The Effect of Forced Choice with Constant Choice Experiment Complexity

Jerrod Penn, Graduate Student, University of Kentucky
Wuyang Hu, Professor, University of Kentucky
Linda J. Cox, Professor, University of Hawaii at Manoa
Jerrod.penn@uky.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014.

In this analysis we compare WTP estimates of responses separately modelling choice sets of forced choices from those of unforced choices in order to measure the effect of forced choice in Choice experiment. This comparison is done while still maintain constant task complexity and evaluated in WTP space. We find evidence to suggest that individual WTP are different in forced and unforced choice sets and in joint tests for parameter equality.

Introduction

Continuous examination and refinement of Choice experiments (CE) have made it a staple tool of nonmarket valuation. Researchers must make a number of decisions to implement a CE, and each affects welfare estimates. One scrutinized choice is whether to make respondents' choices forced or unforced, by including what is often called an "opt-out," "status-quo," or "Choose None" alternative (hereafter, referred to as "opt-out"). Without an opt-out alternative, welfare measurements are inconsistent with demand theory (Hanley, Mourato, & Wright, 2001) and produce biased results.

Considerable work has focused on the opt-out's form (Banzhaf, Johnson, & Mathews, 2001; Kontoleon & Yabe, 2003) and its heterogeneous welfare effects on various segments of people (Pedersen, Kjær, Kragstrup, & Gyrd-Hansen, 2011, 2012), but few specifically considered the effect of including or excluding an opt-out. Dhar and Simonson (2003) show significant differences among consumer products. Z. Kallas and Gil (2012) used a two-step valuation where the respondent made a forced choice in the first step and could opt-out in the second step. They formulated a forced model that ignored the respondent's answer in the second step compared to the same model except the second step opt-out answer was included as an extra alternative in the first step, making it an unforced model. Their results showed some significant differences in forced and unforced WTP for rabbit meat. Both of these studies are for private goods. The most apparent study we found comparing forced and unforced choices was by Carlsson, Frykblom, and Lagerkvist (2007) considering animal welfare in, a quasi-public good. In a mixed logit model, they find that including an opt-out does not have a significant effect on marginal WTP, but significantly affects the number of significant attribute standard deviations.

In Carlsson et al.'s (2007) case, a limitation exists to characterize the opt-out effect. To study opt-out, the authors used a split sample, some respondents were given an opt-out alternative and others were not. The difficulty is that introducing an additional alternative increases the choice complexity.

The purpose of our study is to study is to re-evaluate the effect of the opt-out alternative while holding constant the cognitive complexity in the CE, for the quasi-public good of recreational beach quality in Oahu, Hawaii. The paper proceeds with a short literature review on opt-out and choice experiment complexity effects. We then describe the application and data used, followed by statistical methods. The remainder of the paper presents model results and closes with conclusions and implications.

Literature Review

Opt-out alternatives have become fairly standard in DCE and are generally recommended (Adamowicz & Boxall, 2001; Hensher, 2010), specifically since opt-outs reflect reality (Carson et al., 1994). Z. Kallas and Gil (2012) provide a comprehensive overview of the theory and implications of including (and types) an opt-out alternative and (Boxall, Adamowicz, & Moon, 2009) give a review of potential bias from including an opt-out alternative. Yet recent studies without opt-out alternatives exist (Hasund, Kataria, & Lagerkvist, 2011; Zein Kallas, Escobar, & Gil, 2012; Rigby, Alcon, & Burton, 2010).

Most work has focused on how the opt-out alternative is described such as a "Choose None" or "Status Quo" option or on the respondent's perception of the opt-out alternative across a number of fields such as environment (Banzhaf et al., 2001; Marsh, Mkwara, & Scarpa, 2011; Scarpa, Willis, & Acutt, 2007), food (Kontoleon & Yabe, 2003) and transportation (Hensher, 2010). Some have studied the opt-out alternative's effect on different population segments (De Blaeij, Nunes, & Van den Bergh, 2007; Pedersen et al., 2011; Vujicic, Ryan, & Alfano). Ladenburg and Olsen (2010) used frequent opt-out reminders if the alternatives were too expensive, which reduced total WTP, but not attribute-specific WTP. Lastly, Rose and Hess (2009) studied the effect of a respondent-constructed reference alternative.

In a broader context, the decision to include or what type of opt-out alternative is part of a larger framework of designing an effective choice experiment. Implementation of a meaningful DCE requires a number of decisions, which can affect the choices and consistency of choices among respondents. Researchers have found substantial effects of design differences for multiple factors.

