include distance as an implicit attribute of the sites (see Schaafsma et al. 2013 for more details). Distances from respondents to the three sites ranged from 2 to 160 kilometres. At the same time, the sample adequately reflected the geographical distribution of the population throughout the sampling area. The survey was implemented door-to-door using trained interviewers from July until September 2007 in 46 towns and villages in the sampling area (see Fig. 1).




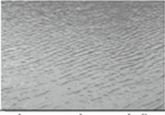
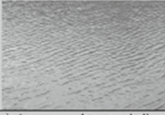



EXAMPLE	Beaches Breskens	Braakman	Land van Saeftinghe	
Walking	Moderate: Natural dunes with vegetation 	Bad: Stone banks without vegetation 	Good: area expansion of 300 ha and natural, overgrown dike 	
Swimming	Moderate: Moderate clarity, occasional algal blooms and odour 	Moderate: Moderate clarity, occasional algal blooms and odour 	Not applicable: Swimming impossible	
Nature	Poor: Low number and diversity of fish and birds, hardly any rare species: only seagulls 	Poor: Low number and diversity of fish and birds, hardly any rare species: only oyster catchers 	Good: High number and diversity of fish and birds, including rare species: spoonbills 	
Extra tax water agency	€ 20 per year	€ 20 per year	€ 10 per year	
Which option do you prefer?	<input type="checkbox"/> A	<input type="checkbox"/> B	<input type="checkbox"/> C	<input type="checkbox"/> none

Fig. 2 Example choice card

5 Results

5.1 General Survey Results

In total, 2322 households were approached, of which 1524 declined to participate. This corresponds to a response rate of 34%. After data cleaning, 96 respondents were removed from the sample because their replies were incomplete (e.g. basic information for the analysis related to income and distances to the sites). The sample includes more female than male respondents (61%). The average age of the respondents is 51. The average household size is 2.7, which is slightly higher than the Dutch average of 2.3. Disposable household income in the sample is, on average, €2143 per month, which is close to the region’s average net household income. Most respondents have a higher secondary education degree. Although women and bigger households are slightly overrepresented in the sample, this should not pose any problems for the methodological aim of this paper. Table 2 presents the key descriptive statistics for the sample used for model estimation (after removal of protest bids, see Sect. 5.2).

As expected given the abundance of water in the study area, most respondents (94%) visit open water for recreation. A fifth of the respondents have never visited any of the three sites in the CE, while another fifth has visited all three sites at least once. Less than 10% had never heard of any of the locations prior to the survey. Breskens is the most popular site: 62% of the respondents have visited this location at least once, compared to 45% for Braakman and 42% for Saeftinghe. Walking is one of the most popular

Table 2 Main descriptive statistics

Variable	Mean (SD)
Mean net household income (£/month)	2143 (821)
Open water recreators (% of respondents)	96
Percentage of users (% of respondents who have visited the site prior to the survey)	
Breskens	60
Braakman	44
Saeftinghe	40
Mean distance from respondent to the survey locations (km)	
Breskens	45 (25)
Braakman	33 (19)
Saeftinghe	46 (20)
Percentage of respondents who visit other water bodies	77

Based on 636 observations used in the model estimation after removing incomplete observations and protest bids

activities at all sites. Breskens is also popular for bathing and Saeftinghe attracts, as expected, many visitors who enjoy natural values, wildlife, peace and quiet.

Half of the respondents believe that current water quality is generally good throughout the catchment, but consider further improvement of water quality in the coming years important. Forty percent believes water quality is not good enough and should be improved further. Among the three sites, water quality is perceived best ('good') at Breskens by most respondents. Water quality at Saeftinghe and Braakman is, on average, perceived as 'moderate', but the number of people who feel informed enough to evaluate water quality at these sites is substantially lower than at Breskens.

The survey results point out that 77% of all respondents also visit other sites besides the three included in the CE. This underlines the importance of accounting for possible substitution effects. Respondents reported substitution behavior in terms of sites and activities. If water quality at their most frequented site decreased to such a low level that their most preferred activity would no longer be possible at that site, two-fifth of the respondents would go to another preferably nearby location. Thirty percent would continue going to their preferred site, but half of them would switch to another recreational activity at that site. One-fifth of the respondents would no longer engage in any water recreational activities.

5.2 Choice Model Results

Out of the 3900 choice occasions, the opt-out was chosen 1026 times. Of the three locations, Breskens was chosen most often in the DCE (36% of all choice occasions), followed by Braakman (24%) and Saeftinghe (22%). 127 Respondents (16%) chose the opt-out in each choice task, of which 66 (6%) were classified as protest votes. These respondents were excluded from the sample. Reasons for refusing to pay that comply with theoretical expectations were kept in the analysis. These reasons include that respondents felt that the proposed alternatives did not give value for money or they never visited the site(s) before and had no intention to do so in the future. The exclusion of respondents after data collection implies that the average number of times that each block was used was 27. The models were estimated in NLOGIT 4.0 using 1000 Halton draws.

Given their categorical nature, all attributes were dummy coded with current quality levels as the baseline level. For the price parameter linear coding was used. To explain the observed preference heterogeneity among respondents, the models include interactions of the ASC with theoretically expected variables for respondent characteristics such as household income and the distance respondents live from the sites. One-way distances are calculated based on the existing road network between a respondent's place of residence and the three locations.

