What are the consequences of consequentiality?

Joseph Herriges [a,*], Catherine Kling [a], Chih-Chen Liu [b], Justin Tobias [c]

[a] Iowa State University, USA
[b] National University of Kaohsiung, Taiwan, ROC
[c] Purdue University, USA

ARTICLE INFO

ABSTRACT

We investigate the extent to which dichotomous choice referenda responses are shaped by whether the individual believes the survey itself will ultimately impact policy. Using survey data from the Iowa Lakes Project, we test this supposition. Specifically, we employ a Bayesian treatment effect model in which the degree of perceived consequentiality, measured as an ordinal response, is permitted to have a structural impact on willingness to pay (WTP) for a hypothetical environmental improvement. We test whether the estimated WTP distributions are the same for each value of the ordinal response.

In our survey data, a subsample of individuals were randomly assigned supporting information suggesting that their responses to the questionnaires were important and will have an impact on policy decisions. In conjunction with a Bayesian posterior simulator, we use this source of exogenous variation to identify the structural impacts of consequentiality perceptions on willingness to pay, while controlling for the potential of confounding on unobservables. We find evidence consistent with a "knife-edge" result, namely that the willingness to pay distributions are equal among those believing the survey to be at least minimally consequential, and different for those believing that the survey is irrelevant for policy purposes.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Despite lingering concerns within the economics literature, stated preferences methods for valuing non-market commodities are frequently used both by policymakers and in the judicial system to assess changes to environmental goods and services. Best practices, starting with the NOAA panel recommendations, have evolved over time aimed at minimizing potential biases in survey responses. These guidelines emphasize the careful design and pre-testing of the survey instrument so as "... to induce respondents to take the [survey] question[s] seriously" [1, p. 4606]. In this context, it is surprising that relatively little attention has been given to assessing whether respondents actually believe that the survey itself will potentially influence future policy. If they do not believe this to be the case, they have little incentive to respond truthfully and their answers may be largely noise. Indeed, Carson and Groves (CG) [2] argue that individuals can be expected to answer the standard dichotomous choice referendum question truthfully *if* two conditions hold. First, the respondent must believe that the results of the survey might influence an outcome they care about, a condition we refer to as *policy consequentiality*. Second, the respondent must perceive that there is some probability that they will have to pay, or *payment consequentiality*. Jointly, these two conditions (*strong consequentiality*) yield a knife-edge implication; i.e., as long

* Corresponding author. Fax: +1 515 294 0221.
  E-mail address: jaherrig@iastate.edu (J. Herriges).

as the respondent believes that their answers will be consequential in both senses with *any* positive probability, their dominant strategy is to answer truthfully. If accurate, this result has clear and important implications regarding how a researcher should handle survey responses: all data arising from respondents who believe the survey is at least minimally consequential can be assumed to provide truthful answers to survey questions.[1]

In this paper, we focus our attention on the role that *policy consequentiality* has on the responses to a dichotomous choice referendum question and the efficacy of an information treatment designed to increase the perceived degree of consequentiality.[2,3] Specifically, we use respondents' perceptions of consequentiality elicited in the 2005 Iowa Lakes Survey to determine whether individuals have different perceptions concerning the degree of consequentiality of the valuation exercise and whether these perceptions affects respondents' willingness to pay (WTP).[4] In the survey, respondents were asked whether they would vote in favor of a referendum to improve water quality at a lake where bid values were varied across the sample. Respondents were also asked to answer, on a scale from 1 to 5, how likely it was that the survey results would influence decisions in the state concerning water quality programs. Thus, a measure of the degree to which respondents perceived the survey as *policy consequential* was directly elicited. Based on the CG arguments, respondents who do not believe that the survey is consequential could be omitted from the sample for estimation purposes. Additionally, the distributions of WTP from respondents with differing views concerning the degree of consequentiality could be tested for equality.

To extract accurate estimates of the impact of consequentiality perceptions on WTP, it is important to recognize that respondents who indicate a high degree of consequentiality may do so because they also place a high value in the proposed water quality improvement project. In other words, there is a potential endogeneity, or unobserved confounding problem. To address this concern, a split sample treatment was administered in the survey. Specifically, half of the sample was provided with a highlighted article from the *Iowa Conservationist*—the magazine of the Iowa Department of Natural Resources (IDNR), the state agency with primary responsibility for water quality control—indicating that IDNR was already using results from the survey in their policy decisions and planned to continue to do so. Our assumption, which is borne out empirically, is that the presentation of this information will positively affect the respondents' perceived degree of consequentiality. This exogenous treatment aids us in estimating the "causal" impacts of consequentiality perceptions on WTP, as we will describe below.

Making use of the information treatment in the survey, we can then explore the impact that perceived *policy consequentiality* has on willingness to pay within the framework of a standard triangular treatment-response model. We proceed using a Bayesian approach and derive and employ a new algorithm that improves upon standard estimation methods, and can be applied by other practitioners seeking to fit models with a similar structure. Specifically, since our consequentiality responses are ordinal, our model must contend with the estimation of cutpoint values, and it is well-documented in the literature that standard Gibbs sampling schemes in such models can suffer from very poor mixing, particularly in moderately large data sets, thus producing imprecise and potentially inaccurate posterior inference (e.g., [3,4]).[5] Our proposed posterior simulator offers significant improvements by sampling the cutpoints, latent willingness to pay and latent consequentiality variables in a single step rather than sampling each component from its corresponding complete posterior conditional distribution.

Finally, before proceeding, it is important to clarify what implications can be drawn from the analysis below. Our survey gathers information as to whether respondents believe the results of the survey "... will affect decisions about water quality in Iowa Lakes." This is, at a minimum, a measure of the perceived *policy conquentiality* of the survey. As such, our econometric analysis identifies the extent to which responses to a dichotomous choice CV question differ as these perceptions differ. Drawing additional implications become difficult. *If* respondents also assume that the statement carries with it an implied obligation to pay, then we have a measure of *strong consequentiality* and Carson and Grove's theoretical results would suggest that all respondents with a perceived consequentiality of 2 or greater would respond truthfully. This is consistent with our findings below that WTP is only different for those individuals who perceive the survey to be completely inconsequential. However, as the survey does not directly ask individuals whether they perceive the survey carries with it a potential obligation to pay, the results are open to interpretation. As one reviewer suggests, *if* no one

---

[1] It is worth noting that the CG "theory" is itself not without controversy. As one reviewer noted, some see a weakness of the CG work in that it fails to predict the direction of bias in an "inconsequential" survey. In practice, hypothetical surveys often generate WTP values that exceed those generated in a real market, suggesting that the hypothetical bias is positive. Since the CG story does not provide a basis for this prediction, it might be argued that the theory is incomplete. While these are important issues, we make no attempt to address them here, rather leaving them for others to ponder.