Results change if a respondent considers a different numbers of alternatives or attributes under in a scenario (DeShazo & Fermo, 2002; Rolfe & Bennett, 2009). Day et al. (2012) found evidence of order effects in which the same CE scenarios presented in different orders can have meaningful effects. Similarly, others have found that the number of choice tasks to complete affects results from learning or fatigue (Savage & Waldman, 2008). Carlsson and Martinsson (2001) showed significantly different preferences comparing choices in the first and second half of scenarios, evidence of a learning effect. Finally, Boxall et al. (2009) show that CE complexity affects the opt-out alternative and the subsequent welfare implications.

Ultimately, these studies show that CE design affects results. While the presence of opt-out alternatives is a straightforward design aspect, comparing its effects exclusively its exclusive effect has gone unnoticed. As such it is pertinent to consider the effect of opt-out while holding constant the changing choice task complexity.

Application and Methods

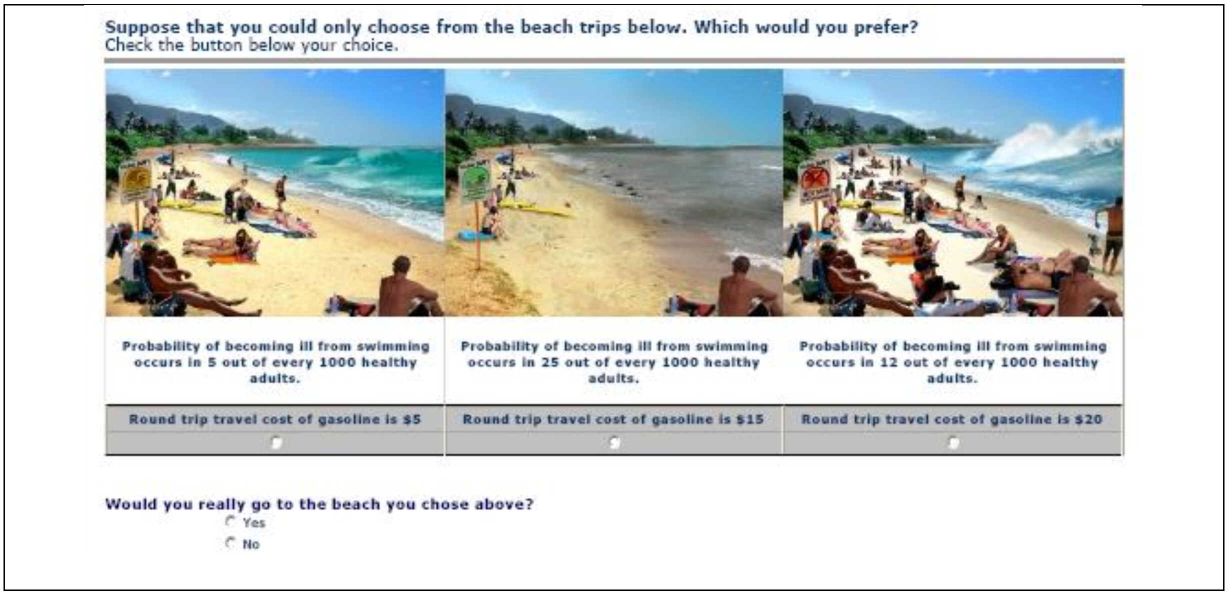
We examine the opt-out effect within the context of a valuation of recreational beaches in Oahu, Hawaii among tourists and residents. Beach attributes and levels were selected and developed primarily to understand recreational beach choice for swimming and wading based on a number of sources including correspondence with the Clean Water Branch within the Hawaii Department of Health, the city and county of Honolulu Ocean Safety and Lifeguard Services, previous scholarly work (Mak & Moncur, 1998; Mourato, Georgiou, Ozdemiroglu, Newcombe, & Howarth, 2003; Murray, Sohngen, & Pendleton, 2001; Oh, Dixon, & Draper, 2006), and focus group feedback.

The attributes of the beach valuation choice experiment are water quality (4 levels), sand quality (4 levels), congestion (3 levels), safety (3 levels), and round trip fuel costs (5 levels); described in Table 1.