Table 3 presents the results of the two models presented in Sect. 3. Only variables are included in the models that have a significant impact on choice behaviour at the 5% level. Model I is the standard mixed logit model, Model II is the mixed logit model extended with the cross-effects. In both models, the attribute parameters β_{ki} are significant at the 1% level and have the expected sign: quality improvements have a positive effect and price a negative effect. The ASCs in the model include the baseline levels of the three categorical attributes (walking, nature and bathing) interacted with the two respondent-characteristics. The estimated parameters for respondents' household income and distance, which are assumed to be the same across the three sites in this study, are also statistically significant and have the expected signs.¹ Income (in its logarithmic form) has a positive effect: the higher net household income, the higher WTP for one of the alternatives. The distance parameter has a negative sign, implying significant distance-decay.

Model I includes site-specific parameters for those attributes that are valued significantly differently between sites based on the LR test. The results show that the attribute coefficients related to good bathing water quality and nature levels are site-specific. Improving bathing water quality at Braakman to a good level (compared to the poor level in the status quo) yields a significantly higher coefficient estimate than improving bathing water quality at Breskens (compared to the moderate status quo level at this site), partially because the baseline level is lower at Braakman. Similarly, the parameter estimate for nature improvements to good levels at Saeftinghe is significantly higher than the parameter estimate for a comparable change at Braakman and Breskens, where nature improvements have a statistically similar value.

Unobserved preference heterogeneity is found for the highest quality levels for walking at all three sites, nature at Braakman and Breskens and bathing at Braakman. A normal distribution was used for the random variables.² No random effects were found for the other attributes at moderate or good levels that significantly improved the model fit according to the Likelihood Ratio test results. Consequently, we exclude the possibility that any significant cross-effects in Model II are the result of unaccounted remaining unobserved preference heterogeneity.

We added additional error-components to Model I for all possible combinations of the three sites. The error-components are significant at the 1% level for the combinations Breskens-Braakman and Braakman-Saeftinghe, but not for Breskens-Saeftinghe (and hence the latter error component is not included in the specification of Model II). Model

¹ When we include site-specific variables for distance and income in Model 2, the cross-effects remain the same. For simplicity reasons, we therefore included these variables as common across all sites.

² We estimated the same models using a uniform distribution because of the categorical nature of the variables (Hensher et al. 2005), where we restricted the spread of the distribution to be the same as the mean of the parameter to ensure that the parameter is positive over the entire distribution. The results are similar to random parameters with normal distributions: the same variables demonstrate significant preferences heterogeneity and in Model II the same cross-effects are significant.

Table 3 Estimated choice model results

Explanatory factor	Model I	Model II
<i>Alternative specific constants (ASC)</i>		
ASC Breskens	-8.322*** (3.142)	-8.348*** (3.170)
ASC Braakman	-9.681*** (3.653)	-9.736*** (3.694)
ASC Saeftinghe	-9.006*** (3.401)	-9.083** (3.438)
<i>Attributes-generic</i>		
Price	-0.023*** (19.245)	-0.024*** (18.321)
Walking—moderate quality	0.418*** (5.178)	0.418*** (5.170)
Walking—good quality (random parameter, normal distribution)	1.186*** (13.339)	1.186*** (13.434)
Standard deviation of the random parameter	0.851*** (6.735)	0.821*** (6.437)
Nature—moderate quality	0.460*** (5.618)	0.492*** (5.917)
<i>Attributes-site-specific</i>		
Nature—good quality (Breskens & Braakman) (random parameter, normal distribution)	1.012*** (10.155)	1.012*** (9.962)
Standard deviation of the random parameter	0.819*** (4.875)	0.837*** (5.085)
Nature—good quality (Saeftinghe)	1.486*** (11.697)	1.524*** (11.812)
Bathing—moderate quality (Braakman)	0.652*** (4.491)	0.646*** (4.437)
Bathing—good quality (Breskens) (random parameter, normal distribution)	0.681*** (6.361)	0.702*** (6.653)
Standard deviation of the random parameter	0.660*** (2.692)	0.573** (2.084)
Bathing—good quality (Braakman)	1.716*** (12.761)	1.855*** (13.167)
<i>Cross-effects</i>		
Saeftinghe * (moderate nature quality at Braakman)		0.314** (2.413)
Saeftinghe * (good bathing quality at Braakman)		0.420*** (3.604)
Saeftinghe * (price of Braakman)		-0.005** (2.048)
<i>Respondent characteristics</i>		
Net household income (logarithmic)	1.450*** (4.104)	1.456*** (4.142)
Distance (in km)	-0.023*** (7.614)	-0.023*** (7.557)
<i>Error-components</i>		
Sigma error-component I (Breskens, Braakman, Saeftinghe)	2.908*** (12.174)	2.921*** (12.330)

Table 3 (continued)

Explanatory factor	Model I	Model II
Sigma error-component 2 (Breskens, Braakman)	1.536*** (13.115)	1.497*** (12.766)
Sigma error-component 3 (Braakman, Saeftinghe)	1.234*** (11.767)	1.274*** (12.062)
<i>Model statistics</i>		
No. of observations	3180	3180
Loglikelihood	-3322	-3314

Explanatory notes: *t*-values are presented between brackets

Significance of the parameters is indicated by asterisks: *** at 1%, ** at 5%, * at 10%

II furthermore extends Model I through the inclusion of cross-effects. All possible combinations of attributes are tested in the probability function of the two other alternatives. In this estimation process, where we retained all cross-effects significant at the 10% level in a backward elimination procedure, three cross-effects remain consistently significant, all in the utility function for Saeftinghe.³ The inclusion of these three variables increases the model fit significantly compared to Model I (LR-test statistic = 16.0, $p < 0.05$).⁴

The first cross-effect reveals a disproportional substitution effect of changes in the nature attribute at Braakman. This cross-effect shows that improving nature to a moderate level at Braakman has a less negative effect on the probability of choosing Saeftinghe than on the probability of choosing Breskens. The parameter value (0.314) reflects the additional effect of a nature improvement at Braakman on the probability of selecting Saeftinghe (higher perceived utility), over and above the proportional effect captured by the main effect of moderate level of the nature attribute (0.492).

