Systematic Non-Response in Stated Preference Choice Experiments: Implications for the Valuation of Climate Risk Reductions

Abdulallah S. Abdulrahman, Clark University, aabdulrahman@clarku.edu
Robert J. Johnston, Clark University, rjohnston@clarku.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Corresponding Author:

Robert J. Johnston
George Perkins Marsh Institute
Clark University
950 Main St.
Worcester, MA 01610
Phone: (508) 751-4619
Email: rjohnston@clarku.edu

Robert J. Johnston is Director, George Perkins Marsh Institute and Professor, Department of Economics, Clark University. Abdulallah S. Abdulrahman is a graduate research assistant, Department of Economics, Clark University. This research is supported by the Northeast Sea Grant Consortium, via prime award NA10AOR4170086 to MIT, Sea Grant (Sub award 5710003190). Opinions do not imply endorsement of the funding agency.

Systematic Non-Response in Stated Preference Choice Experiments: Implications for the Valuation of Climate Risk Reductions

Abstract

Discrete choice experiments (DCEs) addressing adaptation to climate-related risks may be subject to response biases associated with variations in risk exposure across sampled populations. Systematic adjustments for such biases are hindered by the absence of rigorous, standardized selection-correction models for multinomial DCEs, together with a lack of information on non-respondents. This paper illustrates a systematic approach to accommodate risk-related non-response bias in DCEs, where variations in risk exposure may be linked to observable landscape characteristics. The approach adapts reduced form response-propensity models to correct for survey non-response, capitalizing on the fact that indicators of risk exposure may be linked to the geocoded locations of respondents and non-respondents. An application to coastal flood adaptation in Connecticut, USA illustrates implications for welfare estimation. Results demonstrate that the proposed approach can reveal otherwise invisible, systematic effects of survey response patterns on estimated WTP.

Introduction

Discrete choice experiments (DCEs) are increasingly used to draw inferences regarding population preferences and willingness to pay (WTP) for environmental outcomes. The representativeness of the realized sample of respondents determines the validity of these inferences (Edwards and Anderson 1987; Messonnier et al. 2000; Whitehead 1991; Whitehead et al. 1993). Systematic differences between respondents and non-respondents that influence both decisions to respond and preferences elicited by the survey, or "non-passive" sample non-representativeness, is a source of potential bias in welfare estimation (Messonnier et al. 2000). Sample non-representativeness is recognized as an increasing problem for most types of survey research given declining response rates to household surveys (de Leeuw and de Heer 2002; Groves 2006; Groves and Petycheva 2008; National Research Council 2013).

Although the challenges of survey response bias are well-known and the properties of stated preference samples rarely match those of target populations[1], the possibility of non-response bias is overlooked by the substantial majority of published DCEs in environmental economics. Adjustments for non-response biases in DCEs—where made at all—rely primarily on demographic information (e.g., Cameron et al. 2005, 2013; Johnston et al. 2002). Approaches such as these are known to be inadequate when response biases are independent of these indicators (Groves 2006). Consider the example of programs to reduce climate-related hazards, such as the risk of coastal flooding due to sea level rise and coastal storms. It is possible that individuals whose homes are at greater risk of flooding may have higher WTP for flood risk reductions (relative to those with homes at lower risk), and may also be more likely to respond to a DCE addressing WTP for flood risk reductions. In such cases, individuals with higher WTP

[1] For example, see Edwards and Anderson (1987), Loomis (1987), Bockstael et al. (1990), Whitehead (1991), Whitehead et al. (1993), Cameron et al. (1996), Messonnier et al. (2000) and Cameron and DeShazo (2005).

may be over-represented in the final sample, leading to empirical estimates that (unless adjusted for systematic non-response) will not represent preferences of the broader population. Biases such as these can occur regardless of response rates (<100%) or demographic similarities or differences between the survey sample and broader population (Curtin et al. 2000; Groves 2006).

Concerns such as these are rarely addressed by DCEs in environmental economics. One reason for the lack of attention is the absence of a closed-form, Heckman-type corrections (Heckman 1979) for multinomial, multi-response discrete choice models, such as the mixed multinomial logit models commonly applied in stated preference DCEs (Cameron and DeShazo 2005). Another reason is that information on non-respondents sufficient for response bias corrections is rarely available; this is a common challenge for non-response correction methods across the spectrum of survey research (National Research Council 2013). Given these challenges, much of the DCE literature proceeds according to the implied but likely incorrect assumption that non-response biases are (a) not present, (b) inconsequential, or (c) accommodated using simple weighting approaches that rely on demographic data alone.

This paper illustrates a practical approach to assess and correct for non-response bias in multinomial DCEs. The method is developed for cases in which survey response may be modeled as a function of observable landscape characteristics. We illustrate the approach using a DCE on coastal flood adaptation, where survey response is hypothesized to be correlated with observable indicators of household flood exposure such as location within a flood zone. The proposed methods build upon the reduced form, two-step (or two-stage) approach of Cameron and DeShazo (2005, 2013). The first stage models the propensity to respond conditional on observable characteristics of respondents and non-respondents. The second stage incorporates fitted response propensities into a random utility model estimated over the realized sample.

Two additional contributions distinguish the approach developed here. First, given a paucity of data for non-respondents, prior applications rely primarily on demographic indicators in the first stage selection equation, typically proxied over geographic areas rather than individual households. This limits applicability to cases in which the likelihood of survey response can be adequately modeled using such proxies. In contrast, the illustrated models incorporate observable data available for both respondent and non-respondent households, drawn from GIS data layers linked to geocoded respondent and non-respondent addresses. Although we illustrate an application to coastal flood risk, parallel approaches could be applied in any policy context for which response propensity is related to observable environmental indicators that may be linked to geocoded addresses. Second, prior applications of such models note that the use of predicted regressors in the second stage leads to inaccurate standard errors, but nonetheless draw (admittedly suggestive) inferences from these results (Cameron and DeShazo 2005). To address this, we illustrate the use of a semi-parametric two-step bootstrap to evaluate the robustness of results to empirically corrected standard errors.

The result is reduced form approach to test and adjust for sample selection in multinomial DCEs that accounts for response propensity associated with multiple environmental indicators and corrects for inaccurate standard errors associated with the use of predicted regressors. Methods and results are illustrated using an application to coastal flood adaptation in Old Saybrook, Connecticut, USA. Results find significant and sometimes unexpected effects of response propensity on WTP that would remain invisible in the absence of an approach such as that illustrated here. The illustrated methods provide a means to correct for these patterns, thereby providing more representative welfare estimates.

Response Bias in Stated Preference Valuation

A relatively small number of studies have examined the issue of non-response bias in stated preference valuation, and most of these have been conducted within the context of open-ended or discrete choice contingent valuation (Edwards and Anderson 1987; Whitehead et al. 1993). In contrast to the multinomial models typical of contemporary DCEs, contingent valuation analyses are often conducted using simpler binomial or linear regression models for which Heckman or other standard corrections are available (e.g., Whitehead et al. 1993). However, even when corrections of this type are possible in principle, they are rarely applied, often due to the lack of sufficient data on non-respondents (Cameron et al. 1996; Messonnier et al. 2000).

More common in the stated preference literature are approaches that diagnose/discuss but do not correct for non-response bias, provide suggestive evidence based on demographic comparisons between samples and populations, or seek to correct for these differences via the incorporation of demographic or other interactions directly in utility functions or using observation weights (e.g., Johnston et al. 2002; Loomis 1987; Olsen 2009; Whitehead 1991). Rigorous treatments of non-response bias are even more unusual in applied environmental DCEs. Standard procedures for non-response correction (e.g., Dubin and Rivers 1989; Heckman 1979) are not directly applicable to the multinomial discrete choice models typically used for DCE data (Cameron and DeShazo 2005). In the absence of such corrections, most published DCEs fail to consider the possibility of response bias of any type.

Although rigorous Heckman-type corrections have not yet been developed for multinomial DCEs, Cameron and DeShazo (2005, 2013) illustrate a reduced form correction using a two-step modeling approach adapted to a multinomial model (cf. Greene 2003, pp. 183-186). This approach directly models the effect of estimated response propensity on the coefficients of interest, capitalizing on the observation that non-response bias is caused by a systematic relationship between response propensity and the measures under study (Groves 2006). The first stage uses a non-linear model to estimate the propensity of survey response, using data from both respondents and non-respondents. The propensity to respond is predicted using a binary probit model built around sociodemographic variables proxied (for non-respondents) using aggregate population data (Cameron and DeShazo 2005). The resulting propensity scores are incorporated into a second stage random utility model estimated over the realized sample. This approach enables the analyst to assess whether the predicted likelihood of survey response is influenced by variables included in the first stage model, and subsequently whether the predicted response propensity influences parameters estimated in the second stage.