Table 1: Choice Experiment Beach Attributes and Levels	
Attribute	**Level**
Sand Quality	**Excellent** –White; all sand **Good** – Light tan; composed of 75% sand and 25% foreign materials **Average** – Dark tan/light brown; composed of 50% sand and 50% foreign materials **Poor** – Brown/gray; composed of 75% foreign materials and 25% sand
Water Quality	**Excellent** – Clear, aqua colored water; probability of illness from wading occurs in 5 out of every 1000 healthy adults **Good** – Visible particles floating in otherwise clear water, blue in color; probability of illness from wading occurs in 12 out of every 1000 healthy adults **Average** – Cloudier water affecting visibility, green in color probability of illness from wading occurs in 19 out of every 1000 healthy adults **Poor** –Murky water, brownish in color; probability of illness from wading occurs in 25 out of every 1000 healthy adults
Water entry/ swimming safety	**Not Safe** – Conditions safe for the majority of beach recreationists **Safe** – Conditions safe for experienced beach recreationists **Very Safe** – Conditions not safe for any recreationists
Congestion	**Good** – Ample open space, and little noise **Average** – Beach congestion and noise are present, but do not hamper the experience **Below Average** – Overcrowded and extremely noisy
Roundtrip fuel cost	$0, $5, $10, $15, $20

In order to maintain task complexity, the choice experiment used the Dual Response method formulated by Brazell et al. (2006). In this study, dual response is somewhat like a follow-up certainty question in that immediately after a respondent provides their preferred choice in a forced setting, they immediately respond on if they would actually go to their previous forced choice, allowing the opportunity to opt-out. The technique has been used previously by Z. Kallas and Gil (2012), Kallas Escobar and Gill (2012), and Mentzakis, Ryan, and McNamee (2011). An example CE scenario is provided in Figure 1.

Figure 1: Example Choice Experiment Scenario with dual response



As seen above, the CE communicated attribute levels via computer-augmented pictures, except for the risk of illness and round trip cost of fuel, which were presented as text. Each respondent completed ten scenarios within the CE, and each scenario contained three alternatives plus the follow-up question. The survey was fielded in five locations around Oahu by a professional survey firm from late September to mid-October 2009, requiring that respondents be at least 18 years old and a United States citizen[1].

In the CE, every respondent initially made a forced-choice among three alternatives per scenario. The respondent could then opt-out after making the forced choice by saying “No” in the dual response question. “No” in the dual response question represents an opt-out decision, so their answer in the first step represents a *real* forced choice. For those who choose “Yes” in the dual response, it means the beach they selected in the first step represents an unforced choice. Therefore by utilizing the dual response question in the CE, every respondent must answer two questions per scenario such that some are making forced choices and others are making unforced choices, while maintaining choice complexity.

[1] Many of Hawaii’s tourists are from abroad so welfare results are really only representative of US WTP.

With this CE formulation, we can test the potential difference between forced versus non-forced choices but maintain the same choice set complexity; the complexity and the force choice effects are not confounded. To implement the test, we partition the sample scenario responses into two sets: scenarios in which the respondent opted out (said “No”) in the dual, i.e. forced choices, and the second set of scenarios in which the respondent did not opt out (said “Yes”) in the second step (unforced in the first step). This segmentation of the data allows us to test for potential effects of including an opt-out through Random Utility Theory (RUT) that discrete choice experiments rely upon (McFadden, 1973) as in equation (1).

$$\frac{U_{ijt}}{\mu} = \frac{\alpha * C_{ijt}}{\mu} + \frac{\beta * X_{ijt}}{\mu} + \frac{e_{ijt}}{\mu} \tag{1}$$

RUT asserts that the utility for alternative *j* by person *i* in scenario *t* is composed of a systematic component, C and X, which represent the payment vehicle and the non-payment attributes, respectively, and an unobserved component, *e*. By assuming the difference of the error between two alternatives is distributed Extreme Value, Equation (2) can be modelled as a conditional (i.e. multinomial) logit model. With the distributional assumption on *e*, the model’s parameters and standard errors are evaluated using maximum likelihood estimation. Because of the Extreme Value distribution, the random component must be scaled by μ, not directly affecting the estimation of α and β, but prohibits comparison of the parameters across models since multiple models are scaled by different μ.