Probability share simulations based on Model II show that improving nature at Braakman to a moderate level would only draw choices away from Breskens (-4.3%) and would slightly increase Saeftinghe's probability share (+1.5%). Model I without this cross-effect predicts a higher, proportional probability loss for Saeftinghe for such a nature improvement scenario (-1.7%). In other words, when nature at Braakman improves from the current poor to the moderate level, the resulting substitution pattern shows stronger substitutability, as expected, between Breskens and Braakman compared to Saeftinghe. Respondents perceive Breskens as a closer substitute to Braakman, as for these sites access is easier, bathing is possible, and the type of nature amenities are more common, compared to Saeftinghe with its relatively unique nature and wildlife, and which is only accessible under supervision of a guide. However, the increase in the market share of Saeftinghe may suggest a violation of the regularity assumption.

A second cross-effect is found for good bathing water quality at Braakman. The positive coefficient in the utility function of Saeftinghe means that an improvement of bathing water quality at Braakman to a good level draws proportionally fewer choices away from Saeftinghe than from Breskens. Probability share simulations for Model II show that such

³ "Appendix" contains the results of Model II with all cross-effects.

⁴ We also checked whether interaction effects between the attributes within site-specific utility functions were significant, as our design allowed also for their estimation. However, we found no such interaction effects to be significant at the 10% level, while the three cross-effects remained consistently significant.

an improvement in bathing water quality at Braakman would result in only 3.3% fewer choices for Saeftinghe compared to 13.2% fewer choices for Breskens, compared to 6.8% and 10.8% fewer choices respectively according to Model I. This cross-effect is expected to reflect a shift in choices of respondents who are more interested in bathing water quality away from the beach Breskens to Braakman rather than from the natural area Saeftinghe, due to the absence of bathing possibilities at Saeftinghe. It implies again that a disproportional substitution pattern across the three sites can be observed when bathing possibilities at Braakman improve to a good level and suggests that Braakman and Breskens are also in this respect perceived to be close substitutes.

A third cross-effect is found for the price of improvements at Braakman on the probability of choosing Saeftinghe. The negative estimate (-0.005) implies that a price increase for Braakman reduces the probability that respondents are willing to pay for improvements at Saeftinghe more so than for Breskens. For example, if the price for Braakman *ceteris paribus* increases from €0 to €80 per household per year, Model I suggests that this increases Saeftinghe's probability share by 6.9%, while Model II adjusts this downwards to 2.9%. This cross-effect suggests that Saeftinghe and Braakman are weak substitutes in price too.

For the other variables in the models, the estimation results of Models I and II lead to similar conclusions. Overall, the comparison of Model I and II leads us to conclude that disproportional substitution patterns exist that cannot be captured by standard error-components and random parameters alone.

5.3 WTP Results

An important question is whether the cross-effects potentially result in significantly different implicit prices and welfare estimates for environmental improvements at the three sites. Here, the WTP values are estimated using the same approach as for mixed logit models, assuming that this approximation is correct. The estimates for the cross-effects are presented here to illustrate their order of magnitude if this assumption holds, and hence their relevance compared to discrete choice models in which they are not accounted for. Results are presented in Table 4, where the confidence intervals between brackets are calculated following Krinsky and Robb (1986). Tests as described in Poe et al. (2005) are used to test for differences between WTP estimates across sites and models. There are no significant differences in the WTP values for the direct effects of the individual attributes between the two models.

The top part of Table 4 presents the implicit prices of the cross-effects β_{kij} . The WTP values of the cross-effects are €13 and €18 per household per year. These values are the additional marginal WTP for Saeftinghe when Braakman is improved in the bathing quality and nature attributes. The cross-effects can be interpreted as an adjustment of the WTP value due to disproportional substitution between sites when multiple sites change at the same time. The cross-effects not only apply to these scenarios, but to any of the quality improvements, also for an improvement of a single attribute.

Mean household WTP for scenarios in which the conditions of attributes at the three sites change from the status quo to moderate and good quality levels are presented in the lower part of Table 4. We based the estimations on equation 5.17 in Bockstael and McConnell (2007, p. 112). For Model 1, the expected compensating surplus (CS) estimates reflect the WTP of the direct effects only, while for Model 2 the cross-effects are added in the CS estimation. First, we consider the mean WTP estimates for policy interventions that change the levels of attributes associated with the first cross-effects, i.e. an improvement of

Table 4 Mean WTP (€/household/year) for cross-effects and different environmental improvement scenarios

	Model I	Model II
<i>Cross-effects</i>		
Saefthinghe * moderate nature quality at Braakman	n.a.	13 (2–24)
Saefthinghe * good bathing quality at Braakman	n.a.	18 (6–29)
Saefthinghe * price of Braakman	n.a.	–0.2 (–0.4 to –0.01)
<i>Policy scenarios</i>		
(1) Braakman: nature moderate (including cross-effect for Model 2)	20 (13–27)	34** (21–48)
(2) Braakman and Saefthinghe: nature moderate (including cross-effect for Model 2)	40 (26–54)	55 (37–74)
(3) All sites: nature moderate (including cross-effect for Model 2)	60 (39–81)	76 (51–101)
(4) Braakman: all attributes moderate (including cross-effect for Model 2)	67 (51–84)	79 (61–99)
(5) Saefthinghe: all attributes moderate	38 (28–49)	39 (27–50)
(6) Breskens: all attributes moderate	38 (28–49)	39 (27–50)
(7) Braakman: bathing good (including cross-effect for Model 2)	75 (62–90)	96** (76–117)
(8) Braakman and Breskens: bathing good (including cross-effect for Model 2)	105 (88–123)	126* (102–150)
(9) Braakman: all attributes good (including cross-effect for Model 2)	171 (150–194)	189 (163–217)
(10) Saefthinghe: all attributes good	116 (100–135)	115 (97–133)
(11) Breskens: all attributes good	126 (110–144)	123 (106–140)