[2] Throughout the remainder of this paper, we use the term "consequentiality" to refer to the notion of "policy consequentiality," as defined above. When it is necessary to distinguish policy consequentiality from either payment or strong consequentiality, we use the additional qualifier.

[3] Interestingly, Landry and List [10] suggest that surveys could elicit the perceived degree of consequentiality directly from respondents which is precisely the approach we adopted when designing this instrument several years ago.

[4] In practice we must make some assumptions regarding the reliability of our data. Specifically, we do not investigate potential issues of misreporting in the consequentiality reports, but treat these ordinal responses as accurate representations of agents' beliefs. One may argue that the CG model is unable to characterize the validity of WTP responses for those in the inconsequential group, and thus the validity of responses to the consequentiality question itself may also be uncertain. We do not take up this issue in the current paper, but instead, proceed conditionally on the observed consequentiality responses.

[5] "Poor mixing" in simulation based approaches, such as Gibbs sampling, refers to the problem of the algorithm yielding high serial correlation among the draws from the distribution of interest. This high correlation, in turn, implies that a large number of simulation draws are required to accurately estimate characteristics of the distribution, such as its mean, mode or variance.

believes that they will have to pay, then those perceiving the survey to be consequential have an incentive to respond strategically, indicating a higher willingness to pay assuming that they will never have to pay. We leave it to the reader to draw their own conclusions.

2. Related literature

There have been several studies to date testing the impact of "consequentiality" on respondents' preference revelation. These have largely been carried out through laboratory or field experiments (as discussed below, Bulte et al. [5] provide an exception investigating consequentiality within a stated preference framework).[6] In these laboratory settings, "consequentiality" typically takes the form of "strong consequentiality." The outcome of the experiment (i.e., both the provision of and payment for the good begin considered) is binding with a probability set by the researcher. Cummings et al. [7], for example, conducted an experiment involving real and purely hypothetical referenda to compare how respondents behave in these two settings. They find that respondents are more likely to vote "yes" in the hypothetical setting than in the real setting and conclude that hypothetical referenda yield biased estimates of WTP. Cummings and Taylor [8] explore this issue further by investigating how the hypothetical bias of the dichotomous choice referendum varies with the degree of consequentiality. Specifically, their laboratory experiment employs a probabilistic treatment that varies the odds that the referendum will be binding. They find that participants in different treatments, including hypothetical, probabilistic, and real referendum, behave significantly different from one another. Specifically, the probability of a respondent voting "yes" falls if the probability of the referendum being binding rises. They also find that respondents' voting behavior is significantly different from an actual referendum unless there is high probability that the referendum will be binding (e.g., greater than 50 percent). The latter finding is not consistent with the knife-edge result of CG.

A second study within a laboratory setting is reported by Shogren and Tadevosyan [9]. They report that in an induced-value second price auction, participants faced with a consequential treatment (a positive probability of the auction being binding) report an average willingness to pay (and standard deviation) of 4.35 (6.60) compared to 4.16 (2.91) for their real treatment. Despite the apparent overlap of these distributions based on the reported standard deviations, their regression analysis rejects equivalence of the means and they conclude that "... bidding behavior in the consequential auction was neither optimal nor identical to real auction bidding."

Carson et al. [11] conduct a study within the context of a field experiment. In the experiment, participants were informed the probability of the dichotomous choice referendum being binding. In contrast to Cummings and Taylor [8], they find the "knife-edge" theoretical results suggested in CG [2]: as long as the probability of consequentiality exceeds zero even by a small amount, participants respond in the same manner as in an actual referendum. They suggest that results from inconsequential (hypothetical) referendum should not be used to make inference about how CV works in consequential referendum.

Landry and List (2007) also undertake a field experiment, in this case with four treatments: Hypothetical, Cheap Talk, Consequential, and Real. While they find that WTP elicited within the purely hypothetical treatment differs statistically from those in the real treatment, both the Consequential and Cheap Talk designs yield responses that are consistent with real responses. While their findings are consistent with the CG arguments, since they used a single degree of consequentiality in their design (a 50 percent chance that the referendum would be binding) their results do not provide insight on whether lower degrees of consequentiality would generate the same findings.

Though quite useful, these studies largely take place outside of the context of an actual CV exercise and require the analyst to inform participants of the degree of consequentiality associated with the exercise. These laboratory and field experiments have the important advantage that the researcher can directly control the degree of consequentiality, but their disadvantage is that this direct control is not typical of CV surveys. Thus, findings from the lab or field may be difficult to transfer to the survey arena. Within a survey setting, there is only one study we are aware of that directly investigates the consequentiality issue. Bulte et al. [5] explore the impact of "policy consequentiality" by telling respondents that the results of the survey will be made available to policy makers, though they do not directly elicit respondents' reactions to this information. Their WTP estimates obtained from the survey with "cheap talk" and the survey with a consequentiality device are significantly smaller than those obtained in a purely hypothetical survey. In addition, the WTP estimates obtained from the survey with cheap talk are not significantly different from the WTP's obtained from the survey with a consequential device. Their results are also consistent with the CG predictions.

In Bulte et al. [5], survey respondents were directly informed that the findings from the survey will be made available to policy makers. In our study, we take this approach a significant step further: we show survey respondents direct statements from a policy maker indicating that the findings from the survey instrument *will* be used to direct decisions. Moreover, we seek to determine whether (and to what extent) the survey participants are swayed by our arguments.

---

[6] Field experiments are distinct from laboratory experiments in that they involve engaging subjects in real world settings, such as sports card trade shows, rather than in the laboratory. Harrison and List [6] provide an overview of field experiments.

3. The empirical model

To test for the potential existence of differential impacts of perceived consequentiality on willingness to pay, we consider the following two equation system[7]:

$$c_i^* = \mathbf{x}_{ci}\boldsymbol{\beta}_c + \varepsilon_c, \tag{1}$$

$$w_i^* = \mathbf{x}_{wi}\boldsymbol{\beta}_w + \overline{\mathbf{c}}_i\boldsymbol{\delta}_c + \varepsilon_w, \tag{2}$$

where

$$\begin{bmatrix} \varepsilon_i \\ u_i \end{bmatrix} \Bigg| \mathbf{x}_{ci}, \mathbf{x}_{wi} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{cw} \\ \sigma_{cw} & \sigma_w^2, \end{pmatrix} \right].$$