In this study, since the attributes are somewhat qualitative, we convert each attribute’s levels into separate dummies to measure differences in magnitude between each level. For example, there are three indicators for excellent, good and poor water quality, and average water quality is the omitted reference category. Each indicator variable is included in the model of utility in preference space shown in equation (2).

$$\begin{aligned} U_{ijt} = {} & \alpha Cost_{njt} + \beta_1 PoorW + \beta_2 GoodW + \beta_3 ExcellentW + \beta_4 PoorS + \beta_5 GoodS \\ & + \beta_5 ExcellentS + \beta_6 BadC + \beta_7 GoodC + \beta_8 NoteSafe + \beta_9 VerySafe + e_{ijt} \end{aligned} \tag{2}$$

Equation (2) represents standard practice of modelling DCE's in preference space, in which the researcher makes distributional assumptions about parameters estimates, and then subsequently derive WTP estimates by dividing by the negative of α, the payment vehicle (the marginal utility of income).

We instead consider the model in Willingness to Pay (WTP) space. Greene and Hensher (2010) show that with corresponding assumptions, WTP space is a specified form of the Generalized Multinomial Logit model formulated by Fiebig, Keane, Louviere, and Wasi (2010). WTP space models are advantageous versus preference space, especially in mixed logit models, where distributional assumptions in parameter space have led to unrealistic values and distributions once converted into WTP (A. R. Hole & Kolstad, 2012; Scarpa, Thiene, & Train, 2008; Train & Weeks, 2005).

While the number of applications in WTP is much smaller compared to the standard parameter space framework, studies are prevalent and growing across disciplines such as health (A. R. Hole & Kolstad, 2012; Özdemir, Johnson, & Hauber, 2009), transportation (Sonnier, Ainslie, & Otter, 2007; Train & Weeks, 2005), food (Balcombe, Chalak, & Fraser, 2009; de-Magistris, Gracia, & Nayga, 2013), and the environmental issues (Scarpa et al., 2008; Scarpa & Willis, 2010; Thiene & Scarpa, 2009).

Specifically, the model of indirect utility in our case can be re-expressed in Willingness to Pay Space as:

$$\begin{aligned} U_{ijt} = \alpha\big[\big(&Price_{njt} + \gamma_1 PoorW + \gamma_2 GoodW + \gamma_3 ExcellentW + \gamma_4 PoorS + \gamma_5 GoodS \\ &+ \gamma_6 ExcellentS + \gamma_7 BadC + \gamma_8 GoodC + \gamma_9 NotSafe + \gamma_{10} VerySafe\big)\big] \\ &+ e_{ijt} \end{aligned} \tag{2}$$

$$\text{such that } \gamma_i = \frac{\beta_i}{\alpha}$$

which is equivalent to the parameter space model in (2). Beyond the immediate benefit of producing welfare estimates directly from model estimation, this approach also inherently accounts for the scale parameter in each β, such that comparison across models is immediately available, unlike estimation of

(2). In our case, using WTP space accounts for scale so that a LR test comparing Forced and Unforced models is really a test for significantly different parameters. If the same test were used in preference space, it is only capable of testing for significant differences in parameters *or* of the scale parameter, μ.

We can also test for the effect of forced choices by using the pooled data of unforced and forced choice sets and adding an interaction for each attribute multiplied by the response if it was in choice set with a forced choice. This makes all attributes equal to 0 in unforced scenarios. The interactions represent systematic differences in the effect of an attribute in a forced response situation.

Using the WTP space model, we can test the effect of forced choice versus unforced choice using a likelihood ratio test for equality based on Swait and Louviere (1993)[2] with steps outlined in Louviere, Hensher and Swait (2000) (p. 364). The test statistic is:

$$-2(LL_j - \sum LL_i) \tag{3}$$

Distributed X^2 with K(M-1) degrees of freedom, where K is the number of restrictions and M is the number of treatments. LL_j is the log-likelihood of the pooled data, and ΣLL_i is the summation of the log-likelihood for each of the individual treatments. The null hypothesis of the test is that parameters are not significantly different between the two models.

In this study's case, LL_j is the conditional logit model of forced and unforced scenarios combined, which inherently ignores the underlying difference of the two groups. The next step is to run the forced choice responses and unforced choice responses as separate datasets, and then sum each of their own log likelihood, i.e. ΣLL_i. It is distributed X^2 with 10 degrees of freedom (travel cost is constrained to equal one across the models), since the number of treatments (M) (forced and unforced) equals 2, and the number of restrictions (K) equals 10. If the test statistic exceeds the critical value, then there is evidence

[2] For recent applications of the pooled LR test used in a similar manner, see Yang, Hu, Mupandawana, and Liu (2012), de-Magistris et al. (2013), and Tonsor and Shupp (2011).

to suggest that the treatments sufficiently contribute to explaining the model, and reject the null hypothesis of no difference between the forced and unforced datasets.