For scenarios 1–4, the cross-effect ‘Saefthinghe * moderate nature quality at Braakman’ is included. For scenarios 7–9, the cross-effect ‘Saefthinghe * good bathing quality at Braakman’ is included. In scenarios 5, 6, 10 and 11, the cross-effects are not relevant because there are no changes in the attributes of Braakman. For example, WTP for scenario 1 is estimated as $(\beta_{\text{Nature-moderate quality}} + \beta_{\text{Saefthinghe * moderate nature quality at Braakman}}) / \beta_{\text{Price}}$, and for scenario 2 as $(\beta_{\text{Nature-moderate quality}} + \beta_{\text{Nature-moderate quality}} + \beta_{\text{Saefthinghe * moderate nature quality at Braakman}}) / \beta_{\text{Price}}$; for scenario 2, the value of improvements of nature to moderate quality at both Braakman and Saefthinghe is included

The 95% confidence intervals are presented between brackets

n.a. not applicable

Statistical significance of the difference in WTP values between the two models is indicated by asterisks: ** at 5%, * at 10%

the nature attribute to a moderate quality level. In scenario 1, nature at Braakman improves to the moderate level, which leads to a significant difference in WTP at the 5% level between the two Models due to the cross-effect. The parameter estimates of Model I, which are based on proportional substitution between the sites and ignore the cross-effect, differ from Model II estimates by as much as 71%. When nature is also improved at the other

two sites (scenarios 2 and 3), or when bathing and walking conditions at Braakman also improve to a moderate level (scenario 4), the WTP estimates are no longer significantly different between the two models, even if the WTP differences are considerable and between 19 and 38%. However, as this cross-effect may lead to a violation of the regularity assumption (see Sect. 5.2), this result should be interpreted with caution when aiming to use the WTP value for further welfare analysis.

Next, we estimate WTP for improvements of bathing quality at Braakman to a good level (scenarios 7–9). This improvement at Braakman (scenario 7) leads to a difference in WTP between the two Models of 28%, which is significant at the 5% level, due to the cross-effect. Improving bathing quality at both Breskens and Braakman (scenario 8) also leads to a significant difference in WTP estimates at the 10% level. However, for improvements of all environmental attributes to a good level at Braakman (scenario 9), the difference in WTP is not statistically significant. This is in part due to the magnitude of the cross-effects relative to the attributes and the large confidence intervals around the random parameters of the improvements to good quality.

For completeness, we also present the mean WTP estimates for environmental improvement scenarios to moderate and good levels at Saeftinghe and Breskens (scenarios 5, 6, 10 and 11 respectively). This reveals that improvements of all attributes to moderate and good quality levels at Braakman are most beneficial to society, both with and without cross-effects, as these changes generate a significantly higher mean WTP at Braakman than at the two other sites. However, while mean WTP for improvements at Braakman remains highest, all mean household WTP estimates are distance dependent and residents near Breskens or Saeftinghe may prefer the sites closest to them to be improved. For improvements to good quality levels, WTP estimates for Breskens and Saeftinghe are not significantly different, despite the absence of bathing amenities at the latter site (scenarios 10 and 11). This seems to be compensated by the higher WTP for nature improvements at Saeftinghe.

Overall, the cross-effects may lead to significantly different WTP estimates. In the case study presented here, mean WTP for the scenarios of moderate quality changes at Braakman is lower if cross-effects on Saeftinghe are not accounted for. These results suggest that for any set of alternatives where the underlying substitution pattern is disproportional and some alternatives are closer substitutes than others, disproportional substitution patterns may affect WTP values. However, the differences are only significant for a limited set of policy scenarios, and the discussion on the validity of these welfare estimate due to inconsistency with RUM needs to be considered.

6 Conclusions

The specification of substitution effects between environmental goods and services provided at different locations in a confined geographical area is often neglected in both the design and modelling of stated preferences in site selection DCEs. This paper addresses spatial substitution patterns for environmental quality improvements across different sites in the same area using a labelled site choice DCE. The results of the improvements at these sites lead to different benefits, suggesting that they are imperfect substitutes. A novel modelling framework was developed with the aim to estimate possible disproportional substitution patterns among alternatives in the respondents' choice set, combining the advantages of mixed and universal logit models. The study shows that accounting for more complex substitution patterns across sites improves our model fit and leads to different probability shares, resulting

from changes in the characteristics of these alternative sites. Besides common random terms and error-components, cross-effects are included in the site-specific utility functions to identify additional disproportional substitution patterns across the three sites. Cross-effects reflect disproportional changes in the utility attached to an alternative based on changes in the characteristics of its substitutes. Cross-effects may occur when options vary in similarity, but such similarity and substitutability are ultimately based on perceptions of responsiveness and not on scientifically defined functionality, neither in terms of ecological functionality nor in terms of ecosystem benefits. A cautious approach would therefore be to check whether cross-effects are present if the experimental design allows for that. The model proposed in this paper is relevant to any case study which compares multiple sites in a confined geographical area that are considered as potential, but imperfect substitutes.