Econometricians have routinely favored Heckman approaches to reduced form, structurally ad hoc corrections such as this (Cameron and DeShazo 2005; Cuddeback et al. 2004). The reason is obvious; where assumptions of the model hold and selection equations are correctly specified, it can be formally demonstrated that Heckman approaches eliminate bias due to systematic survey non-response. Outside of economics, the survey literature treats Heckman corrections with more caution, given the low likelihood that selection equations are correctly specified in practice (National Research Council 2013). When selection equations are not correctly specified, the ideal properties of Heckman corrections no longer hold, and these models can magnify rather than reduce bias (Cuddeback et al. 2004). The direct use of predicted response propensities (as illustrated here, e.g., to test and adjust statistical models in various ways) is more common outside of economics. Moreover, given the advantages and disadvantages of all non-response mitigation methods, the survey literature increasingly advocates the use of multiple approaches to diagnose and correct for these biases (Groves 2006;

National Research Council 2013; Schouten et al. 2009). The presented model is developed with this intent, as a feasible (if theoretically ad hoc) means to identify and adjust for non-response bias in multinomial DCEs, when formal Heckman-type approaches are unavailable.

The Theoretical Model

To develop the model, we begin with a random utility framework similar to those that underlie most environmental DCEs (Hanemann 1984). We assume that the utility of household n is determined by the choice of a multi-attribute coastal flood adaptation plan from a set of j alternatives (j = A, B, N). These include two multi-attribute adaptation options (A, B), and a status quo option (N) with no adaptation and zero cost. The household's utility, $U_{nj}(\cdot)$ is decomposed into a deterministic and stochastic component. The deterministic component, $V_{nj}(\cdot)$, includes observable attributes X_{nj} and C_{nj}, where X_{nj} is a vector of adaptation outcomes and C_{nj} is monetary cost. The resulting utility function can be represented

$$U_{nj}(\cdot) = U_{nj}\left(X_{nj}, C_{nj}\right) = V_{nj}\left(X_{nj}, C_{nj}\right) + \varepsilon_{nj}, \qquad (1)$$

where ε_{nj} represents the stochastic component of the utility function, modeled as a random error. Common specifications assume that (1) may be estimated using an additively separable, linear-in-the-parameters function such as

$$U_{nj}\left(X_{nj}, C_{nj}\right) = \beta_1 X_{nj} + \beta_2 C_{nj} + \varepsilon_{nj} \qquad (2)$$

where β_1 is a vector of parameters on policy outcomes and β_2 is the parameter on household cost. When making a choice between policy alternatives (j = A, B, N) the household is assumed to choose the alternative that provides the greatest anticipated utility. Utility parameters may be estimated using a variety of the maximum likelihood models for discrete dependent variables, with likelihood functions determined by assumptions regarding such factors as the unobservable components of utility and preference heterogeneity among respondents (Train 2009). If estimated in preference space as implied above (in contrast to WTP space, see Train and Weeks 2005), the estimated parameters represent measures of attribute marginal utilities.

Incorporating Response Propensity

Non-response bias occurs in stated preference analysis when the preferences of respondents differ systematically from those of non-respondents. This may be due to self-censoring behavior of non-respondents or active culling by researchers, e.g., to remove outliers (Edwards and Anderson 1987). To illustrate the effect on the model developed above, we adapt prior examples such as Edwards and Anderson (1987) and Messonnier et al. (2000). Although these were developed for contingent valuation with WTP as the dependent variable, parallel approaches apply to random utility models (additional estimation challenges notwithstanding; see below).

The model begins with an observation that equation (2) can only be estimated for those who provide usable data in response to the DCE, or survey respondents. Hence, the estimated utility equation is

$$U_{nj}\left(X_{nj}, C_{nj}\right) = \beta_1 X_{nj} + \beta_2 C_{nj} + \varepsilon_{nj} \;\; (n = 1 \ldots N_R < N) \tag{3}$$

where N_R is the number of respondents providing usable data and N is the total set of respondents and non-respondents (assumed to be a true random sample of the population). Although equation (3) is linear in the parameters, it is estimated within DCEs using a non-linear model (e.g., conditional or mixed multinomial logit). We further assume that the latent propensity that household *n* responds to the survey may be specified

$$P_n^* = \alpha_1 Z_n + e_n, \tag{4}$$

where Z_n is a vector of exogenous variables determining the individual's response propensity, α_1

is a conforming vector of parameters, and e_n is an independently and identically distributed equation error. Although P_n^* is unobservable, the observable counterpart is P_n is a discrete variable taking on a value of 1 if the individual responds to the survey and 0 if the individual does not respond, such that $P_n = 1$ when $P_n^* \geq 0$ and $P_n = 0$ otherwise (Messonnier et al. 2000). Equation (4) may be estimated using common approaches such as binary logit or probit, and is interpreted as the sample selection rule (SSR).

Combining (3) and (4),

$$E\big(U_{nj}(\cdot)\big|X_{nj}, C_{nj}, SSR\big) = \beta_1 X_{nj} + \beta_2 C_{nj} + E(\varepsilon_{nj}|SSR) \tag{5}$$

where $E(\cdot)$ is expected value. Given (5), if ε_{nj} is conditional on the decision rule (e.g., if individuals who are more/less likely to respond have systematically different preferences, such that ε_{nj} and e_n are not independent), then $E(\varepsilon_{nj}|SSR) \neq 0$, leading to potential bias in the estimation of model parameters. If, in contrast, $E\big(\varepsilon_{nj}\big|SSR\big) = 0$, then non-response is "passive" and does not influence welfare estimation (Messoniere et al. 2000).

Heckman approaches may be used to mitigate non-response biases of this type when the primary estimating equation is linear, but parallel approaches are unavailable for non-linear, multi-attribute, multinomial response models such as those applicable to most DCEs. In the absence of rigorous Heckman corrections, the approach illustrated by Cameron and DeShazo (2005, 2013) takes the form of a two-step (or two-stage) nonlinear regression (Greene 2003, p. 183; Murphy and Topel 1985) that directly estimates the effect of response propensity (the SSR) on utility parameters.[2]

Within this model, the first stage estimates (4) using data on respondents and non-

[2] One may also use the (inverse of) estimated response propensity as an observation weight prior to estimation of the second stage model (Groves 2006).

respondents (e.g., via a binary logit model) to obtain

$$\hat{P}_n = E(P_n|Z_n) = f(\hat{\alpha}_1 Z_n) = \frac{e^{\hat{\alpha}_1 Z_n}}{1 + e^{\hat{\alpha}_1 Z_n}}. \tag{6}$$

The resulting $\hat{P}_n$ represent estimates of individuals' response propensities conditional on observable characteristics Z_n. Note that this requires suitable data on Z_n for both respondents and non-respondents. The $\hat{P}_n$ are then incorporated as alternative invariant covariates into utility function (3) via interactions with X_{nj} and C_{nj}. This second-stage specification enables marginal utilities associated with policy outcomes to vary directly according to individuals' predicted response propensities,

$$U_{nj}\left(X_{nj}, C_{nj}\right) = \beta_1 X_{nj} + \beta_2 C_{nj} + \beta_3 \hat{P}_n X_{nj} + \beta_4 \hat{P}_n C_{nj} + \varepsilon_{nj} \;\; (n = 1 \ldots N_R < N) \tag{7}$$

with β_3 and β_4 representing conforming vectors of parameters to be estimated. Within (7), the marginal utilities of vector X_{nj} are given by $(\beta_1 + \beta_3 \hat{P}_n)$ and the marginal utility of income is given by $(\beta_2 + \beta_4 \hat{P}_n)$, such that they are a linear function of response propensity. If $\beta_3 \neq 0$ or $\beta_4 \neq 0$, then variation in the predicted likelihood of being in the sample is associated with variations in marginal utility within the estimation sample, allowing the presence of response bias to be inferred (Cameron and DeShazo 2005). One can subsequently adjust marginal utilities and WTP to mitigate this bias by setting $\hat{P}_n$ equal to either the mean or median from the entire sample of respondents and non-respondents (Cameron and DeShazo 2013).

Implementation of this model faces two additional challenges not addressed by the extant literature, and for which we propose practical solutions. First, data required to predict response propensity (Z_n) are required, and sufficient data of this type are rarely observable for non-respondents. Second, although the use of $\hat{P}_n$ as a predicted regressor in (7) leads to consistent estimates of coefficients in the second stage utility equation, the estimated standard errors are inaccurate (Cameron and DeShazo 2005; Greene 2003, p. 183-186). To address the first challenge, we illustrate an approach whereby Z_n may be quantified by linking geocoded addresses of respondents and non-respondents to GIS data layers that quantify primary variables affecting survey response. This provides a means to obtain directly relevant and observable information on non-respondents unavailable in prior efforts. To address the second concern, we develop a two-step semi-parametric bootstrap that empirically corrects standard errors for the use of predicted regressors in (7), thereby allowing more reliable inference. The goal is a broadly applicable and reliable reduced-form means to identify and offset WTP variations associated with differences in response propensity across respondents and non-respondents.