Instead, the same studies concluded that making distributional assumptions in WTP space produces more realistic values of WTP relative to preference space. We utilize Fiebig et al.'s GMNL based on the formulation by A. Hole, Gu, and Knox (2013). Based on the preceding model structure, we can compare the effect of forced versus unforced choice in WTP space.

Results

The summary statistics describing the tourists and residents in the sample are provided in Table 2. The sample is generally more well-educated and appears to be much younger for both groups. Sampled tourists tend to have a greater household income compared to the United States average. This is unsurprising since the funds required to vacation in Hawaii are comparatively high to other alternative recreational trips within the United States. Sampled residents tend to be below Hawaii norms. Lastly, tourists appear to be representative in terms of the length of time spent on Hawaii.

Table 2: Sample Summary Statistics

Characteristics	Hawaii Average[1]	Resident Sample	Tourist Sample	US Average[1]
Median Household Income [2]	$67, 116	$56,930	$75,932	$52,762
Female	49.9%	52.6%	48.3%	50.8%
Associate Degree or More	39.1%	59.1%	68.8%	36.3%
Age 18-25	10.99%	30.7%	27.8%	11.23%
Age 55 or Older	27.8%	12.4%	16.9%	25.5%
Days on Oahu[2]			8.0	7.37
# of Socioeconomic Responses		371	373	

[1]Based on information from the Hawaii Tourism Authority and the US Census
[2]Based on midpoint of each available response.

The conditional logit results in WTP space of the separate models are shown in Table 3[3]. Tourists completed 3510 choice sets (10 per person), and 341 were forced choices, 9.7%[4]. Column I represents results of a choice experiment without an opt-out alternative such as in Carlsson and Martinsson (2001). Each attribute is significant with the anticipated magnitude and direction. If we consider the model that only contains Unforced Choices (Column II), the results are fairly similar to the pooled results, with anticipated signs and magnitudes, as well as all variables remaining statistically significant. The WTP of appealing attributes decreased slightly compared to the pooled model, while WTP for the bad attribute levels slightly increased and decreased.

The model of Forced choices is starkly different from the Pooled and Unforced results. None of the attribute levels are statistically significant. WTP changes dramatically changing, at least doubling in most case, as well as a change in sign (Poor Sand). These results appear to be extremely unrealistic.

Table 3: Tourist Conditional Logit Results of Forced (I), Unforced (II), Pooled (III) and Pooled Responses with attribute interactions of forced decisions in Willingness to Pay Space

	I	II	III	IV	
Attribute	**Pooled WTP Responses**	**Unforced WTP Responses**	**Forced WTP Responses**	**Pooled WTP Responses**	**Forced Choice Interactions**
Round Trip Travel Cost	1	1	1	1	
Poor Sand	-17.92** (3.44)	-19.43** (3.55)	16.91 (36.20)	-21.50** (3.99)	25.66** (8.27)
Good Sand	6.29** (2.27)	5.89** (2.19)	20.56 (38.19)	6.31** (2.42)	.270 (7.52)
Excellent Sand	13.18** (2.84)	12.37** (2.74)	39.24 (59.64)	13.62** (3.06)	-2.20 (7.47)
Poor Water	-38.26** (5.85)	-35.98** (5.56)	-110.35 (158.24)	-39.77** (6.28)	10.10 (9.43)
Good Water	21.74** (3.46)	19.19** (3.17)	94.85 (135.22)	21.07** (3.56)	4.50 (6.96)
Excellent Water	46.83** (6.21)	42.60** (5.61)	184.21 (260.05)	46.90** (6.39)	2.51 (7.44)

[3] We attempted and were successful in some mixed logit models in WTP space, but convergence was problematic for the forced choice models. In the (immediate) future, we intend to pool the resident and tourist forced datasets to achieve model stability.

[4] It may be that there are significant differences in the characteristics of those who opt-out and those who do not as shown by Pedersen. A more complete scrutiny of this possibility should be considered.

Very Congested	-10.30** (2.43)	-9.71** (2.37)	-26.58 (41.83)	-10.74** (2.65)	2.40 (6.42)
Little Congestion	8.17** (2.15)	7.62** (2.07)	22.12 (38.78)	8.51** (2.31)	-3.17 (6.01)
Unsafe Waters	-21.14** (3.41)	-21.22** (3.37)	-29.85 (47.85)	-23.40** (3.80)	15.49* (6.69)
Very Safe Waters	11.09** (2.23)	9.94** (2.09)	40.32 (60.64)	10.97** (2.33)	-.26 (5.94)
n	3510	3169	341	3510	
LL	-2807.191	-2481.540	-310.577	-2794.387	

**p-value<.01 *p-value<.05
Standard Error Reported in Parentheses; clustered per respondent.
N is the number of choice sets
Note: The base category of each attribute is Average Quality

We then test for differences in the forced and unforced responses based on the likelihood ratio test from (3), equal to $[2807.191 - (310.577 + 2481.540)] * 2 = 30.148$.