In this paper, possible cross-effects are estimated in a mixed logit model and we demonstrate that this results in a modest but significant improvement in model fit. Despite the inclusion of error-components as suggested by Brownstone and Train (1999), and random parameters to allow for flexible substitution patterns between alternatives, the cross-effects are found to be significant in the extended mixed logit model. This leads us to conclude that when using standard mixed logit models, some part of the complexity of the substitution effects caused by differences in the similarity of the alternatives in choice behaviour remains unaccounted for. The results from our study point out the necessity to pay adequate attention in spatial choice studies to the complexity of possible substitution patterns, as also explored in recent work by Hess et al. (2017) and Hess and Train (2017).

In the absence of clear guidance in the discrete choice literature regarding their consistency with RUT, we estimated the WTP estimates for the cross-effects based on the extended model in the same way as for standard mixed logit models. Although these estimates have to be interpreted with the necessary caution, accounting for these additional cross-effects produces significantly different welfare estimates for environmental improvement scenarios of up to 70% in our case study. This may question the validity and reliability of existing welfare estimates in cases where environmental policy targets multiple sites, not only from traditional single-site studies in which such substitution effects are ignored, but also more recent multiple site DCEs (e.g. Lizin et al. 2016; Logar and Brouwer 2018). Unlabelled DCEs do not allow for the estimation of cross-effects between sites, so such designs are limited in the extent to which substitution can be accounted for anyway. However, also for multiple site studies elicited in labelled DCEs, our results suggest that adding up the values found for single sites to estimate the welfare effects of a policy that affects preferences for multiple sites may not be reliable.

The main problem of cross-effects is that in some cases they can lead to a violation of the regularity assumption. We tested this here empirically through market share simulations, which confirmed disproportional substitution patterns. More guidance is needed as to how *actual* rather than *potential* violation of the regularity assumption can be detected and affects the validity of cross-effects for economic welfare estimation (Hess et al. 2018). Moreover, more empirical applications are needed to further evaluate the impact of disproportional substitution patterns on WTP estimates. At this point, whether to apply UL type models or standard mixed logit models may be a choice between describing choice behaviour as well as possible or having confidence that welfare estimates are RUM-consistent.

In this study we used a limited number of alternatives and attributes. The identification of significant cross-effects may become more complicated as the size of the choice set increases as well as the number of attributes (for a choice set consisting of J alternatives and K attributes, there are $(J-1)*K$ cross-effects when using continuous attribute levels), for instance as in multi-site travel cost studies. A related challenge is the large

experimental design needed to estimate cross-effects under such circumstances, which may cause practical implementation problems related to sample size and survey length. A practical shortcoming of using stated DCEs in site-selection studies is that only a limited number of alternatives can be included in the design, not least to keep the cognitive burden to a minimum, which can be less than the number of relevant alternatives in an area. In our study, we limited the choice set to three sites and we did not find a significant effect of alternative sites visited by respondents outside the choice set on choice behaviour for these three sites. However, this result has to be interpreted with the necessary care as we have no information about the impact of changes at these alternative sites outside the choice set on demand for the quality improvements at the sites in the CE and vice versa.

Options for future work include the comparison of our modelling approach to models that account for attribute or alternative non-attendance or other behavioural ‘anomalies’ (Hensher et al. 2005b; Scarpa et al. 2009; Hess et al. 2013; Campbell et al. 2014; Hess et al. 2018). Another possible comparison would be against the modelling approach suggested by Hess and Train (2017) and operationalised in Hess et al. (2017). A major difference between these models and our model is that they, unlike our model presented here, ultimately retain the IPS assumption at individual level.

Finally, in terms of policy implications, our results show that the non-market benefits of environmental quality improvements of a WFD water body are site-specific and depend on changes in and around other water bodies located in the same catchment. This complicates the practice of benefits transfer based on generic attribute values which presumes similar values for environmental quality objectives across sites and points at a potential limitation of unlabelled DCEs, which cannot assess the effect of context, such as the site-specific manifestation of the ecological policy objectives as articulated in the WFD. In this case study, improvements in nature and bathing water quality conditions are valued significantly different across sites, indicating that ecosystem services provided at one site are not perfect substitutes for the same type of services provided elsewhere. Furthermore, we argue that cross-effects are especially relevant when water agencies and local governments could opt to improve multiple sites, cost permitting. The cross-effects found in this study suggest that synergy effects in water management investments can be achieved if both Braakman and Saeftinghe are improved, when comparing the WTP results of Model I (without cross-effects) with Model II (with cross-effects). Gaining insight in these different site-specific values and substitution effects is expected to help policymakers prioritize the allocation of their limited budgets based on welfare maximization principles.

Moreover, the substitution effects in this study show that even if sites are physically separated, they may still be perceived by users and non-users as (imperfect) substitutes in economic terms. Investments in the improvement of one site may have implications for the welfare associated with substitute sites. Such substitute effects may be site-specific. Policy and decision-makers who base their decision on information from SP studies should therefore be careful and ensure that they use studies that have accounted for such substitution effects.

Acknowledgements This study was carried out and financed under the EU DG Research F6 Project Aqua-Money (SSPI-022723).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Table 5 presents the extended mixed logit model with the full set of cross-effects alongside with Models 1 and Model 2, all without the income and distance parameters. We also estimated a model with those parameters and including them does not change the results. The cross-effects that remain significant in Model II after backward elimination are also consistently significant in this full model.