Eq. (1) corresponds to the latent "consequentiality" equation, while Eq. (2) corresponds to the latent "willingness to pay" equation. The latent c_i^* and w_i^* are not observed, but their values generate an observed binary willingness to pay indicator w_i and reported consequentiality variable c_i. In our data, c_i takes on five different ordered values, ranging from regarding the survey as being completely irrelevant to the making of policy ($c_i = 1$) to regarding the survey as having a definite impact on policy ($c_i = 5$). We thus relate the observables w and c to their latent values as follows:

$$w_i = \begin{cases} 1 & \text{if } w_i^* \geq B_i, \\ 0 & \text{if } w_i^* < B_i, \end{cases} \quad c_i = \begin{cases} 1 & \text{if } -\infty < c_i^* \leq 0, \\ 2 & \text{if } 0 < c_i^* \leq \alpha_3, \\ 3 & \text{if } \alpha_3 < c_i^* \leq \alpha_4, \\ 4 & \text{if } \alpha_4 < c_i^* \leq \alpha_5, \\ 5 & \text{if } \alpha_5 < c_i^* < \infty, \end{cases}$$

where B_i is a hypothetical price (bid) proposed to individual i and $\overline{\mathbf{c}}_i$ in (2) is a 1×5 vector with a one in the c_ith column and zeros elsewhere. (Note, with this formulation of the model, an intercept must be excluded from $\mathbf{x_{wi}}$). The parameter vector $\boldsymbol{\delta_c}$ (within the treatment-response framework of (1) and (2) and with a valid exclusion restriction or instrument in $\mathbf{x_c}$) is commonly interpreted as capturing the "causal" impact of consequentiality categories on WTP. We similarly denote δ_{cj} as the jth element of $\boldsymbol{\delta_c}$, the parameter associated with the event that $c_i = j$.

It is also worth noting that, although the observed w_i responses are binary, and (2) represents a latent variable equation that generates this binary response, we do not restrict the variance of (2) to unity. Variation in the bids B_i across individuals enables identification of σ_w^2 and thereby enables us to estimate the entire WTP distribution. The consequentiality equation in (1), however, is a "standard" ordered choice equation, whence we restrict the variance of (1) to unity for identification purposes.

In practice, a variety of methods can be employed to fit this model. We have experimented with some traditional alternatives, such as simulated maximum likelihood, and found it difficult to get successful convergence and good performance, even under ideal conditions with correct model specification. To this end, we choose to employ a Bayesian estimation approach with data augmentation (e.g., [12,13, Chapter 14]), which is quite similar in spirit to classical likelihood-based techniques, and is indistinguishable from these alternatives in large sample sizes when suitably "diffuse" or "non-informative" priors are employed.

Simply adopting the Bayesian methodology, however, is not a guarantee of good performance, and indeed, some important complications arise concerning how best to estimate the parameters of our equation system under this approach. With typical conjugate priors,[8] for example, fitting the model in (1) and (2) using Gibbs sampling (i.e., an algorithm that iteratively samples from the complete posterior conditionals for each set of parameters) is reasonably straightforward, following the important work of Albert and Chib [14]. However, unless care is taken in the development and application of these samplers, slow mixing and unreliable inference can occur. Thus, unless something is done to improve the "baseline" Gibbs algorithm that samples from each parameter's conditional posterior distribution, use of Bayesian techniques offers little or no improvement over these other alternatives. To this end, we offer a refinement of such "off-the-shelf" simulation methods which offers substantial performance improvements.

First, we introduce a reparameterization, following the suggestion of Nandram and Chen [4] in the context of a single-equation ordered probit model. Specifically, we let $\gamma = \alpha_5^{-1}$ and define $\widetilde{x} \equiv \gamma x$. [That is, the $\widetilde{}$ notation simply denotes the operation of taking the original variable and multiplying it by γ.] Multiplying (1) by γ and adjusting the rule mapping the

---

[7] As pointed out by a referee, it is also possible that w^* in (2) should appear as right-hand side variable in the structural representation of c^* in (1). In this case, the parameters in (1) are interpreted as reduced-form and no type of structural impact should be ascribed to these effects. For purposes of the present study, the key parameters are those appearing in Eq. (2) and those in (1) are, with the notable exception of the instrument's coefficient, not of primary interest. Of course, the parameters of (2) are both identifiable and consistently estimable provided an instrument is available in (1). Finally, we adopt the convention throughout this paper of using boldface to denote vector or matrix quantities and capital letters to denote matrices.

[8] A (conditionally) *conjugate* prior is a prior that yields a (conditional) posterior distribution of the same functional form.

latent c^* into the observed c produces the following equivalent model:

$$\widetilde{c}_i^* = \mathbf{x}_{ci}\widetilde{\boldsymbol{\beta}}_c + \widetilde{\varepsilon}_{ci}, \tag{3}$$

$$w_i^* = \mathbf{x}_{wi}\boldsymbol{\beta}_w + \overline{\mathbf{c}}_i\boldsymbol{\delta}_c + \varepsilon_{wi}, \tag{4}$$

where

$$\begin{bmatrix} \widetilde{\varepsilon}_{ci} \\ \varepsilon_{wi} \end{bmatrix} \Big| \mathbf{x}_{ci}, \mathbf{x}_{wi} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma^2 & \gamma\sigma_{cw} \\ \gamma\sigma_{cw} & \sigma_u^2 \end{pmatrix} \right],$$

$$w_i = \begin{cases} 1 & \text{if } w_i^* \geq B_i \\ 0 & \text{if } w_i^* < B_i \end{cases} \quad \text{and} \quad c_i = \begin{cases} 1 & \text{if } -\infty < \widetilde{c}_i^* \leq 0, \\ 2 & \text{if } 0 < \widetilde{c}_i^* \leq \widetilde{\alpha}_3, \\ 3 & \text{if } \widetilde{\alpha}_3 < \widetilde{c}_i^* \leq \widetilde{\alpha}_4, \\ 4 & \text{if } \widetilde{\alpha}_4 < \widetilde{c}_i^* \leq 1, \\ 3 & \text{if } 1 < \widetilde{c}_i^* < \infty. \end{cases}$$

Let

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \widetilde{\sigma}_c^2 & \widetilde{\sigma}_{cw} \\ \widetilde{\sigma}_{cw} & \sigma_w^2 \end{pmatrix} = \begin{pmatrix} \gamma^2 & \gamma\sigma_{cw} \\ \gamma\sigma_{cw} & \sigma_w^2 \end{pmatrix}.$$

We choose to work with the parameters[9]:

$$\widetilde{\boldsymbol{\beta}} = [\widetilde{\boldsymbol{\beta}}_c' \ \ \boldsymbol{\beta}_w' \ \ \boldsymbol{\delta}_c']' \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}}^{-1}$$

rather than the original parameterization of the model. The primary reasons for doing this are threefold. First, the reparameterization helps to improve the mixing of the posterior simulations (e.g., [4]), as freeing up the variance parameter rather restricting it to unity tends to reduce the lagged autocorrelations among the parameter draws. Second, the reparameterization eliminates one component of the cutpoint vector (which is important in our case since $\widetilde{\boldsymbol{\alpha}}$ is two-dimensional, and routines for calculating bivariate, but not trivariate, normal probabilities are readily available).[10] Finally, the reparameterization eliminates the diagonal restriction on the 2×2 covariance matrix, enabling the use of a Wishart prior on $\widetilde{\boldsymbol{\Sigma}}^{-1}$ and a standard Gibbs step for sampling its elements.[11]