Restrictions (K) equal 10 since the forced and unforced parameters must be equal in the pooled model, and there are two treatments (M), so the test statistic exceeds critical value for $X^2_{(10)}$ with a p-value of .0008. There is evidence to reject a pooled model and that separate models and parameters is appropriate.

For the second specification; the forced and unforced data were again pooled, but with interactions on of each attribute level multiplied by an indicator if it was in a forced choice. The forced interactions represent potential differences in the effect of an attribute in a forced choice relative to an unforced choice, with interaction parameter estimates appearing in the second column of model IV. Two of the forced choice attribute interactions are significant, Poor Sand Quality and Unsafe Waters. The Forced Unsafe Water interaction reduces the WTP to avoid the attribute Unsafe Water to $-7.91 (-23.40+15.49). For Poor Sand Quality, the forced choice actually reverses the sign of WTP from negative to positive (-21.50+25.66= 4.16), meaning that among those forced to choose, they have greater WTP for a beach with poor sand quality, similar to the result in III.

We turn our attention to the same analysis of residents in the sample, with results reported in Table 4. Residents sampled completed 3313 choice sets and 459 were ultimately forced choices, 13.9% of all decisions. The sign and magnitude of the Pooled and Unforced are as expected, each with significance. Differences in WTP between I and II appear small than for tourists. The Forced model in III yields similar mostly insignificant and sometimes counterintuitive results as it did for tourists. In this case though, good water and excellent water remain significant.

Table 4: Resident Conditional Logit Results of Forced (I), Unforced (II), Pooled (III) and Pooled Responses with attribute interactions of forced decisions in Willingness to Pay Space

	I	II	III	IV	
Attribute	**Pooled WTP Responses**	**Unforced WTP Responses**	**Forced WTP Responses**	**Pooled WTP Responses**	**Forced Choice Interaction WTP**
Round Trip Travel Cost	1	1	1	1	
Poor Sand	-11.99** (2.84)	-16.40** (3.41)	12.87 (8.16)	-16.83** (3.41)	28.37** (7.38)
Good Sand	7.50** (2.36)	7.45** (2.49)	9.59 (7.45)	7.58** (2.53)	1.12 (6.46)
Excellent Sand	12.65** (2.83)	12.70** (3.01)	11.03 (8.26)	12.98** (3.01)	-3.16 (6.83)
Poor Water	-25.38** (4.30)	-27.41** (4.88)	-10.72 (8.59)	-28.08** (4.78)	18.73* (7.89)
Good Water	16.77** (3.07)	16.13** (3.23)	20.29* (9.70)	16.45** (3.22)	1.69 (6.45)
Excellent Water	43.31** (5.80)	40.84** (5.92)	58.42** (21.36)	41.69** (5.74)	10.23 (6.37)
Very Congested	-15.76** (2.89)	-16.60** (3.17)	-8.77 (6.59)	-16.97** (3.14)	9.21 (5.71)
Little Congestion	9.20** (2.14)	10.09** (2.36)	2.96 (5.47)	10.35** (2.37)	-7.67 (5.25)
Unsafe Waters	-21.72** (3.41)	-24.04** (3.92)	-7.49 (6.31)	-24.57** (3.84)	17.93** (6.02)
Very Safe Waters	7.66** (1.95)	7.80** (2.08)	6.95 (5.70)	7.96** (2.10)	-1.75 (5.00)
n	3313	2854	459	3313	
LL	-2834.424	-2374.819	-427.760	-2802.392	

**p-value<.01 *p-value<.05
Standard Error Reported in Parentheses; clustered per respondent.
n is the number of choice sets
Note, the base category of each attribute is Average Quality

As before, we use the likelihood ratio test and calculate a test statistic to compare to the critical value for $X^2_{(10)}$, which equals $[2834.424 - (427.760 + 2374.819)] * 2 = 63.69$. The p-value is less than .0001. Therefore, there is evidence to reject the pooled model which imposes equal parameter values for forced and unforced choices.