Table 5 Model estimates of Model 1, Model 2 and a Model with all cross-effects (without socio-demographic variables)

Explanatory factor	Model I	Model II	Full cross-effects model
<i>Alternative specific constants (ASC)</i>			
ASC Breskens	−8.322*** (3.142)	−8.348*** (3.170)	1.413*** (3.024)
ASC Braakman	−9.681*** (3.653)	−9.736*** (3.694)	0.343 (0.724)
ASC Saeftinghe	−9.006*** (3.401)	−9.083** (3.438)	0.634 (1.377)
<i>Attributes-generic</i>			
Price	−0.023*** (19.245)	−0.024*** (18.321)	−0.275*** (13.227)
Walking—moderate quality	0.418*** (5.178)	0.418*** (5.170)	0.324** (2.396)
Walking—good quality (random parameter, normal distribution)	1.186*** (13.339)	1.186*** (13.434)	1.265*** (9.703)
Standard deviation of the random parameter	0.851*** (6.735)	0.821*** (6.437)	0.793*** (6.017)
Nature—moderate quality	0.460*** (5.618)	0.492*** (5.917)	0.583*** (4.826)
<i>Attributes-site-specific</i>			
Nature—good quality (Breskens & Braakman) (random parameter, normal distribution)	1.012*** (10.155)	1.012*** (9.962)	1.308*** (7.868)
Standard deviation of the random parameter	0.819*** (4.875)	0.837*** (5.085)	0.900*** (5.261)
Nature—good quality (Saeftinghe)	1.486*** (11.697)	1.524*** (11.812)	1.577*** (6.854)
Bathing—moderate quality (Braakman)	0.652*** (4.491)	0.646*** (4.437)	0.553** (2.264)
Bathing—good quality (Breskens) (random parameter, normal distribution)	0.681*** (6.361)	0.702*** (6.653)	0.819*** (4.015)
Standard deviation of the random parameter	0.660*** (2.692)	0.573** (2.084)	0.610** (2.249)
Bathing—good quality (Braakman)	1.716*** (12.761)	1.855*** (13.167)	2.108*** (9.271)
<i>Respondent characteristics</i>			
Net household income (logarithmic)	1.450*** (4.104)	1.456*** (4.142)	

Table 5 (continued)

Explanatory factor	Model I	Model II	Full cross-effects model
Distance (in km)	-0.023*** (7.614)	-0.023*** (7.557)	
<i>Error-components</i>			
Sigma error-component 1 (Breskens, Braakman, Saeftinghe)	2.908*** (12.174)	2.921*** (12.330)	3.384*** (12.996)
Sigma error-component 2 (Breskens, Braakman)	1.536*** (13.115)	1.497*** (12.766)	1.705*** (13.553)
Sigma error-component 3 (Braakman, Saeftinghe)	1.234*** (11.767)	1.274*** (12.062)	1.254*** (11.358)
<i>Cross-effects of final Model II</i>			
Saeftinghe * (moderate nature quality at Braakman)		0.314** (2.413)	0.693*** (2.742)
Saeftinghe * (good bathing quality at Braakman)		0.420*** (3.604)	0.371** (2.026)
Saeftinghe * (price of Braakman)		-0.005** (2.048)	-0.043*** (2.667)
<i>All other cross-effects</i>			
Breskens * (moderate nature quality at Braakman)			-0.177 (1.056)
Breskens * (good nature quality at Braakman)			0.316* (1.743)
Breskens * (moderate bathing quality at Braakman)			-0.118 (0.530)
Breskens * (good bathing quality at Braakman)			0.377* (1.704)
Breskens * (moderate walking quality at Braakman)			-0.175 (1.020)
Breskens * (good bathing quality at Braakman)			0.088 (0.516)
Breskens * (price at Braakman)			-0.24* (0.727)
Breskens * (moderate nature quality at Saeftinghe)			0.490*** (2.697)
Breskens * (good nature quality at Saeftinghe)			0.354 (1.547)
Breskens * (moderate walking quality at Saeftinghe)			0.149 (0.723)
Breskens * (good walking quality at Saeftinghe)			0.381** (1.991)
Breskens * (price at Saeftinghe)			-0.17 (1.246)
Braakman * (moderate nature quality at Breskens)			0.143 (0.745)
Braakman * (good nature quality at Breskens)			0.312 (1.566)
Braakman * (good bathing quality at Breskens)			0.189 (0.863)

Table 5 (continued)

Explanatory factor	Model I	Model II	Full cross-effects model
Braakman * (moderate walking quality at Breskens)			-0.143 (0.745)
Braakman * (good walking at Breskens)			0.065 (0.357)
Braakman * (price at Breskens)			-0.035** (2.349)
Braakman * (moderate nature quality at Saeftinghe)			0.469** (2.296)
Braakman * (good nature quality at Saeftinghe)			0.259 (1.020)
Braakman * (moderate walking quality at Saeftinghe)			-0.024 (0.188)
Braakman * (good bathing quality at Saeftinghe)			0.214 (1.066)
Braakman * (price at Saeftinghe)			-0.021 (1.422)
Saeftinghe * (moderate nature quality at Breskens)			0.164 (0.866)
Saeftinghe * (good nature quality at Breskens)			0.376* (1.793)
Saeftinghe * (good bathing quality at Breskens)			0.137 (0.658)
Saeftinghe * (moderate walking quality at Breskens)			-0.053 (0.268)
Saeftinghe * (good walking at Breskens)			0.044 (0.224)
Saeftinghe * (price at Breskens)			-0.015 (0.973)
Saeftinghe * (good nature quality at Braakman)			0.334 (1.518)
Saeftinghe * (moderate bathing quality at Braakman)			-0.086 (0.351)
Saeftinghe * (moderate walking quality at Braakman)			0.023 (0.116)
Saeftinghe * (good bathing quality at Braakman)			0.247 (1.180)
<i>Model statistics</i>			
No. of observations	3180	3180	3180
Loglikelihood	-3322	-3314	-3332