Since the structural parameters $\boldsymbol{\beta}_c$, σ_{cw}, σ_w^2 and $\boldsymbol{\alpha} = [\alpha_3 \ \ \alpha_4 \ \ \alpha_5]'$ (rather than their transformations) are ultimately of interest, we can, of course, recover these by using the appropriate inverse transformations at each iteration of the sampler:

$$\boldsymbol{\beta}_c = \widetilde{\boldsymbol{\beta}}_c/\gamma, \quad \sigma_{\varepsilon u} = \widetilde{\sigma}_{\varepsilon u}/\gamma, \quad \boldsymbol{\alpha} = [(\widetilde{\alpha}_3/\gamma) \ \ (\widetilde{\alpha}_4/\gamma) \ \ \gamma^{-1}]'.$$

3.1. Priors and the joint posterior

As fully described in Appendix A, we fit this model using Gibbs sampling coupled with the Metropolis-Hastings algorithm and auxiliary variable Gibbs (e.g., [17]). Before talking about the specifics of this algorithm, we must first derive our *augmented posterior* distribution. This augmented posterior involves adding the latent $\widetilde{\mathbf{c}}^*$ and $\mathbf{w}^*$ to the joint posterior distribution. Under prior independence this joint distribution can be represented as

$$p(\widetilde{\mathbf{c}}^*, \mathbf{w}^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\alpha}}|\mathbf{c}, \mathbf{w}) \propto p(\mathbf{c}, \mathbf{w}|\widetilde{\mathbf{c}}^*, \mathbf{w}^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\alpha}})p(\widetilde{\mathbf{c}}^*, \mathbf{w}^*|\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\alpha}})p(\widetilde{\boldsymbol{\beta}})p(\widetilde{\boldsymbol{\Sigma}})p(\boldsymbol{\alpha})$$

$$= p(\boldsymbol{\alpha})p(\widetilde{\boldsymbol{\beta}})p(\widetilde{\boldsymbol{\Sigma}})\prod_{i=1}^n p(\widetilde{c}_i, w_i|\widetilde{c}_i^*, w_i^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha})p(\widetilde{c}_i^*, w_i^*|\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}), \tag{5}$$

where the product term follows from the assumed (conditional) independence across observations. For the first term on the right-hand side of this product, note

$$p(c_i, w_i|\widetilde{c}_i^*, w_i^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\alpha}}) = p(c_i|w_i, \widetilde{c}_i^*, w_i^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\alpha}})p(w_i|\widetilde{c}_i^*, w_i^*, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}}) = p(c_i|\widetilde{c}_i^*, \widetilde{\boldsymbol{\alpha}})p(w_i|w_i^*),$$

---

[9] We do not, however, take up the issue of the induced priors on the original structural parameters.

[10] Appendix A provides details regarding the specifics of the required calculations for sampling the transformed cutpoints. All of the appendices for this paper are available at JEEM's online archive of supplementary material and can be accessed at http://www.aere.org/journals/.

[11] This latter point should probably be regarded as a minor contribution in light of the fact that one can simply write, for example, $\varepsilon_{wi} = \sigma_{cw}\varepsilon_{ci} + u_i$, with u_i independent of ε_{ci}, substitute this into Eq. (2), and then develop a posterior simulator. See, e.g., [15,16] for examples of this strategy.

where

$$p(c_i|\widetilde{c}_i^*, \widetilde{\boldsymbol{\alpha}}) = I(\widetilde{\alpha}_{c_i} < \widetilde{c}_i^* \leq \widetilde{\alpha}_{c_i+1}) \tag{6}$$

with $I(\cdot)$ denoting the standard indicator function, and

$$p(w_i|w_i^*) = I(w_i = 0)I(w_i^* < B_i) + I(w_i = 1)I(w_i^* \geq B_i). \tag{7}$$

Apart from the priors, the remaining piece in our joint posterior in (5) follows immediately from our normality assumption:

$$\widetilde{c}_i^*, w_i^* | \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Sigma}} \overset{ind}{\sim} N\left(\begin{bmatrix} \mathbf{x}_{ci}\widetilde{\boldsymbol{\beta}}_c \\ \mathbf{x}_{wi}\boldsymbol{\beta}_w + \overline{\mathbf{c}}_i\boldsymbol{\delta}_c \end{bmatrix}, \widetilde{\boldsymbol{\Sigma}} \right).$$

The model is completed by choosing priors of the forms:

$$\widetilde{\boldsymbol{\beta}} \sim N(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta), \tag{8}$$

$$\widetilde{\boldsymbol{\Sigma}}^{-1} \sim W(\mathbf{R}, v), \tag{9}$$

$$p(\widetilde{\boldsymbol{\alpha}}) = p(\widetilde{\alpha}_3)p(\widetilde{\alpha}_4|\widetilde{\alpha}_3) = I(0 < \widetilde{\alpha}_3 < 1)\frac{1}{1 - \widetilde{\alpha}_3}I(\widetilde{\alpha}_3 < \widetilde{\alpha}_4 < 1). \tag{10}$$

For the last of these three priors, we impose the ordering restriction on the cutpoints. Unconditionally, we specify a prior for the smallest transformed cutpoint, $\widetilde{\alpha}_3$ which is uniform over its support, while $\widetilde{\alpha}_4$ is specified to be conditionally uniform over $(\widetilde{\alpha}_3, 1)$.

A second contribution of our algorithm is that we utilize a blocking step to sample a group of parameters jointly, rather than conditionally, thus significantly improving the mixing of our posterior simulations. That is, our algorithm employs a blocking procedure wherein the 2×1 vector of transformed cutpoints, $\widetilde{\boldsymbol{\alpha}}$ and latent data $\widetilde{\mathbf{c}}^*$, $\mathbf{w}^*$ are sampled together in a *single step* rather than sampled from their respective conditional posterior distributions. The clear need for such an algorithm was motivated by generated data experiments suggesting poor mixing of the simulated cutpoints when separately sampling from each parameter's conditional posterior distribution.[12] In cases where objects of interest are functions of the cutpoints, such as the WTP distribution described below, the improved mixing offered by our algorithm would seem to offer a substantial benefit.