Empirical Application

The model is illustrated using a DCE addressing coastal flood adaptation in the town of Old Saybrook, Connecticut, USA. The DCE questionnaire, *Adapting to Coastal Storms and Flooding*, elicited preferences for measures to protect physical and ecological assets such as homes, tidal marshes and beaches from loss due to coastal flooding and erosion (Johnston et al. 2015). The DCE was developed over more than two years in a process that engaged economists, coastal ecology experts, engineers, municipal official, and other stakeholders. Thirteen focus groups were held with town residents to inform and test the questionnaire and model, employing ethnographic methodology as described by Johnston et al. (1995). Survey language and graphics were subject to extensive pretesting in focus groups and cognitive interviews (Kaplowitz et al. 2004), including the use of verbal protocols to gain insight into respondents' comprehension and decision processes (Schkade and Payne 1994).

The data used to inform DCE scenarios were obtained from sources including Columbia University's Center for Climate Systems Research, NASA's Goddard Institute for Space Studies, The Nature Conservancy (TNC), and the National Oceanic and Atmospheric Administration (NOAA), as reflected in coastal flooding scenarios for TNC's Coastal Resilience platform (www.coastalresilience.org). Choice attributes were selected based on a conceptual model combining input from focus groups; scientists with expertise in sea level rise and coastal resilience; coastal flooding scenarios; and interviews with municipal officials and stakeholders.

Following the theoretical model outlined above, each choice question allowed the respondent to choose among three possible multi-attribute adaptation plans ($j = A, B, N$), including two multi-attribute adaptation options (A, B) and a status quo (N) with no new adaptation and zero household cost. Prior to administration of choice questions, the survey provided information describing tradeoffs associated with alternative approaches to coastal adaptation and projected inundation scenarios in the mid-2020s and baseline (status quo) effects with no new adaptation. This and other information was conveyed via a combination of text, custom graphics, geographic information system (GIS) maps and photographs. Detailed instructions were also provided, including reminders to consider budget constraints and specific statements highlighting consequentiality (Carson and Groves, 2007; Johnston, 2006).

Choice options are characterized by six attributes: (1) the percentage and number of homes expected to flood in a Category 3 storm, (2) wetland acreage lost, (3) beach and dune acreage lost, (4) the length of coastline that is hard-armored, (5) the general emphasis of adaptation efforts (whether there is additional emphasis on hardened coastal defenses), and (6) unavoidable household cost (Table 1). All outcomes are forecast as of the mid-2020s. Following the general approach of Johnston et al. (2012), attributes represent each adaptation method and effect in relative (percentage) terms with regard to upper and lower reference conditions (i.e., best and worst possible in Old Saybrook) as defined in the survey. Scenarios also present the cardinal basis for relative levels where applicable. Table 1 provides summary statistics and definitions of each attribute. Table 2 illustrates attribute levels, chosen based on feasible adaptation outcomes identified using the data sources identified above.

Grounded in these attributes and levels, a fractional factorial experimental design was generated using a D-efficiency criterion (Ferrini and Scarpa 2007; Sándor and Wedel 2001, 2002; Scarpa and Rose 2008) for main effects and selected two-way interactions, yielding 72 profiles blocked in 24 booklets. Although optimized for D-efficiency, other measures of efficiency were also reviewed, e.g., S-efficiency, to evaluate potential sample sizes required for assumed utility specifications (Bliemer et al. 2009, Rose and Bliemer 2008; Scarpa and Rose 2008). Design efficiency was reevaluated using alternative assumptions for utility structure. Each respondent was provided with three choice questions and was instructed to consider each as independent and non-additive. A sample choice question is illustrated in Figure 1.

The DCE was implemented from May-June 2014 over a random sample of Old Saybrook households. The self-administered questionnaire was distributed via U.S mail, with follow-up mailings to increase response rates (Dillman et al. 2009). Three different versions of the DCE were implemented, distinguished by slight variations in the attribute set provided to respondents.[3] As these minor intra-version variations are assumed to have no substantive impact on response propensity, data from all versions were used to estimate the response propensity model in equation (6). However, to ensure the direct comparability of all attributes, the random utility model in equation (7) is estimated using data from the single, primary survey version. For

[3] These variations were used to test hypotheses unrelated to non-response bias.

all versions combined, 1,729 surveys were mailed, from which 1,489 were deliverable. Of these, 489 were returned, for a 32.8% response rate, with 423 providing sufficient data for the response propensity model. For the survey version used for random utility modelling, 576 surveys were mailed, 488 were deliverable and 163 were returned, for a 33.4% response rate. Returned surveys yielded 408 complete choice responses, from which the random utility model is estimated.

Data for Response Propensity Modeling

As noted above, development of the response propensity model requires data for respondents and non-respondents. Given the focus of the DCE on coastal flooding, we expect that variables influencing response propensity would include indicators of each household's flood exposure, e.g., distance from the shoreline, property elevation, and whether the home is in a flood hazard zone. Multiple indicators of this type may be extracted from readily accessible GIS data layers, combined with geocoded home location of both respondents and non-respondents (from their physical mailing addresses).[4] Maps showing the approximate geocoded location of homes for all targeted households, respondents and non-respondents are shown in Figure 2.

This approach enabled the extraction of multiple characteristics related to each household's flood risk exposure, such that the response propensity model can be populated with multiple exposure indicators observable for each household. We also include the gender of respondents/non-respondents, with the gender of non-respondents inferred using the first name and gender prefix in the mailing list. The gender of respondents is available from survey responses. Table 3 summarizes variables included in the response propensity model.

[4] Surveys were only mailed to physical addresses, not including post office boxes.

Model and Welfare Estimation

As a precursor to estimation of the response propensity model, Table 4 compares risk exposure variables (Table 3) across respondent and non-respondents, and tests for the equivalence of means across the two groups. A significant difference between respondents and non-respondents across these variables would provide a preliminary though not definitive suggestion that response propensity is related to these variables. As shown in Table 4, we fail to reject the null hypothesis of equal means (for all risk indicators) across the two groups. This result suggests that the univariate risk-exposure properties of respondents are similar to those of non-respondents. However, these initial results must be treated with caution, as comparisons based on univariate t-tests cannot identify variations in response propensity based on multivariate patterns.

The first-stage response propensity model follows equations (4) and (6), with Z_n (including the exposure variables in Table 3) as main effects and interactions. The model is estimated using data on respondents and non-respondents to all survey versions.[5] Parameters are estimated using a binomial logit model, with results used to generate fitted response probabilities for each respondent in the realized sample. These fitted response propensities are incorporated into structural estimation of the utility function in the second stage, following equation (7).[6] Predicted propensities enter the model via multiplicative interactions with policy attributes X_{nj}. The model does not include an interaction with cost, C_{nj} because the resulting coefficient estimate is highly insignificant, and doing so hence leads to undefined welfare estimates.[7]

[5] Slight variations in attributes between survey versions are not expected to influence response propensity patterns.
[6] Fitted response propensities are scaled up by 100 before incorporation into the second stage model.
[7] This interaction, when included, leads to an insignificant coefficient estimate with a large standard error. Given a random coefficient on *Cost* in the mixed logit model (here, bounded triangular), this large standard error on the interaction causes a portion of the estimated distribution of the marginal utility of income ($\beta_2 + \beta_4\hat{P}_n$) to overlap zero. As discussed by Daly et al. (2012) and Hole (2007), the result is an undefined mean welfare estimate for the sample. To avoid this problem, we do not include the interaction between the response propensity and program cost in the random utility model. The exclusion of this interaction has no significant effect on other aspects of the model.

The random utility model is estimated using a simulated-likelihood mixed logit (ML) in preference-space with 2000 Halton draws, using data from the realized sample of respondents to the primary survey version.[8] The final model was chosen based on the results of preliminary models with varying specifications. Except for the coefficient on *Seawalls*, which is specified as fixed, coefficients on all main effects are specified as random and independent. We assume a normal distribution for coefficients on *Neither*, *Hard*, *Homes*, *Wetlands*, and *Beaches*. The coefficient on *Cost* is specified with a bounded triangular distribution, with the sign reversed prior to the estimation to ensure a positive marginal utility of income (Hensher and Greene 2003; Campbell et al. 2008; Johnston et al. 2012; Johnston and Ramachandran 2014). Coefficients on the interactions with response propensities are specified as fixed.