References

- Agarwal MK (2002) Asymmetric price effects in the telecommunications services markets. *J Bus Res* 55:671–677
- Arrow K, Solow R, Portney PR et al (1993) Report of the NOAA Panel on contingent valuation. *Fed Reg* 58:4601–4614
- Bateman IJ, Carson RT, Day B et al (2002) *Economic valuation with stated preference techniques: a manual*. Edward Elgar, Cheltenham

- Bateman JJ, Lovett AA, Brainard JS (2003) Applied environmental economics—a GIS approach to cost-benefit analysis. Cambridge University Press, Cambridge
- Ben-Akiva M (1974) Note on the specification of a logit model with utility functions that include attributes of competing alternatives. Department of Civil Engineering, MIT, Cambridge
- Block HD, Marschak J (1960) Random orderings and stochastic theories of responses. In: Marschak J (ed) Economic information, decision and prediction: selected essays (volume 1). D. Reidel, Dordrecht
- Bockstael N, McConnell K (2007) Environmental and resource valuation with revealed preferences a theoretical guide to empirical models. Springer, Berlin
- Bos DM, Molin EJE (2006) Is there a 'stick' bonus? A stated choice model for P&R patronage incorporating cross-effects. *Eur J Transp Infrastruct Res* 6:275–290
- Brouwer R (2008) The potential role of stated preference methods in the Water Framework Directive to assess disproportionate costs. *J Environ Planning Manag* 51:597–614
- Brouwer R, Martín-Ortega J, Berbel J (2010) Spatial preference heterogeneity: a choice experiment. *Land Econ* 86:552–568
- Brownstone D, Train KE (1999) Forecasting new product penetration with flexible substitution patterns. *J Econom* 89(1-2):109–129
- Campbell D, Hensher DA, Scarpa R (2014) Bounding WTP distributions to reflect the 'actual' consideration set. *J Choice Model* 11:4–15
- Carson RT, Louviere JJ, Anderson DA et al (1994) Experimental analysis of choice. *Mark Lett* 5:351–367
- Carson R, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence. *Environ Resource Econ* 19:173–210
- Chowdhury S, Meenakshi JV, Tomlins KI, Owori C (2011) Are consumers in developing countries willing to pay more for micronutrient-dense biofortified foods? Evidence from a field experiment in Uganda. *Am J Agric Econ* 93:83–97
- Crouch GI, Oppewal H, Huybers T et al (2007) Discretionary expenditure and tourism consumption: insights from a choice experiment. *J Travel Res* 45:247–258
- De Valck J, Broekx S, Liekens I et al (2017) Testing the influence of substitute sites in nature valuation by using spatial discounting factors. *Environ Resource Econ* 66:17–43
- Hahn J, Hausman J, Lustig J (2007) Specification test on mixed logit models. Centre for Microdata Methods and Practice, Institute for Fiscal Studies, No. CWP58/17
- Hailu A, Adamowicz WL, Boxall PC (2000) Complements, substitutes, budget constraints and valuation. *Environ Resource Econ* 16:51–68
- Hanley N, Wright RE, Alvarez-Farizo B (2006) Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive. *J Environ Manag* 78:183–193
- Hensher DA, Rose JM, Greene WH (2005a) Applied choice analysis: a primer. Cambridge University Press, Cambridge
- Hensher DA, Rose J, Greene WH (2005b) The implications on willingness to pay of respondents ignoring specific attributes. *Transportation* 32:203–222
- Herriges JA, Phaneuf DJ (2002) Inducing patterns of correlation and substitution in repeated logit models of recreation demand. *Am J Agric Econ* 84:1076–1090
- Hess S, Train K (2017) Correlation and scale in mixed logit models. *J Choice Model* 23:1–8
- Hess S, Stathopoulos A, Campbell D et al (2013) It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation* 40:583–607
- Hess S, Murphy P, Le H, Leong WY (2017) Estimation of new monetary valuations of travel time, quality of travel, and safety for Singapore. *Transp Res Rec J Transp Res Board* 2664:79–90
- Hess S, Daly A, Batley R (2018) Revisiting consistency with random utility maximisation: theory and implications for practical work. *Theor Decis* 1:181–204
- Hoehn JP, Loomis JB (1993) Substitution effects in the valuation of multiple environment programs. *J Environ Econ Manag* 25:56–75
- Hole AR, Kolstad JR (2012) Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empir Econ* 42:445–469
- Huber J, Payne JW, Puto C (1982) Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *J Consum Res* 9:90–98
- Kanninen BJ (2007) Valuing environmental amenities using stated choice studies; a common sense approach to theory and practice. Springer, Dordrecht
- Kaoru Y (1995) Measuring marine recreation benefits of water-quality improvements by the nested random utility model. *Resour Energy Econ* 17:119–136