3.2. The WTP predictive

As we noted in the Introduction, a primary focus of our paper involves the calculation and comparison of WTP distributions for individuals with varying degrees of perceived consequentiality. To this end, we consider a posterior predictive exercise and seek to obtain the WTP distribution for a hypothetical out-of-sample agent with observed characteristics $\mathbf{x}_{cf}$ and $\mathbf{x}_{wf}$ and consequentiality response $c_f = j$. The model in (3) and (4), assumed to hold for this agent's outcomes, implies that the posterior predictive willingness to pay distribution for this representative agent, conditioned on the model parameters $\boldsymbol{\Gamma}$, is given as

$$p(w_f^*|c_f = j, \boldsymbol{\Gamma}) = p(w_f^*|\widetilde{\alpha}_j < \widetilde{c}_f^* \leq \widetilde{\alpha}_{j+1}, \boldsymbol{\Gamma}) = [\Pr(\widetilde{\alpha}_j < \widetilde{c}_f^* \leq \widetilde{\alpha}_{j+1})|\boldsymbol{\Gamma}]^{-1} \int_{\widetilde{\alpha}_j}^{\widetilde{\alpha}_{j+1}} p(w_f^*, \widetilde{c}_f^*|\boldsymbol{\Gamma}) d\widetilde{c}_f^*.$$

Factoring the latter distribution into the conditional $\widetilde{c}_f^*|w_f^*$ and the marginal for w_f^* enables evaluation of the integral. Specifically, we obtain[13]

$$p(w_f^*|c_f = j, \boldsymbol{\Gamma}) = \left[\Phi\left(\frac{\widetilde{\alpha}_{j+1} - \mathbf{x}_{cf}\boldsymbol{\beta}_c - [\widetilde{\sigma}_{cw}/\sigma_w^2](w_f^* - \mathbf{x}_{wf}\boldsymbol{\beta}_w - \delta_{cj})}{\widetilde{\sigma}_c\sqrt{1 - \widetilde{\rho}_{cw}^2}} \right) \right.$$
$$\left. - \Phi\left(\frac{\widetilde{\alpha}_j - \mathbf{x}_{cf}\boldsymbol{\beta}_c - [\widetilde{\sigma}_{cw}/\sigma_w^2](w_f^* - \mathbf{x}_{wf}\boldsymbol{\beta}_w - \delta_{cj})}{\widetilde{\sigma}_c\sqrt{1 - \widetilde{\rho}_{cw}^2}} \right) \right] \frac{\phi(w_f^*; \mathbf{x}_{wf}\boldsymbol{\beta}_c + \delta_{cj}, \sigma_w^2)}{\Phi\left(\frac{\widetilde{\alpha}_{j+1} - \mathbf{x}_{cf}\boldsymbol{\beta}_c}{\widetilde{\sigma}_c} \right) - \Phi\left(\frac{\widetilde{\alpha}_j - \mathbf{x}_{cf}\boldsymbol{\beta}_c}{\widetilde{\sigma}_c} \right)}. \tag{11}$$

In practice, we would like to calculate a variety of statistics associated with this willingness to pay distribution. For example, quantities like the posterior mean, posterior standard deviation, and posterior median (which is critical for

---

[12] Again, we refer here to the case where no blocking steps are employed, and the cutpoints are drawn element-by-element from their (uniform) posterior conditional distributions.

[13] Note that this specification reduces to the marginal for w_f^* when ρ_{cw} (and thus $\widetilde{\sigma}_{cw}$) equals zero.

determining whether or not a particular referendum is likely to pass under majority vote) are certainly of interest and are policy relevant.

Since these quantities are not easily calculated analytically, given the rather messy form of (11), it seems desirable to pursue a numerical alternative by obtaining draws from (11) and using these draws to calculate any statistic of interest. Unfortunately, (11) is not of an immediately recognizable functional form, thus calling into question the feasibility of this numerical scheme. However, it can be shown (a proof is available upon request) that a draw from (11) can be obtained by the following procedure:

First, sample

$$z_f^* \sim TN_{(\underline{z}_f, \overline{z}_f)}(0, 1)$$

where $TN_{(a,b)}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated to lie in the interval (a, b),

$$\underline{z}_f = \frac{\widetilde{\alpha}_j - \mathbf{x_{cf}}\boldsymbol{\beta_c}}{\widetilde{\sigma}_c} \quad \text{and} \quad \overline{z}_f = \frac{\widetilde{\alpha}_{j+1} - \mathbf{x_{cf}}\boldsymbol{\beta_c}}{\widetilde{\sigma}_c}.$$

Then, set

$$w_f^* = \pi_{0f} + \pi_1 z_f^* + \pi_2 \varepsilon,$$

where

$$\varepsilon \sim N(0, 1), \quad \pi_{0f} = \mathbf{x_{wf}}\boldsymbol{\beta_c} + \delta_{cj}, \quad \pi_1 = \sigma_w \widetilde{\rho}_{cw} \quad \text{and} \quad \pi_2 = \sigma_w \sqrt{1 - \widetilde{\rho}_{cw}^2}$$

as a draw from (11).[14] The above procedure provides a simple to implement series of steps (i.e., they only require drawing from standard distributions) for sampling from the WTP distribution directly, whence any feature of this distribution can be calculated. In practice, the steps above are repeated for each post-convergence simulation $\boldsymbol{\Gamma}$, and the collection of the resulting draws are then used to calculate the desired quantity or quantities. To our knowledge, a description of such a method in the presence of an endogeneity problem offers a new and hopefully valuable contribution to the literature.

Beyond these specific statistics, it is also of interest to provide a plot of the full posterior predictive distribution. While (11) is conditioned on the parameters $\boldsymbol{\Gamma}$, the Bayesian approach handles parameter uncertainty by integrating the parameters out of the conditional posterior predictive. In the context of our problem, we would like to obtain

$$p(w_f^* | c_f = j, \mathbf{c}, \mathbf{w}) = \int p(w_f^* | c_f = j, \boldsymbol{\Gamma}) p(\boldsymbol{\Gamma} | \mathbf{c}, \mathbf{w}) \, d\boldsymbol{\Gamma}, \tag{12}$$

since w_f^*, in the case of random sampling, is independent of the past data $\mathbf{c}, \mathbf{w}$ given $\boldsymbol{\Gamma}$. Simply plotting the left-hand side of (12) can be done without having to take draws from the (conditional) posterior predictive, as described above. Instead, we can provide a "Rao–Blackwell" approximation of this posterior predictive by noting

$$p(w_f^* | c_f = j, \mathbf{c}, \mathbf{w}) \approx \frac{1}{R} \sum_{r=1}^{R} p(w_f^* | c_f = j, \boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{(r)}),$$

where $\boldsymbol{\Gamma}^{(r)}$ is the rth post-convergence simulation from the posterior distribution, R denotes the total number of simulations, and $p(w_f^* | c_f = j, \boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{(r)})$ is known from (11). In our empirical results provided in Section 4, we use this method to plot various posterior predictives and also sample directly from (11) to calculate specific features of this distribution whose values are not known analytically.