Estimates on the response propensity interactions capture the variability in mean marginal utilities of policy attributes associated with variations in the predicted likelihood of being in the realized sample, as a function of flood exposure indicators. Here, the null hypothesis of interest is whether the marginal utilities of X_{nj} vary as a function of response propensity, or whether $\beta_3 = \partial^2 U_{nj}(.)/\partial X_{nj}\,\partial \hat{P}_n = 0$. Rejecting the null would imply that the variation in the likelihood of completing and returning the survey questionnaire explains at least some of the heterogeneity around the mean estimate of marginal utility. Consequently, the unadjusted utility estimates would likely not reflect the true preference parameters in the targeted population.

However, the presence or absence of response bias in marginal utilities does not necessarily imply corresponding impacts on WTP, and it is often these estimates that are of greatest policy relevance (Johnston et al. 2005). Hence, we continue the analysis through the

---

[8] Parallel WTP–space models would not converge with the response propensity interactions, hence all models are estimated in preference space. Models were tested with alternative numbers of Halton draws to evaluate stability.

estimation of implicit prices using the welfare simulation approach described by Johnston and Duke (2007), following Hensher and Greene (2003). We first conduct a parameter simulation following Krinsky and Robb (1986), with R=1000 draws taken from the mean parameter vector and associated covariance matrix. For each of these R draws, the resulting parameters characterize asymptotically normal empirical densities for fixed and random coefficients. For each draw, a coefficient simulation is conducted for each random coefficient (capturing preference heterogeneity), with S=1000 draws taken from simulated empirical densities. Welfare measures are calculated for each draw, resulting in a combined distribution of $R{\times}S$ observations from which summary statistics are derived (Hensher and Greene 2003). Statistical significance is determined by the percentiles on these empirical distributions (Poe et al. 2005).

The resulting estimates may be used in a variety of different ways to evaluate and potentially adjust for the effect of response propensity on WTP. First, we evaluate the marginal effect of response propensity on the implicit price for each attribute, here given by $\partial WTP / \partial \hat{P}_n = \hat{\beta}_3/\hat{\beta}_2$ (recall that $\hat{\beta}_3$ is a vector of estimated coefficients associated with the interaction of $\hat{P}_n$ and the various X_{nj}). The result indicates whether and to what extent, on average, response propensity influences WTP for different coastal adaptation attributes.

These results indicate the effect of a unit change in $\hat{P}_n$ on implicit prices, or the predicted difference in marginal WTP associated with a one percentage point increase in response propensity. However, of greater policy relevance is the effect on mean welfare estimates reported over the entire realized sample, reflecting the potential effect of non-response bias. That is, to what extent are WTP estimates influenced by the variation in response propensities between the realized sample (of respondents), and the combined sample of both respondents and non-respondents (with the latter assumed to reflect the broader population)? To answer this question, we calculate differences between implicit prices evaluated at the mean value of predicted response propensities for the entire sample (both respondents and non-respondents) and analogous implicit prices evaluated at the mean value of response propensity for the realized sample of respondents (Cameron and DeShazo 2013). This is calculated

$$WTP_{diff} = \frac{\hat{\beta}_{1} + \hat{\beta}_3 \hat{P}_{ma}}{\hat{\beta}_2} - \frac{\hat{\beta}_{1} + \hat{\beta}_3 \hat{P}_{mr}}{\hat{\beta}_2} = \frac{\hat{\beta}_{1} + \hat{\beta}_3 \left(\hat{P}_{ma} - \hat{P}_{mr}\right)}{\hat{\beta}_2}, \tag{8}$$

where $\hat{P}_{ma}$ is the mean of fitted values of response probabilities calculated over the entire sample of respondents and non-respondents, and $\hat{P}_{mr}$ is the corresponding mean calculated using the realized sample (respondents). This difference (WTP_{diff}) is a reduced form estimate of the extent to which implicit prices differ between the realized sample and the population, based on risk-related variations in response propensity.

Standard Error Corrections

A well-known consequence of the use of fitted (or predicted) regressors in (7) is biased estimates of coefficient standard errors (Cameron and DeShazo 2005; Greene 2003; Murphy and Topel 1985). This leads to the potential for incorrect statistical inferences regarding the effect of response propensities. In cases such as this, bootstrapping can provide an empirical means to estimate corrected standard errors without the assumptions implied by alternative approaches (Davidson and Mackinnon 1999; Guan 2003; King and Roberts 2015; Skrondal and Rabe-Hesketh 2009). In the present case, however, the application of a simple two-stage non-parametric bootstrap is complicated by (1) the use of different samples to estimate the first- and second-stage models [9], and (2) the use of a complex and computationally intensive mixed logit

[9] Recall, the first stage is estimated using data on respondents and non-respondents from all survey versions. The second stage is estimated using data on respondents to a single survey version.

estimator in the second stage that may not converge over all iterations of the bootstrap.

To address these complications, we develop an iterative, semi-parametric bootstrap, applied to a simplified version of the two-step model with a conditional rather than mixed logit model as the second stage. This enables us to evaluate the extent to which standard errors are affected by the use of predicted response propensities, for a simplified model over which bootstrapping is feasible. To develop the illustrated approach, the first-stage response propensity model is estimated in identical fashion to that above, leading to parameter estimates $\hat{\alpha}_i$ and an associated covariance matrix which reflects the statistical precision of these estimates. Drawing from these first-stage estimates, we implement a parametric bootstrap to account for sampling variation in $\hat{P}_n$. The procedure begins by repeatedly taking *c* random draws of $\hat{\alpha}_i$ (*c*=1000 or 10,000 draws) using the mean parameter vector and the associated covariance matrix. Each c^{th} draw of the parameters ($\hat{\alpha}_{ic}$) is used to calculate $\hat{P}_{nc}$ for each respondent, leading to an empirical distribution of *c* estimates of predicted response propensity for each respondent, reflecting sampling variation in the underlying parameters.

The second-stage then draws *c* independent non-parametric bootstrap samples from the second-stage model data, reflecting DCE responses from realized sample of respondents. The c^{th} bootstrap sample of the DCE data is then paired with $\hat{P}_{nc}$ from first stage parametric draws, with a conditional logit model estimated over the resulting data. Identical to the specification in our original mixed logit model, we incorporate predicted propensities into the model through interaction with all attributes except the *Cost*. The result is *c* bootstrap estimates of the conditional logit model, each estimated over a unique bootstrap sample of the DCE data and a different parametric bootstrap draw of $\hat{P}_{nc}$ for each respondent. These estimates are used to generate robust, semi-parametric bootstrap standard errors for each of the coefficients (estimated marginal utilities) in the random utility model, that account for the fact that response propensity is predicted rather than observed. For comparison, we estimate the model using c=1000 and 10,000 bootstrap replications. The resulting robust standard errors are compared to the original (uncorrected) standard errors from the conditional logit model, providing insight on the extent to which standard errors are influenced by the use of a predicted regressor in the second-stage.

Results

Results of the first stage selection model are presented in Table 5. As expected based on initial findings in Table 4, logit model results suggest that observable indicators of flood exposure explain a relatively small proportion of the variation in response propensity. This is a positive finding in terms of non-response bias, as it suggests that response likelihood is not heavily influenced by the household's flood exposure.

Three variables have a statistically significant impact, however. Results for these variables suggest at least some patterns that may defy prior expectations. Coefficients on *Distance×Elevation* and *Distance×SFHZ* are statistically significant and positive. The former suggests that households that are both at greater distance from the coastline and at higher elevation—and hence at *lower* flood risk—are *more* likely to respond to the survey. (Coefficient estimates associated with the main effects *Distance* and *Elevation* are negative, but not statistically significant.) The positive and significant coefficient estimate for *Distance×SFHZ* further implies that for households in a special flood hazard zone, greater distance to the coastline implies a greater propensity to respond. Both of these lead to the conclusion that survey recipients whose homes are less vulnerable to flood risks are more likely to respond, although the magnitude of these effects are relatively small. These results are robust across multiple specifications of the response propensity model. The estimate on *Male* also indicates that men are more likely to respond than others (Table 3).

Although these results might seem unexpected, focus group results suggest an intuitive explanation. Within Old Saybrook, many current and potential adaptation solutions involve restrictions placed on those whose properties are at high risk of flooding—for example, restrictions that prevent rebuilding of homes subject to repeated and extensive flood damage (Town of Old Saybrook 2015). As a result, focus group results suggest that those at higher risk of flooding may be more skeptical of community-level coastal adaptation decisions and may distrust in the policy process. Such lack of trust can have significant effects on stated preference survey responses (Johnston et al. 1999). In the present case, it may cause homeowners at higher risk (and with potentially less trust in the policy process) to be less likely to return the DCE.