- Kotchen MJ, Reiling SD (1999) Do reminders of substitutes and budget constraints influence contingent valuation estimates? Another comment. *Land Econ* 75:478–482
- Krinsky I, Robb AL (1986) On approximating the statistical properties of elasticities. *Rev Econ Stat* 68:715–719
- Kuhfeld WF (2010) In: Discrete choice. SAS Institute. http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html (Marketing Research Procedures)
- Lipovetsky S, Conklin M (2014) Finding items cannibalization and synergy by BWS data. *J Choice Model* 12:1–9
- Lizin S, Brouwer R, Liekens L, Broeckx S (2016) Accounting for substitution and spatial heterogeneity in a labelled choice experiment. *J Environ Manag* 181:289–297
- Logar I, Brouwer R (2018) Substitution effects and spatial preference heterogeneity in single- and multiple-site choice experiments. *Land Econ* 94:302–322
- Loomis J, Gonzalez-Caban A, Gregory R (1994) Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Econ* 70:499
- Louviere JJ, Hensher DA, Swait J (2000) Stated choice methods, analysis and applications. Cambridge University Press, Cambridge
- McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, New York
- McFadden D (1975) On interdependence, structure and simultaneity in transportation demand analysis. Institute of Transport Studies, University of California, Berkeley. Working paper
- McFadden D (2000) Disaggregate behavioral travel demand's RUM side. *Travel behaviour research*, pp 17–63
- McFadden D, Train K (2000) Mixed MNL models of discrete response. *J Appl Econom* 15:447–470
- Needelman MS, Kealy MJ (1995) Recreational swimming benefits of New Hampshire Lake Water Quality Policies: an application of a repeated discrete choice model. *Agric Resour Econ Rev* 24:78–87
- Nielsen ASE, Lundhede TH, Jacobsen JB (2016) Local consequences of national policies—a spatial analysis of preferences for forest access reduction. *For Policy Econ* 73:68–77
- Oppewal H, Holyoake B (2004) Bundling and retail agglomeration effects on shopping behavior. *J Retail Consum Serv* 11:61–74
- Oppewal H, Timmermans H (1991) Context effects and decompositional choice modeling. *Pap Reg Sci* 70:113–131
- Parsons GR (2003) The travel cost model. In: Champ PA, Boyle KJ, Brown TC (eds) *A primer on non-market valuation*. Kluwer, Dordrecht, pp 269–330
- Parsons GR, Massey DM (2003) A RUM model of beach recreation. In: Hanley N et al (eds) *The new economics of outdoor recreation*. Edward Elgar, Cheltenham
- Pellegrini PA, Fotheringham AS (2002) Modelling spatial choice: a review and synthesis in a migration context. *Prog Hum Geogr* 26:487–510
- Poe GL, Giraud KL, Loomis JB (2005) Computational methods for measuring the difference of empirical distributions. *Am J Agric Econ* 87:353–365
- Rolfe J, Bennett J, Louviere JJ (2002) Stated values and reminders of substitute goods: testing for framing effects with choice modelling. *Austral J Agric Resour Econ* 46:1–20
- Rooderkerk RP, Van Heerde HJ, Bijmolt THA (2011) Incorporating context effects into a choice model. *J Mark Res* 48:767–780
- Sawtooth Software (2008) Inc. Sequim, WA. <http://www.sawtoothsoftware.com/>
- Scarpa R, Ferrini S, Willis K (2005) Performance of error component models for status-quo effects in choice experiments. In: Scarpa R, Alberini A (eds) *Applications of simulation methods in environmental and resource economics*. Springer, Berlin, pp 247–273
- Scarpa R, Gilbride TJ, Campbell D, Hensher DA (2009) Modelling attribute non-attendance in choice experiments for rural landscape valuation. *Eur Rev Agric Econ* 36:151–174
- Schaafsma M, Brouwer R (2013) Testing geographical framing and substitution effects in spatial choice experiments. *J Choice Model* 8:32–48
- Schaafsma M, Brouwer R, Rose J (2012) Directional heterogeneity in WTP models for environmental valuation. *Ecol Econ* 79:21–31
- Schaafsma M, Brouwer R, Gilbert A et al (2013) Estimation of distance-decay functions to account for substitution and spatial heterogeneity in stated preference research. *Land Econ* 89:514–537
- Steenburgh TJ (2008) The invariant proportion of substitution property (IPS) of discrete-choice models. *Mark Sci* 27:300–307
- Steenburgh TJ, Ainslie A (2008) Taste heterogeneity, IIA, and the similarity critique. No. 09-049. Harvard Business School

- Sutherland RJ, Walsh RG (1985) Effect of distance on the preservation value of water quality. *Land Econ* 61:281–291
- Swait J, Louviere JJ (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res* 30:305–314
- Termansen M, Zandersen M, McClean CJ (2008) Spatial substitution patterns in forest recreation. *Reg Sci Urban Econ* 38:81–97
- Timmermans HJP, Molin EJE (2009) Dealing with increased complexity in conjoint experiments: background and overview of alternate approaches. *Transp Rev* 29:557–567
- Timmermans H, Borgers A, van der Waerden P (1991) Mother logit analysis of substitution effects in consumer shopping destination choice. *J Bus Res* 23:311–323
- Train KE (2002) *Discrete choice methods with simulation*. Cambridge University Press, Cambridge
- Tversky A (1972) Elimination by aspects: a theory of choice. *Psychol Rev* 79:281–299
- Walker JL, Ben-Akiva M, Bolduc D (2007) Identification of parameters in normal error component logit-mixture (NECLM) models. *J Appl Econom* 22:1095–1125
- Walker JL, Wang Y, Thorhauge M, Ben-Akiva M (2018) D-efficient or deficient? A robustness analysis of stated choice experimental designs. *Theor Decis* 1:215–238
- Whitehead JC, Blomquist GC (1999) Do reminders of substitutes and budget constraints influence contingent valuation estimates? Reply to another comment. *Land Econ* 75:483–484

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.