Lastly, the pooled model with forced choice interactions with the attribute levels is in column IV. Each of the attributes is significant, and three interactions, Poor Sand, Poor Water and Unsafe Water attribute levels were significant. Unsafe Waters and Poor Water Quality drastically reduce the WTP to avoid those attributes, while Poor Sand again reverses the sign, such that forced responses have a positive WTP for Poor Sand quality. Based on the number of significant interactions, the difference in the results appears moderate.

Conclusion

The results from each of the models for tourists and residents show considerable evidence on the effect of forced choice. Specifically, joint tests for significant differences in the attributes were significant. Furthermore, there are some attribute-specific differences in forced vs unforced selection. This result is different from Carlsson, Frykblom & Lagerkvist (2007) who found no significant differences in attribute-level WTP, but did find significantly greater unobservable heterogeneity for their models of unforced choice. The result coincides somewhat with the slightly different models of Kallas and Gil (2012) with significantly different WTP for forced vs unforced model.

To extend this work, we intend to analyze the data using a mixed logit model in WTP space, which relaxes the assumption of Independence of Irrelevant Alternatives, and would allow for closer comparison to the work of Carlsson et al. (2007).

This works provides evidence supporting the importance of including an opt-out alternative, even after maintaining task complexity. Forcing a choice in a scenario of undesirable alternatives produces different welfare estimates, and in our case, some attributes with the theoretically incorrect sign. Including an opt-out should be standard, with only strong justification in the particular application to convince a researcher otherwise.

Adamowicz, W. L., & Boxall, P. (2001). *Future directions of stated choice methods for environment valuation.* Paper presented at the workshop on Choice Experiments: A New Approach to Environmental Valuation.

Balcombe, K., Chalak, A., & Fraser, I. (2009). Model selection for the mixed logit with Bayesian estimation. *Journal of environmental economics and management, 57*(2), 226-237.

Banzhaf, M. R., Johnson, F. R., & Mathews, K. (2001). Opt-out alternatives and anglers' stated preferences. *The choice modelling approach to environmental valuation*, 157-177.

Boxall, P., Adamowicz, W. L., & Moon, A. (2009). Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement*. *Australian Journal of Agricultural and Resource Economics, 53*(4), 503-519.

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P. F. (2006). The no-choice option and dual response choice designs. *Marketing Letters, 17*(4), 255-268.

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2007). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics, 34*(3), 321-344.

Carlsson, F., & Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment. *Journal of environmental economics and management, 41*(2), 179-192.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., . . . Swait, J. (1994). Experimental analysis of choice. *Marketing Letters, 5*(4), 351-367.

Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., . . . Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of environmental economics and management, 63*(1), 73-91.

de-Magistris, T., Gracia, A., & Nayga, R. M. (2013). On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments. *American Journal of Agricultural Economics, 95*(5), 1136-1154.

De Blaeij, A. T., Nunes, P. A., & Van den Bergh, J. C. (2007). 'No-choice'options within a nested logit model: one model is insufficient. *Applied Economics, 39*(10), 1245-1252.

DeShazo, J., & Fermo, G. (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of environmental economics and management, 44*(1), 123-143.

Dhar, R., & Simonson, I. (2003). The effect of forced choice on choice. *Journal of Marketing Research*, 146-160.

Greene, W. H., & Hensher, D. A. (2010). Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation, 37*(3), 413-428.

Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice Modelling Approaches: A Superior Alternative for Environmental Valuatioin? *Journal of economic surveys, 15*(3), 435-462.

Hasund, K. P., Kataria, M., & Lagerkvist, C. J. (2011). Valuing public goods of the agricultural landscape: a choice experiment using reference points to capture observable heterogeneity. *Journal of Environmental Planning and Management, 54*(1), 31-53.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological, 44*(6), 735-752.

Hole, A., Gu, Y., & Knox, S. (2013). Fitting the generalized multinomial logit model in Stata.

Hole, A. R., & Kolstad, J. R. (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. *Empirical Economics, 42*(2), 445-469.

Kallas, Z., Escobar, C., & Gil, J. M. (2012). Assessing the impact of a Christmas advertisement campaign on Catalan wine preference using Choice Experiments. *Appetite, 58*(1), 285-298.

Kallas, Z., & Gil, J. M. (2012). A dual response choice experiments (DRCE) design to assess rabbit meat preference in Catalonia: A heteroscedastic extreme-value model. *British Food Journal, 114*(10), 1394-1413.