The estimated model is used to predict response propensity scores for both respondents and non-respondents, a summary of which is provided in Table 6. These predicted propensities are incorporated into the second-stage random utility model estimated over the realized sample of respondents. The mean of predicted propensity scores for the entire sample (respondents and non-respondents to all survey versions) is 24.46%. This mean is approximately one percentage point lower than the parallel response propensity calculated over the realized sample of respondents for the main survey (25.23%). Hence, in the present case, there is a difference in the predicted response propensity of respondents and non-respondents, but this difference is small.

Second Stage: Estimated Utility Functions

Table 7 reports the results of the estimated random utility model. Two specifications are illustrated. The first is a restricted model that omits interactions with response propensities—this is akin to standard, non-corrected models reported in the literature. The second model is an unrestricted model that includes interactions between non-cost choice attributes and predicted response propensities for each respondent in the realized sample. Specification of the main effect coefficients in both models is identical. Estimated coefficients are jointly significant at $p <$ 0.0001 for both models, with Pseudo R^2s of 0.19 and 0.21 for restricted and unrestricted model, respectively. A likelihood ratio test rejects the null hypothesis that the joint effect of response propensity interactions is zero ($\chi^2 = 23.08$, df. 6, p<0.001), implying that estimated marginal utilities vary as a function of fitted response propensities.

As noted above, although the second stage model provides consistent estimates of mean coefficient values (and hence implied effects of response propensity on marginal utilities), estimates of standard errors are inaccurate. Hence, inferences regarding statistical significant should be interpreted with caution. Below we present estimates of semi-parametric bootstrap-corrected standard errors for simpler conditional logit model, suggesting that most inferences drawn from the primary model are robust. Prior to presenting these results, however, we discuss initial results of the unrestricted mixed logit model with uncorrected standard errors.

For the unrestricted model, coefficients on all attribute main effects are statistically significant except for those on *Beaches* and *Seawalls*. The signs of statistically significant coefficient estimates match prior expectations and are identical to the corresponding estimates in the restricted model.[10] All estimated standard deviations/spreads of random parameter distributions are statistically significant at p < 0.10 or better in both models.

[10] Magnitudes of main effect coefficients cannot be compared across the two models, because (1) the effect of each attribute in the unrestricted model includes both the main effect coefficient and the coefficient on the response propensity interaction, and (2) scale confounds parameter magnitudes in mixed logit models (Fiebig et al. 2010).

Response Bias in Estimates of Marginal Utilities

Results of the unrestricted model suggest that simple comparisons between respondents' and non-respondents' characteristics (Table 4) may lead to misleading conclusions about the potential presence of response bias. Here, parameters on response propensity interactions may be interpreted as the deviation in mean marginal utilities associated with one percentage point change in the predicted likelihood of response. In contrast to the implications of simple difference-in-means tests (Table 4), results of the unrestricted model (Table 7) indicate the presence of statistically significant effects of response propensity on marginal utilities, suggesting the presence of non-response bias. Initial (uncorrected) standard errors suggest statistically significant ($p<0.05$) coefficients on $Score \times Neither$, $Score \times Hard$ and $Score \times Wetlands$.

The positive sign on ($Score \times Neither$) implies that the baseline disutility from no action (*Neither*, the ASC on the status quo) is lower for those who have a higher predicted response propensity, conditional on risk exposure (those with higher response propensities are more likely to choose the status quo). The estimate on $Score \times Hard$ indicates that the marginal disutility from an emphasis on hard adaptation (*Hard*) is lower for those with higher predicted response propensities (those with higher response propensities are more likely to support plans that focus on *Hard* or engineered adaptation). The estimate on ($Score \times Wetlands$) implies that those who are more likely to respond obtain less utility from wetlands protection (respondents with higher response propensity lose less utility when wetlands losses increase).

Taken together with the first-stage response propensity model, these results provide an intuitive perspective on survey non-response patterns. As noted above (Table 5), those with greater physical exposure to flood risks are less likely to respond to the survey. Hence, it is not surprising that those who are more likely to respond (and are both further from the coast and less vulnerable to flood risks): (1) are more likely to choose the no-action status quo, (2) are less likely to have negative preferences for hardened shoreline structures and (3) have lower values for the protection of coastal wetlands. The latter two findings are intuitive based on distance decay intuition (Bateman et al. 2006), assuming that households that are more distant from the coast are less effected by the (negative) amenities of hardened shoreline structures and the (positive) amenities of coastal wetlands.

These findings suggest the presence of significant non-response bias in marginal utility estimates. These results correspond to focus group results, but are not of a type that might be initially expected in the context of environmental hazards. That is, those who are less vulnerable to flood risks are more likely to return the survey, and are hence slightly over-represented in the sample. Combined with the effect of response propensity on marginal utilities, this tends to reduce the magnitude of certain marginal utilities (positive or negative) related to coastal adaptation outcomes. We emphasize, however, the conclusions regarding the statistical significance of these and other estimates in the second-stage model are subject to further verification in models that correct for inaccurate standard errors.

Response Bias in WTP Estimates

Implicit price estimates simulated from mixed logit results are shown in Tables 8. Risk-related response bias in marginal WTP estimates is reported in the second column, measured as the WTP difference associated with a unit change in the predicted response propensity. As expected, these differences show similar patterns to those found in marginal utility estimates (Table 7), with higher (less negative) WTP for *Neither*, *Hard* and *Wetlands*.

Results indicate that, holding all other adaptation attributes fixed, a household WTP to avoid the status quo (*Neither*) is \$21.39 less than a comparable household with one point lower predicted response propensity. The estimate on *Hard* indicates that a one point increase in predicted propensity is associated with \$30.90 lower WTP to avoid a flood protection plan that emphasizes hard defenses relative to the status quo. Finally, the estimate on *Wetlands* indicates that per household annual WTP to avoid the loss of one percentage point of currently existing wetland acres is \$2.60 lower for a household with a one point higher response propensity. Interestingly, despite the fact that the response propensity score is constructed based on the exposure of individuals' homes to flood risks, the difference in WTP for *Homes* is insignificant.

Table 8 also presents mean implicit prices evaluated at the mean propensity score for respondents only (column 4), and at the mean score for all sampled individuals (column 5). The differences between these estimates may be interpreted as bias in mean implicit prices related to survey non-response. Results indicate that per household WTP to avoid the status quo (*Neither*) is \$140.79 when evaluated using the mean propensity of the realized sample. This increases by \$16.35 (to \$157.14) when evaluated using mean predicted propensity of both respondents and non-respondents. The parallel difference in WTP for *Hard* is \$23.62. Implicit price estimates for *Wetlands* indicate that annual WTP to prevent the loss of one percentage point of wetland acres is \$7.47 when evaluated using mean propensity of respondents, and not statistically significant. This estimate increases to \$9.45 and becomes statistically significant ($p<0.06$) when evaluated at mean of predicted propensity of the entire sample.

Taken together, these results suggest that estimates of WTP for coastal flood adaptation outcomes are influenced by response propensity, and that uncorrected estimates may lead to misleading conclusions regarding these values. The direction and magnitude of these effects, although consistent with focus group results, may not always match prior expectations. This observation highlights the importance of obtaining empirical estimates of these patterns.

Accuracy of Statistical Inferences

To evaluate the extent to which standard errors are influenced by the use of predicted rather than observed response propensities in the second-stage, Table 9 illustrates results of the semi-parametric bootstrap over conditional logit estimates. Results for both 1,000 and 10,000 bootstrap iterations are reported. Results of the conditional logit model are generally consistent with those of the mixed logit model (Table 7). Hence, we focus our discussion on interactions with predicted propensities and the resulting coefficient standard errors. As found in the mixed logit model, estimates on $Score \times Hard$ and $Score \times Wetlands$ are positive and statistically significant at $p < 0.01$ based on the conditional logit coefficient estimates and uncorrected standard errors. Semi-parametric bootstrap-corrected standard errors on these coefficients increase as expected, but these estimates remain statistically significant, albeit at a reduced level of $p < 0.05$. Other standard errors from the model show similar patterns, with modest differences (generally increases as expected) between the uncorrected and corrected standard errors, but general inferences and significance remaining unchanged. The coefficient on the interaction $Score \times Neither$ is not significant in the conditional logit, based on either the uncorrected or corrected standard errors. Results are robust over both 1,000 and 10,000 bootstrap samples.

These results suggest that inferences regarding the effect of response propensity continue to hold when standard errors are adjusted for the use of predicted regressors in the second-stage. Semi-parametric bootstrap results suggest that the degree of inaccuracy in the original (uncorrected) standard errors is generally of small magnitude, such that inferences from the uncorrected model can still provide useful insights into the potential for non-response bias.