4. The data and study limitations

This study employs data from the 2005 survey of the "Iowa Lakes Project," a four-year study and panel data collection effort aimed at understanding recreational use and the value of water quality in the primary recreational lakes of Iowa. The project began in 2002 with mail surveys sent to a random sample of 8,000 Iowa residents, obtaining detailed information regarding their visitation patterns to approximately 130 lakes, as well as standard socio-demographic data (e.g., age, education, income and gender). In subsequent years, surveys were sent to those households completing a survey in the prior year.[15] Standard follow-up procedures were followed in each year of the survey, including a postcard reminder mailed two weeks after the initial mailing and a second copy of the survey mailed one month later. Households were provided a $10 incentive for completing the survey each year, helping to provide for response rates ranging from 62–72 percent across the four years of the project.

---

[14] It is also worth noting that (11) is closely related to the *skew normal* random variable, and our procedure for sampling from it can be regarded as a slight generalization of the additive construction representation of this random (e.g., [18,19]).

[15] A second random sample was added in the panel in 2003 to fill in for non-deliverable surveys in 2003 and return the sample to a total of 8,000 households. No additional households were added after 2003.

The contingent valuation (and consequentiality) responses modeled in the current paper were included only in the 2003 and 2005 versions of the questionnaire. In 2003, a referendum style CV question was posed to estimate WTP for a water quality improvement project at one of eight focus lakes targeted in the study. These lakes were selected in consultation with the Iowa Department of Natural Resources. In addition to being geographically dispersed, the target lakes are each of policy interest since various restoration projects are being considered for them. Each respondent was asked a CV question for only one of the focus lakes. The survey described the current water quality information of the lake, including water clarity, water color, water odor, health concerns from algae blooms and bacteria level, and variety and quantity of fish. A photograph, illustrating the water clarity and water color, was provided to help respondents picture the current water quality of the focus lake vividly. Prior to the main valuation question, cheap talk text was used to remind respondents about the incentive and information properties of stated preference question in a further attempt to elicit truthful responses to the CV questions.[16]

A water quality improvement project regarding the focus lake was then proposed that outlined the methods to achieve the water quality improvement at each lake, such as dredging and building protection strips around the perimeter. Respondents were asked whether they would vote in favor of the referendum to improve the lake where bid values (ranging from $100 to $600 payable over a five year period) were randomly assigned to households in the sample receiving the survey. Finally, the perceived consequentiality of the survey was elicited from respondents by asking the following question: "How likely do you think it is that the results of surveys such as this one will affect decisions about water quality in Iowa lakes?" Possible responses to this question ranged from 1 to 5 where a 1 denoted "no effect at all" (i.e., completely inconsequential) and 5 denoted "definite" effects.

The 2003 survey responses indicated that the more consequential the respondents viewed the survey, the higher the observed "yes" rate to the proposed referendum.[17] The "yes" rate was lowest (13 percent) among respondents who believe the survey is inconsequential, and ranged from 27 percent to 38 percent when respondents believe the survey has some degree of consequentiality. In addition, the voting pattern of respondents who reported any degree of consequentiality (i.e., anyone who reported a 2, 3, 4, or 5 to the consequentiality question) was consistent with economic theory: the "yes" rate fell as the bid value rose.[18] In this sample, it was also the case that the "yes" rate of the inconsequential group was clearly lower than that of consequential group at each bid value.

The results of the 2003 survey were promising in terms of indicating the relationship between the responses of survey participants to the dichotomous choice CV referendum and the individual's perceptions regarding the survey's policy consequentiality. To measure the extent of this relationship, we could estimate a simple, single-equation WTP function for each level of reported consequentiality and test whether the distributions are statistically different from one another. However, as noted in the Introduction, there is the potential endogeneity of respondents' responses to the consequentiality question. In particular, respondents who have a high WTP may believe the survey is consequential due to the importance of the project. Likewise, respondents who indicated a low degree of consequentiality may do so because they place a low value in the proposed water quality improvement project.

To control for this potential endogeneity, a split sample treatment was added to the 2005 Iowa Lakes survey. To do so, we took advantage of a unique opportunity in the form of a magazine article and letter from the director of the state's Department of Natural Resources that had recently been published in the *Iowa Conservationist*, the magazine of the DNR. In the introductory letter, the DNR director (Jeff Vonk) stated that "... information from these two studies is going to be critical in prioritizing lake restoration projects not only from an environmental perspective, but from a return on investment standpoint." The presence of this article provided us with a unique opportunity to provide compelling evidence of the consequentiality of a survey directly to survey participants. Rather than simply telling respondents that the survey results would be made available to policy makers, we were able to provide direct evidence from a widely circulated source indicating that the answers to the survey would directly affect policy decisions. Specifically, half of the sample was provided with a copy of the director's letter and the associated article indicating that the IDNR was already using results from previous surveys in their policy decisions and planned to continue to do so.

It is possible, however, that the presentation of this supporting material will have affected the perception of the public good, thereby reducing the value of the split sample treatment as an instrument.[19] If this material is presented in such a way, however, that does not emphasize the importance of the project itself, but rather, only emphasizes the importance of the survey responses in decision making, then the instrument would remain legitimate.

To focus participants attention on the consequentiality aspect of the letter, in the materials provided to the treated survey respondents, a particular passage was highlighted and its message summarized (by us) in bold red print in the margin of the document. It is our view that the information treatment will have its primary affect on individuals through this highlighted passage, as it should be the first (and potentially only) passage that the respondent will read. That specific passage states: "Director Vonk agrees that your input is important in prioritizing restoration projects."[20] This statement

---

[16] The specific wording used in the cheap talk is provided in Appendix C, which is available at JEEM's online archive of supplementary material.

[17] Summary statistics for the 2003 survey are provided in Appendix table B1, which is available at JEEM's online archive of supplementary material.

[18] See Appendix B, Figure 1.

[19] Ideally, the published letter from the DNR director would have said nothing about the water quality of the lakes other than explaining the importance of the survey results to decisions; however, this could hardly be expected.

[20] The full text of the letter from the Director as sent to survey respondents and the accompanying cover letter is contained in Appendix C, available at JEEM's online archive.

Table 1
2005 summary statistics by level of consequentiality.