Kontoleon, A., & Yabe, M. (2003). Assessing the impacts of alternative 'opt-out'formats in choice experiment studies: consumer preferences for genetically modified content and production information in food. *Journal of Agricultural policy and Resources, 5*(1), 1-43.

Ladenburg, J., & Olsen, S. B. (2010). Augmenting short cheap talk scripts with a repeated opt-out reminder in Choice Experiment surveys: University of Copenhagen, Institute of Food and Resource Economics.

Mak, J., & Moncur, J. (1998). Political economy of protecting unique recreational resources: Hanauma Bay, Hawaii. *Ambio*, 217-223.

Marsh, D., Mkwara, L., & Scarpa, R. (2011). Do Respondents' Perceptions of the Status Quo Matter in Non-Market Valuation with Choice Experiments? An Application to New Zealand Freshwater Streams. *Sustainability, 3*(9), 1593-1615.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Mentzakis, E., Ryan, M., & McNamee, P. (2011). Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. *Health economics, 20*(8), 930-944.

Mourato, S., Georgiou, S., Ozdemiroglu, E., Newcombe, J., & Howarth, A. (2003). Bathing water directive revisions: What are the benefits to England and Wales? A stated preference study: CSERGE Working Paper ECM 03-12. Center for Social and Economic Research on the Global Environment.

Murray, C., Sohngen, B., & Pendleton, L. (2001). Valuing water quality advisories and beach amenities in the Great Lakes. *Water Resources Research, 37*(10), 2583-2590.

Oh, C., Dixon, A., & Draper, J. (2006). *Visitor Needs Assessment and Economic Analysis of South Carolina Beaches*: Department of Parks, Recreation and Tourism Management.

Özdemir, S., Johnson, F. R., & Hauber, A. B. (2009). Hypothetical bias, cheap talk, and stated willingness to pay for health care. *Journal of health economics, 28*(4), 894-901.

Pedersen, L. B., Kjær, T., Kragstrup, J., & Gyrd-Hansen, D. (2011). Does the Inclusion of a Cost Attribute in Forced and Unforced Choices Matter?: Results from a Web Survey Applying the Discrete Choice Experiment. *Journal of Choice Modelling, 4*(3), 88-109.

Pedersen, L. B., Kjær, T., Kragstrup, J., & Gyrd-Hansen, D. (2012). Do General Practitioners Know Patients' Preferences? An Empirical Study on the Agency Relationship at an Aggregate Level Using a Discrete Choice Experiment. *Value in Health, 15*(3), 514-523.

Rigby, D., Alcon, F., & Burton, M. (2010). Supply uncertainty and the economic value of irrigation water. *European Review of Agricultural Economics, 37*(1), 97-117.

Rolfe, J., & Bennett, J. (2009). The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics, 68*(4), 1140-1148.

Rose, J. M., & Hess, S. (2009). Dual-Response Choices in Pivoted Stated Choice Experiments. *Transportation Research Record: Journal of the Transportation Research Board, 2135*(1), 25-33.

Savage, S. J., & Waldman, D. M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics, 23*(3), 351-371.

Scarpa, R., Thiene, M., & Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics, 90*(4), 994-1010.

Scarpa, R., & Willis, K. (2010). Willingness-to-pay for renewable energy: Primary and discretionary choice of British households' for micro-generation technologies. *Energy Economics, 32*(1), 129-136.

Scarpa, R., Willis, K. G., & Acutt, M. (2007). Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management, 50*(4), 449-466.

Sonnier, G., Ainslie, A., & Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics, 5*(3), 313-331.

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

Thiene, M., & Scarpa, R. (2009). Deriving and testing efficient estimates of WTP distributions in destination choice models. *Environmental and Resource Economics, 44*(3), 379-395.

Tonsor, G. T., & Shupp, R. S. (2011). Cheap Talk Scripts and Online Choice Experiments:"Looking Beyond the Mean". *American Journal of Agricultural Economics, 93*(4), 1015-1031.

Train, K., & Weeks, M. (2005). *Discrete choice models in preference space and willingness-to-pay space*: Springer.

Vujicic, M., Ryan, M., & Alfano, M. (Unpublished). *The Effect of Forced Choice on Choice: an Application to Nurse Location Decisions in Liberia*.

Yang, S.-H., Hu, W., Mupandawana, M., & Liu, Y. (2012). Consumer Willingness to Pay for Fair Trade Coffee: A Chinese Case Study. *Journal of Agricultural and Applied Economics, 44*(1), 21.