Conclusions

This paper presents a systematic, two-step modeling approach to accommodate risk-related response bias in DCEs addressing climate risk reductions. The approach diagnoses and corrects for response bias attributable to systematic differences in risk exposure between respondents and non-respondents, where indicators of climate-related risks may be extracted from readily available GIS data layers, combined with geocoded home locations of both respondents and non-respondents. A simplified version of this model is then coupled with a semi-parametric bootstrap approach to generate corrected standard errors in the second-stage model.

Results demonstrate the ability of the proposed approach to illuminate otherwise invisible biases in preference and welfare estimates related to risk-related response propensity. Effects on welfare estimates, while intuitive, may not always conform to prior expectations. In contrast to ex ante expectations, present results suggest that individuals who are more exposed to climate-related risks (and presumably have higher levels of awareness and interest in the topic) are less likely to respond to the survey. These findings illustrate a case in which confounding factors (here, trust in the policy process) can lead to patterns in which people who may be less interested in a survey topic may be more likely to respond—a possibility that is generally unacknowledged within the past literature on non-response bias.

We emphasize that all reported empirical results pertain to our case study, and must be viewed within this context. Results, for example, might differ across different intensities and/or types of environmental hazards. Moreover, there may be other causes of non-response bias in DCE results that are not modeled here. We also reiterate that the proposed corrections are reduced form and empirical, in contrast to the formal Heckman corrections available for other types of models. These and other caveats aside, model results highlight the potential of the illustrated approach to detect and correct for otherwise invisible response patterns in environmental valuation results, ameliorating potential biases in welfare analysis. They also illustrate that uncorrected DCE results can lead to misleading inferences regarding WTP for climate-risk reductions—a concern not widely appreciated in the applied valuation literature.

References

Bateman, I.J., Day, B.H., Georgiou, S., and Lake, I. (2006). The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. *Ecological Economics* 60(2): 450-460.

Bliemer, M.C.J., Rose, J.M., and Hensher, D.A. (2009). Constructing efficient stated choice experiments allowing for differences in error variances across subsets of alternatives. *Transportation Research Part B* 43(1): 19–35.

Bockstael, N.E., Strand Jr I.E., McConnell, K.E., and Arsanjani, F. (1990). Sample selection bias in the estimation of recreation demand functions: An application to sportfishing. *Land Economics* 66(1):40-49.

Cameron, T., DeShazo, J.R., and Dennis, M. (2005). Correcting for self-selection bias using the Heckman selection correction in a valuation survey using Knowledge Networks. Presented at the 2005 Annual Meeting of the American Association of Public Opinion Research. Knowledge Networks.

Cameron, T.A, Shaw, W.D, Ragland, S.R, Keef, S., and Callaway, J.M. (1996) Using distance and zip code census information for nonresponse correction in the analysis of mail survey data. UCLA Department of Economics, Working paper # 751.

Cameron, T.A., and DeShazo, J.R. (2005). Comprehensive selectivity assessment for a major consumer panel: Attitudes toward government regulation of environment, health and safety risks. Unpublished paper, Department of Economics, University of Oregon.

Cameron, T.A., and DeShazo, J.R. (2013). Demand for health risk reductions. *Journal of Environmental Economics and Management* 65: 87-109.

Cameron, T.A., Shaw, W.D., Ragland, S.E., Gallaway, J.M., and Keefe, S. (1996). Using actual and contingent behavior data with differing levels of time aggregation to model recreation demand. *Journal of Agricultural and Resource Economics* 21(1): 130-149.

Campbell, D., Scarpa, R., and Hutchinson, W.G. (2008). Assessing the Spatial Dependence of Welfare Estimates Obtained from Discrete Choice Experiments. *Letters in Spatial and Resource Sciences* 1: 117-126.

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

Cuddeback, G., Wilson, E., Orme, J.G., and Combs-Orme, T. (2004). Detecting and statistically correcting sample selection bias. *Journal of Social Service Research* 30 (3): 19-33.

Curtin, R., Presser, S., and Singer, E. (2000). The effects of response rate changes on the Index of Consumer Sentiment. *Public Opinion Quarterly* 64: 413-428.

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficients models. *Transportation* 39(1): 19-31.

Davidson, R., and Mackinnon, J.G. (1999). Bootstrap testing in nonlinear models. *International Economic Review* 40(2):487-580.

de Leeuw, Edith, and Wim De Heer. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In *Survey Nonresponse*, eds. R. Groves, D. Dillman, J. Eltinge, and R. J. A. Little. New York: Wiley.

Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell,J., Berck, J., and Messer, B.L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and internet. *Social Science Research* 38 (1): 1-18.

Dubin, J.A., and Rivers, D. (1989). Selection bias in linear regression, logit and probit models. *Sociological Methods and Research* 18(2 &3): 360-390.

Edwards, S.F., and Anderson, G.D. (1987). Overlooked biases in contingent valuation surveys: some considerations. *Land Economics* 63(2): 168-178.

Ferrini, S., and Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management* 53: 342-363.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3): 393-421.

Greene, W.H. (2003). *Econometric Analysis, 5th Edition*. Upper Saddle River, NJ: Prentice Hall.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *The Public Opinion Quarterly* 70(5): 646-675.

Groves, R.M., and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly* 72(2):167–189.

Guan, W. (2003). From the help desk: Bootstrap standard errors. *The Stata Journal* 3 (1): 71-80.

Hanemann, W. M. (1984). Welfare evaluation in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics* 66(3): 332-341.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1):153-161.

Hensher, D., and Greene, W. (2003). The mixed logit model: The state of practice. *Transportation* 30(2): 133-176.

Hole, A.R. (2007). A Comparison of Approaches to Estimating Confidence Intervals for Willingness to Pay Measures. *Health Economics* 16: 827–840.

Johnston, R.J. (2006). Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum. *Journal of Environmental Economics and Management* 52(1): 469-481.

Johnston, R.J., and Duke, J.M. (2007). Willingness to pay for agricultural land preservation and policy process attributes: Does the method matter? *American Journal of Agricultural Economics* 89(4):1098-1115.

Johnston, R.J., Opaluch, J.J., Mazzotta, M.J., and Magnusson, G. (2005). Who Are Resource Nonusers and What Can They Tell Us About Nonuse Values? Decomposing User and Nonuser Willingness to Pay for Coastal Wetland Restoration. *Water Resources Research* 41(7): W07017.

Johnston, R.J., and Ramachandran, M. (2014). Modeling Spatial Patchiness and Hot Spots in Stated Preference Willingness to Pay. *Environmental and Resource Economics* 59(3): 363-387.

Johnston, R.J., Schultz, E.T., Segerson, K., Besedin, E.Y., and Ramachandran, M. (2012). Enhancing the content validity of stated preference valuation: the structure and function of ecological indicators. *Land Economics* 88(1): 102-120.

Johnston, R.J., Swallow, S.K., Allen, C.W., and Smith, L.A. (2002). Designing multidimensional environmental programs: assessing tradeoffs and substitution in watershed management plans. *Water Resources Research* 38(7):1099-1105.

Johnston, R.J., Swallow, S.K., and Weaver, T.F. (1999). Estimating Willingness to Pay and Resource Trade-offs With Different Payment Mechanisms: An Evaluation of a Funding Guarantee for Watershed Management. *Journal of Environmental Economics and Management* 38(1): 97-120.

Johnston, R.J., Weaver, T.F., Smith, L.A., and Swallow, S.K. (1995). Contingent valuation focus groups: insights from ethnographic interview techniques. *Agricultural and Resource Economics Review* 24 (1): 56-69.

Kaplowitz, M., Frank, L., and John, P.H. (2004). Multiple methods for developing and evaluating a stated-choice questionnaire to value wetlands. In: Methods for testing and evaluating survey questionnaires, Ed. Stanley, P., Jennifer M.R., Mick, P.C., Judith, L.L., Elizabeth, M., Jean, M., and Eleanor, S., New York: Wiley.

King, G., and Roberts, M.E. (2015). How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis* 23:159-179.

Loomis, J.B. (1987). Expanding contingent value sample estimates to aggregate benefit estimates: Current practices and proposed solutions. *Land Economics* 63(4): 396-402.

Messonnier, M.L., Bergstrom, J.C., Cornwell, C.M., Teasley, R.J., and Cordell, H.K. (2000). Survey response-related biases in contingent valuation: Concepts, remedies, and empirical application to valuing aquatic plant management. *American Journal of Agricultural Economics* 82(2): 438-450.

Murphy, K., and Topel, R. (1985). Estimation and inference in two step econometric models. *Journal of Business and Economic Statistics* 3: 370–379.