Variable	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$
Num. observations (n)	83	266	782	689	176
Percentage YES	0.23	0.42	0.47	0.53	0.48
Percent not visited	0.50	0.25	0.29	0.31	0.26
Percent male	0.53	0.69	0.64	0.66	0.61
College	0.33	0.42	0.41	0.49	0.37
Income/$1,000	58.05	56.71	54.72	53.73	57.71
Std. income	(16.60)	(15.62)	(15.31)	(14.35)	(14.98)
Age	54.00	63.43	63.75	64.95	59.30
Std. age	(38.64)	(37.28)	(38.12)	(37.38)	(41.07)

appeals directly to the individual's perceptions regarding the consequentiality of the survey responses, but makes no reference to the underlying importance of the projects themselves. In addition, this point was reiterated in the cover letter sent to respondents. Thus, we feel that the instrument used is indeed valid in practice.

There are two other aspects of the 2005 survey worthy of note. First, the percentage "yes" response rates from the 2003 survey never exceeded fifty percent. Typically, one wants the "yes" responses rates to range from 25 percent to 75 percent over the various bid values used, reducing the need for the model (rather than the data) to characterize the shape of the willingness-to-pay distribution. To alleviate this problem, the bid values in 2005 were once again randomly assigned, but this time they were allowed to range from $10 to $600. The percentage "yes" response rates in 2005 ranges from less than 15 percent to nearly 80 percent. Second, due to other project needs, only a subset of 2005 survey participants were used as part of the split treatment consequentiality investigation. Specifically, just under 2000 of the households were randomly divided between the treatment and control groups used to estimate the model developed in Section 3 above.[21]

Table 1 provides a summary of the data obtained in 2005. The "yes" response rates from the 2005 survey (as depicted in Fig. 1) are similar to those observed in 2003 survey. Specifically, though the sample sizes in each cell are smaller than in the 2003 survey (See Appendix B, Figure 1), there is a clear trend of decreasing "yes" rates as the bid value rises for the groups of respondents who indicated some degree of consequentiality (i.e., with responses of 2, 3, 4, or 5). But this pattern is not observed for those in the inconsequential group. Moreover, as observed in the 2003 survey, the "yes" rate amongst the inconsequential group is lower than that of consequential group even in lower bid values.

5. Empirical results

Using the model described Section 3 and the algorithm fully documented in the Appendix, we fit our two equation triangular treatment-response model. We run our posterior simulator for 50,000 simulations and discard the first 5,000 simulations as the burn-in. Numerous generated data experiments, which are not reported here for the sake of brevity, revealed that our algorithm mixed reasonably well (i.e., the lagged autocorrelations among our parameter simulations were not severe, and for some parameters, resembled what would be obtained under iid sampling), and consistently recovered parameters of the data generating process in a variety of experimental designs (i.e., those with high and low degrees of endogeneity and those with balanced and non-balanced frequencies for the observed consequentiality responses). For our priors, we set $\mu_\beta = \mathbf{0}$, $\mathbf{V}_\beta = (200^2)\mathbf{I}_k$, $v = 5$ and $R = \mathbf{I}_2$, which are reasonably non-informative choices, suggesting that information coming from the data will dominate information added through our prior.

5.1. Diagnostic checking

Before diving into our empirical results, we first provide some information regarding the overall performance of our model. Our assumptions, significantly stronger than those required for popular alternatives like standard IV, involve a complete description of the joint distribution of observables $p(\mathbf{c}, \mathbf{w}|\mathbf{\Gamma})$. To this end, it is important to assess the appropriateness of these assumptions and to document potential deficiencies associated with the model specification.

There are numerous diagnostic checks for investigating the reasonableness of a model's assumptions in the Bayesian paradigm, including, for example, the use of QQ plots (e.g., [20, Chapter 2]), posterior predictive p-values (e.g., [21, Section 6.3]), and other comparisons of specific features of the model to their counterparts in the observed data (e.g., [13, Chapter 11]). We focus here on one such exercise, which involves simulation from the posterior predictive distribution.

The idea behind this exercise is to try and *replicate* the distribution of the observed data, or specific features of the data that are important to the problem at hand, based on output from the model itself. If the model is legitimate for the analysis of a given data set, then it is reasonable to require that the model will produce a distribution of predicted outcomes that mimics what is found in the raw data. For example, if a simple wage regression were run under the assumption that wages

---

[21] Appendix Table B2 provides a breakdown of the consequentiality responses for the treatment and control groups.
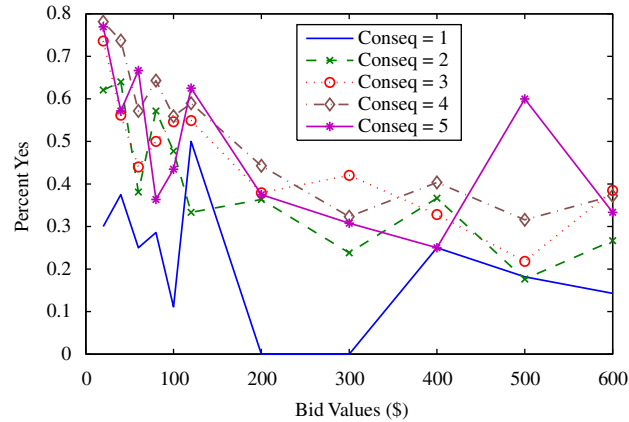
Fig. 1. 2005 percentage "Yes" by bid value for the five consequentiality groups.

were normally distributed, the posterior predictive density of wages would be symmetric, while the observed density of hourly wages would have a pronounced right-skew. This exercise would therefore reveal that an assumption of the model—namely that of normality—is inappropriate for the observed data.

To this end, we obtain a vector $\mathbf{y_j^{rep}} = [\mathbf{c_j^{rep'}} \, \mathbf{w_j^{rep'}}]'$ from $p(\mathbf{y^{rep}}|\mathbf{y})$ where

$$p(\mathbf{y^{rep}}|\mathbf{y}) = \int_{\mathbf{R_\Gamma}} p(\mathbf{y^{rep}}|\Gamma, \mathbf{y}) p(\Gamma|\mathbf{y}) \, d\Gamma, \tag{13}$$

and $\mathbf{y_j^{rep}} \sim \mathbf{y^{rep}}|\Gamma = \Gamma_j, \mathbf{y}$, with Γ_j representing the jth post-convergence draw from our posterior simulator. The density $p(\mathbf{y^{rep}}|\Gamma, \mathbf{y})$ is simply the likelihood function assumed by the given model (which does not depend on $\mathbf{y}$ given Γ). In these simulations, we obtain a vector $\mathbf{y_j^{rep}}$ from the conditional density $p(\mathbf{y^{rep}}|\Gamma = \Gamma_j, \mathbf{y})$ by choosing exactly the same $\mathbf{x_c}$ and $\mathbf{x_w}$ values as those that our found in our sample of data. This motivates our use of the notation "rep" to denote *replications* of the observed data from the posterior predictive, i.e., they are "a re-run of history on the assumption that the model is what generates histories" [20, pp. 90–91].