National Research Council. (2013). Nonresponse in Social Science Surveys: A Research Agenda. Roger Tourangeau and Thomas J. Plewes, Editors. Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.Panel on a Research Agenda for the Future of Social Science Data Collection; Committee on National Statistics, Division on Behavioral and Social Sciences and Education; National Research Council. Nonresponse in social science surveys: a research agenda. Washington, DC: The National Academy Press 2013.

Olsen, S.B. (2009). Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods. *Environmental and Resource Economics* 44:591–610.

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

Rose, J.M., and Bliemer, M.C.J. (2008). Stated preference experimental design strategies. In Handbook of Transport Modelling, Ed. Hensher, D.A. and Button, K.J.,151, 151-180, Elsevier, Oxford, UK.

Sándor, Z., and Wedel, M. (2001). Designing conjoint experiments using manager's prior beliefs. *Journal of Marketing Research* 38: 430-444.

Scarpa, R., and Rose, J.M. (2008). Experimental designs for environmental valuation with choice experiments: Monte Carlo investigation. *Australian Journal of Agricultural and Resource Economics* 52: 253-282.

Schkade, D.A., and Payne, J.W. (1994). How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation. *Journal of Environmental Economics and Management* 26(1): 88-109.

Schouten, B., Cobben, F., and Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology* 35: 101-113.

Skrondal, A., and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society* 172: 659-687.

Town of Old Saybrook (2015). Report of Findings from a Study of the Effects of Sea Level Rise and Climate Change on Old Saybrook, Connecticut. Sea Level Rise Climate Adaptation Committee, Town of Old Saybrook, CT. December.

Train, K. (2009). *Discrete Choice Methods with Simulation, 2nd Edition*. Cambridge University Press, New York.

Whitehead, J.C. (1991) Environmental interest group behavior and self-selection bias in contingent valuation mail surveys. *Growth and Change* 22(1): 10-21.

Whitehead, J.C., Groothuis, P.A., and Blomquist, G.C. (1993). Testing for non-response and sample selection bias in contingent valuation: Analysis of a combination phone/mail survey. *Economics Letters* 41: 215-220.

Table 1. Variables and Descriptive Statistics*

Variable	Description	Mean (Std.Dev)**
Neither	Alternative specific constant (ASC) associated with no new action, or a choice of neither protection plan.	0.33 (0.47)
Hard	Binary (dummy) variable indicating whether the protection plan place more emphasis on hard defenses relative to the omitted condition (similar emphasis on hard and soft defenses).	0.23 (0.42)
Homes	Number of homes expected to flood in a Category 3 storm in the mid-2020s; presented as a percentage of the total number of homes in the town (Range 0-100%).	48.97 (7.27)
Wetlands	Number of acres of wetlands expected to be lost by the mid-2020s due to flooding or erosion; presented as a percentage of the total number of acres of the town coastal marshes (Range 0-100%).	4.71 (2.62)
Beaches	Number of acres of beaches and dunes expected to be lost by the mid-2020s due to flooding or erosion; presented as a percentage of the total acres of the town beaches and dunes that currently exist (Range 0-100%).	9.46 (4.05)
Seawalls	Mileage of the town coast shielded by hard defenses by the mid-2020s; presented as a percentage of the total mileage of the town coastline (Range 0-100%).	24.99 (6.11)
Cost	Household annual cost, presented as unavoidable increase in taxes and fees required to implement the coastal protection plan. A choice of neither protection plan is associated with zero cost (Range 0-$155).	62.16 (55.84)

* Reported statistics are drawn from the realized sample of respondents to the survey version used in random utility modelling.
** Means and standard deviations include status quo option of no new action.

Table 2. Attribute Levels in Choice Experiment Design

Attribute	Levels
Homes	36% of 5034 homes expected to flood in a Category 3 storm.
	43% of 5034 homes expected to flood in a Category 3 storm.
	51% of 5034 homes expected to flood in a Category 3 storm*.
	59% of 5034 homes expected to flood in a Category 3 storm.
Wetlands	2% of 497 wetland acres expected to be lost due to flooding or erosion.
	5% of 497 wetland acres expected to be lost due to flooding or erosion*.
	10% of 497 wetland acres expected to be lost due to flooding or erosion.
Beaches	4% of 30 beach acres expected to be lost due to flooding or erosion.
	10% of 30 beach acres expected to be lost due to flooding or erosion*.
	16% of 30 beach acres expected to be lost due to flooding or erosion.
Seawalls	15% of 50 miles of coast armored.
	24% of 50 miles of coast armored*.
	35% of 50 miles of coast armored.
Cost	$0 (cost to household per year) *.
	$35 (cost to household per year).
	$65 (cost to household per year).
	$95 (cost to household per year).
	$125 (cost to household per year).
	$155 (cost to household per year).

* Status quo value.

Table 3. Summary of Variables in the Response Propensity Model

Variable	Description	Mean / %* (Std. Dev)
Distance	Distance (in feet) of geocoded property centroids to the nearest coastline.	886.80 (793.84)
Elevation	Elevation (in feet) of geocoded property centroids relative to the sea level.	39.41 (35.60)
Flood Zone	Binary variable indicating whether the property centroid lies within U.S. Federal Emergency Management Agency zones A, AE, V, VE or X (default is a centroid not within one of these zones).	85.71% (0.35)
SFHZ	Binary variable indicating whether the property centroid lies within a U.S. Federal Emergency Management Agency Special Flood Hazard Zone, defined as zones A, AE, V or VE (default is a centroid not located in a special flood hazard zone).	17.29% (0.38)
Male	Binary variable indicating whether the subject is a male. For respondents, this is determined by responses to a question requesting the gender of the respondent. For non-respondents, this is inferred based on the individual to which the survey was mailed (default is females and those whose gender is not identifiable or is ambiguous based on the mailing address prefix and name).	52.98% (0.50)

* Summary statistics are based on the entire sample of both respondents and non-respondents (N= 1729).

Table 4. Comparison of Flood Risk Exposure Variables: Respondents versus Non-respondents[a]

Variable	Respondents		Non-respondents		Test statistic (P-value)
	Mean/ %	Std. Dev	Mean/ %	Std. Dev	
Distance	920.13	841.27	876.01	778.89	t = -0.99 (0.32)
Elevation	40.32	37.35	39.11	35.02	t = -0.61 (0.54)
Flood Zone	0.84	-	0.86	-	χ^2 = 1.05 (0.29)
SFHZ	0.17	-	0.17	-	χ^2 = 0.17 (0.87)
Obs. (N)	423		1306		

[a] Comparisons of means are made between respondents and non-respondents to all survey versions (see main text).

Table 5. Response Propensity Model (Binomial Logit)

Variables	Coefficient	Std. Error
Distance	-0.0005	0.0004
Elevation	-0.0064	0.0260
Distance × *Elevation*	**0.0000004***	**0.0000002**
Flood Zone	-0.4845	0.5612
SFHZ	0.1055	0.3955
Male	**0.2115***	**0.1136**
Distance × *Flood zone*	0.0003	0.0004
Distance × *SFHZ*	**0.0008****	**0.0004**
Elevation ×*Flood zone*	0.0028	0.0262
Elevation ×*SFHZ*	-0.0260	0.0186
Constant	-0.6540	0.5451
Observations	1729	
Pseudo R^2	0.008	

**, & * indicate 5%, and 10% significance level respectively.

Table 6. Summary of Predicted Propensity Scores

Sample/Sub-sample	Number of Cases	Mean (%)	Min (%)	Max (%)
Entire sample of respondents and non-respondents	1729	24.46	9.97	44.90
Non-respondents.	1306	24.25	9.97	44.90
Respondents	423	25.14	13.51	41.45
Respondents to the analysis version	142	25.23	17.07	41.45

Table 7. Mixed Logit Results: Preference-Space Model with Response Propensity Interactions

Attributes and Interactions	Restricted		Unrestricted	
	Coefficient (Std. Error)	Std. Dev or Spread (Std. Error)	Coefficient (Std. Error)	Std. Dev or Spread (Std. Error)
Random Parameters:				
Neither	-1.5556***	3.0943***	-7.4029**	3.6339***
	(0.4858)	(0.4692)	(2.9866)	(0.6695)
Homes	-0.1362***	0.1494***	-0.3521**	0.1627***
	(0.0311)	(0.0396)	(0.1790)	(0.0436)
Hard	-1.0311***	1.3255**	-9.4468***	1.3866**
	(0.3709)	(0.6403)	(2.7920)	(0.6615)
Wetlands	-0.0877	0.2333**	-0.7940**	0.2161*
	(0.0549)	(0.0952)	(0.3170)	(0.1208)
Beaches	-0.0998***	0.1854***	-0.1583	0.1861***
	(0.0370)	(0.0602)	(0.2097)	(0.0642)
Cost[b]	0.0159***	0.0159***	0.0167***	0.0167***
	(0.0044)	(0.0044)	(0.0048)	(0.0048)
Non-random Parameters:				
Seawalls	-0.0264		0.0020	
	(0.0255)		(0.1546)	
Interactions:				
Score×*Neither*			0.2291**	
			(0.1116)	
Score×*Homes*			0.0083	
			(0.0067)	
Score×*Hard*			0.3341***	
			(0.1075)	
Score×*Wetlands*			0.0279**	
			(0.0122)	
Score×*Beaches*			0.0020	
			(0.0081)	
Score×*Seawalls*			-0.0014	
			(0.0061)	
χ^2	166.627***		189.703***	
Pseudo R^2	0.186		0.212	
Respondents.(N)	136		136	
Choice responses	408		408	
Observations	1224 (408 questions × 3 options per question)		1224 (408 questions × 3 options per question)	

[a] Except for cost, parameters are specified as random with a normal distribution.
[b] Cost coefficient is specified with bounded triangular distribution with the sign inversed prior to estimation.
***, **, & * indicate 1%, 5%, and 10% significance level respectively.

Table 8. Estimates of WTP Differences Associated with One Percent Change in Response Propensity Score, and Estimates of Mean WTP

Attributes	WTP Difference	Mean WTP		
		Restricted	Unrestricted Respondents	Unrestricted All
Neither	21.39** (0.02)	-140.99*** (<0.01)	-140.79*** (<0.01)	-157.14*** (<0.01)
Homes	0.74 (0.20)	-12.66*** (<0.01)	-12.71*** (<0.01)	-13.28*** (<0.01)
Hard	30.90*** (<0.01)	-99.97*** (<0.01)	-96.15*** (<0.01)	-119.77*** (<0.01)
Wetlands	2.60** (0.02)	-7.43 (0.16)	-7.47 (0.14)	-9.45* (0.06)
Beaches	0.16 (0.86)	-9.05*** (<0.01)	-9.57*** (<0.01)	-9.69*** (<0.01)
Seawalls	-0.14 (0.80)	-2.42 (0.32)	-3.02 (0.24)	-2.91 (0.28)

***, **, and * indicate 1%, 5%, and 10% significance levels, respectively. *P*-values in parentheses.

Table 9. Semi-Parametric Bootstrap Results: Conditional Logit Model

Attributes and Interactions	Coefficient	Uncorrected Standard Errors	Bootstrap Standard Errors	
			1000 Samples	10,000 Samples
Neither	-0.8614	0.7999	0.8810	0.8695
Homes	-0.0991	0.0763	0.0710	0.0730
Hard	-4.9006***	1.2407	1.6759***	1.6701***
Wetlands	-0.4406***	0.1565	0.1808**	0.1805**
Beaches	-0.0575	0.1059	0.1252	0.1243
Seawalls	0.0621	0.0877	0.0947	0.0967
Cost	-0.0077***	0.0022	0.0020***	0.0020***
Interactive terms:				
Score×Neither	0.0117	0.0309	0.0347	0.0344
Score×Hard	0.1694	0.0483***	0.0670**	0.0666**
Score×Homes	0.0008	0.0030	0.0030	0.0028
Score×Wetlands	0.0159	0.0060***	0.0071**	0.0071**
Score×Beaches	0.0002	0.0041	0.0050	0.0050
Score×Seawalls	-0.0030	0.0035	0.0038	0.0039
χ^2	76.56***			
Pseudo R^2	0.085			
Respondents(N)	136			
Choice responses	408			
Observations	1224			

***, **, and * indicate 1%, 5%, and 10% significance levels, respectively.

PLEASE VOTE

Question 4.

PROTECTION OPTION A and **PROTECTION OPTION B** are possible protection options for Old Saybrook. **NO NEW ACTION** shows what is expected to occur with no additional protection. All plans involve hard and soft defenses in different areas. Given a choice between the three, how would you vote?

Methods and Effects of Protection	Result in 2020s with NO NEW ACTION	Result in 2020s with PROTECTION OPTION A	Result in 2020s with PROTECTION OPTION B
	No Change in Existing Defenses	More Emphasis on HARD Defenses	More Emphasis on HARD Defenses
Homes Flooded	**51%** 2,585 of 5,034 homes expected to flood in a Category 3 storm	**51%** 2,585 of 5,034 homes expected to flood in a Category 3 storm	**51%** 2,585 of 5,034 homes expected to flood in a Category 3 storm
Wetlands Lost	**5%** 25 of 497 wetland acres expected to be lost	**5%** 25 of 497 wetland acres expected to be lost	**2%** 10 of 497 wetland acres expected to be lost
Beaches and Dunes Lost	**10%** 3 of 30 beach acres expected to be lost	**4%** 1 of 30 beach acres expected to be lost	**4%** 1 of 30 beach acres expected to be lost
Seawalls and Coastal Armoring	**24%** 12 of 50 miles of coast armored	**24%** 12 of 50 miles of coast armored	**24%** 12 of 50 miles of coast armored
$ Cost to Your Household per Year	**$0** Increase in annual taxes or fees	**$95** Increase in annual taxes or fees	**$125** Increase in annual taxes or fees
HOW WOULD YOU VOTE? (CHOOSE ONLY ONE) I vote for	☐ I vote for **NO NEW ACTION**	☐ I vote for **PROTECTION OPTION A**	☐ I vote for **PROTECTION OPTION B**

Figure 1. Sample Choice Question

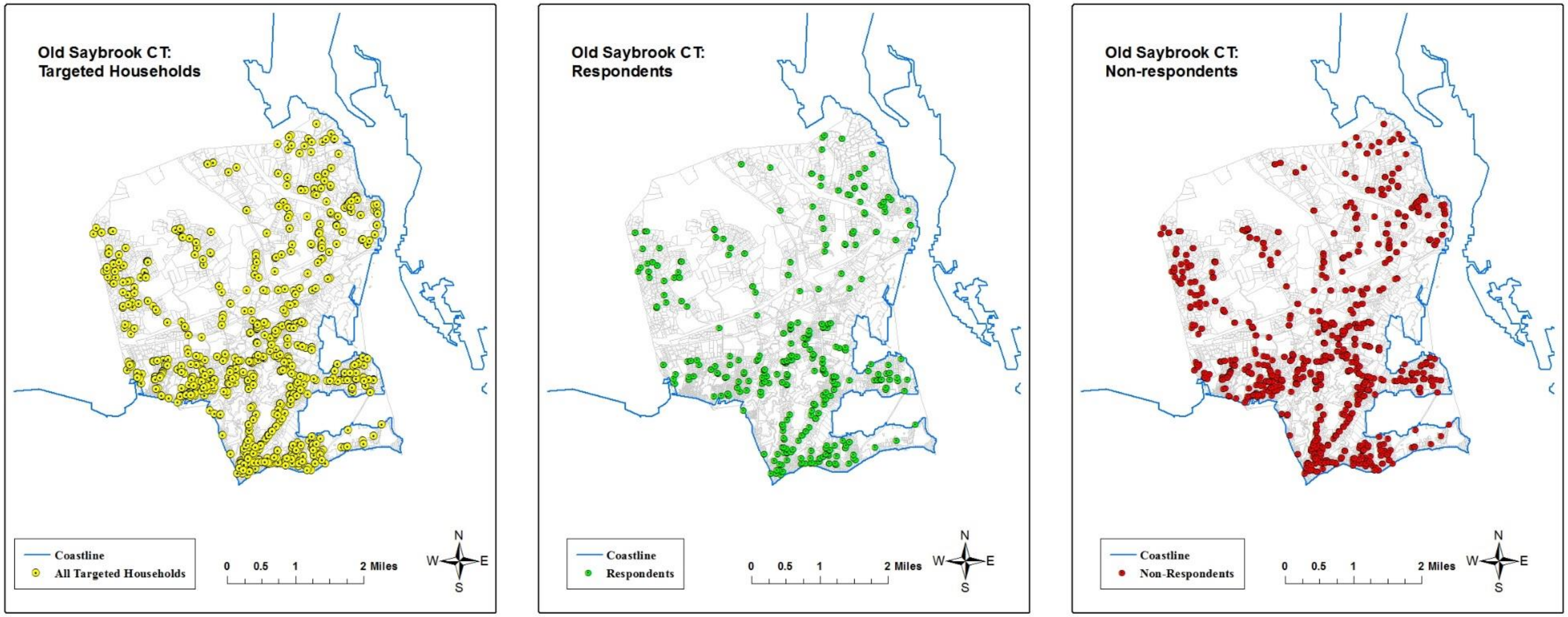

Figure 2: Spatial Distribution of Respondents and Non-respondents Based on Geocoded Locations of Homes