The act of generating a series of y^{rep} variates in this way can be used to reveal how well the model fares in reproducing the actual outcome distributions in our sample. Thus, for each post-convergence simulation Γ_j, we draw the latent vector $\tilde{\mathbf{c}}^{*,\mathbf{rep}}$ using the $\mathbf{x_c}$ values in the sample, calculate the discrete indicator $\mathbf{c^{rep}}$ from this latent data, and then generate a latent willingness to pay $\mathbf{w^{*,rep}}$ from the conditional distribution $p(\mathbf{w^{*,rep}}|\tilde{\mathbf{c}}^{*,\mathbf{rep}}, \Gamma_j, \mathbf{x_{w,rep}} = \mathbf{x_w})$. From this final draw, we calculate the replicated willingness to pay as $\mathbf{w^{rep}} = \max\{0, \mathbf{w^{*,rep}}\}$. We then compare features of these replicated simulations to those same features from the observed data.

As we should expect, the posterior predictive frequencies associated with the consequentiality responses closely match those found in the raw data. Specifically, the posterior means associated with the fraction of observations for $c^{rep} \in \{1, 2, 3, 4, 5\}$ were 4.4, 13.2, 38.9, 34.5 and 8.9, while those found in the raw data were 4.2, 13.3, 39.2, 34.5 and 8.8, respectively. At the individual level, however, the results are less encouraging. The posterior mean (and posterior standard deviation) of the fraction of $\mathbf{c^{rep}}$ values that correctly predict their counterpart in the observed data is 0.31 (0.011). This, of course, is an improvement over randomly guessing based on a uniform prior (which would achieve a success rate of 20 percent), but does suggest that our covariates will not play a strong role in describing variation in consequentiality perceptions. In particular, and as will be discussed in the following subsection, the instrument will only play a small role in tracing out variation in consequentiality responses. This will lead to reduced posterior precision associated with the "causal" effect parameters $\delta_\mathbf{c}$ in (4), which will make it more difficult to get conclusive information regarding tests of parameter equality. It is important to note, however, that these results do not necessarily signal a deficiency of the model's assumptions, but simply document that it is difficult to determine why some individuals believe the surveys are useful for policy purposes, while others believe they have little or no value for policymaking.

To assess the adequacy of the willingness to pay distribution, we consider two primary statistics: $\Pr(w^{rep} \geq 400|\mathbf{y})$ and $\Pr(w^{rep} \leq 60|\mathbf{y})$. The motivation for this focus is to determine if our model can adequately recover these "tail" probabilities, while recognizing that our model is not constructed to necessarily recover these quantities.[22] In practice, it is often difficult for a model to predict such "extreme" outcomes, making statistics such as these a useful metric for assessing the performance of our model. Though the largest bid in our sample is \$600 and the smallest is \$10, we focus our calculations instead on more "interior" bid points due to small samples sizes and observed non-monotonicity in our sample. Specifically, just looking at the

---

[22] A focus on the mean, for example, is a comparison that should typically be avoided, since parameters of the model often capture the mean of the distribution of interest. The above quantities clearly rely on the accuracy of the normality assumption.

Table 2
Posterior means, standard deviations and probabilities of being positive.

Variable	$E(\cdot|y)$	$Std(\cdot|y)$	$Pr(\cdot > 0|y)$
Consequentiality equation			
Constant	1.64	0.059	1.00
Iowa Conservationist Article	0.103	0.046	0.984
Age	−0.020	0.025	0.185
Female	0.003	0.051	0.521
College	0.071	0.053	0.914
Income	−0.008	0.026	0.397
Willingness to pay equation			
Constant	51.34	68.56	0.779
Not consequential	−192.1	106.3	0.032
Moderately consequential	34.12	60.39	0.719
Consequential	57.03	98.31	0.720
Definitely consequential	−63.63	150.44	0.331
Age	26.48	19.30	0.909
Female	72.91	39.77	0.974
College	135.99	41.56	0.999
Income	91.79	22.16	1.00
Covariance matrix parameters and cutpoints			
σ_u^2	3.82×10^5	7.46×10^4	1.00
$\rho_{\varepsilon u}$	0.099	0.076	0.925
α_3	0.785	0.045	1.00
α_4	1.89	0.051	1.00
α_5	3.07	0.060	1.00

raw data, we find that 34.6, 28.4 and 34.5 percent of our sample say they are willing to pay bid amounts of $400, $500 and $600, respectively, for the proposed water quality improvement. This non-monotonicity represents a daunting challenge for replication purposes, leading us to average these frequencies for all bids greater than or equal to $400 and use this as a basis for comparison. Performing the required calculations, we find a posterior mean (and standard deviation) associated with the event $w^{rep} \geq 400$ equal to 0.35 (0.02), which is close to and within two standard deviations of the fraction in the raw data, 0.32. On the other end of the WTP distribution, we find a posterior mean (and standard deviation) associated with the event $w^{rep} \leq 60$ equal to 0.44 (0.19), while the observed fraction in the raw data is quite close to this value at 0.39. Taken together, the results of these calculations do not seem to provide any direct evidence against our model's assumptions or that any type of re-modeling beyond what is asserted in (3) and (4) is obviously necessary.

5.2. Parameter estimates and discussion of results

Presented in Table 2 are posterior means and standard deviations associated with the parameters of our model. The continuous variables age and income are standardized to have mean zero and unit variance for interpretation purposes and to suitably scale the values of the associated slope parameters. In the willingness to pay equation, we include an intercept parameter and omit the $c_i = 2$ ("Barely consequential") category so that all other elements of δ_c are interpreted relative to this excluded group.

The coefficients associated with demographic variables in the willingness to pay equation are generally consistent with our prior expectations. Females, older respondents, more educated individuals, and those with higher incomes are more willing to pay for a given water quality improvement. Specifically, females have a (latent) willingness to pay that is approximately $73 more than males, on average, while the college educated have a latent WTP approximately $136 higher than their non-college counterparts. Similarly, a one-standard deviation increase in age (approximately 15 years) is associated with an expected increase in latent willingness to pay of approximately $26, while a similar standard deviation increase in income (approximately $38,000) is associated with an expected increase in latent willingness pay of about $92. Perhaps with the exception of the age coefficient (whose posterior probability of being positive is still reasonably high at 0.91), the marginal posteriors associated with these coefficients place nearly all mass over positive values.

As suggested previously in our section on diagnostic checking, few variables emerge as influential predictors of consequentiality perceptions, and again, these parameters should not be given any kind of structural interpretation. Fortunately, our instrument, the indicator denoting the receipt of the highlighted article from the *Iowa Conservationist*, is positively associated with the perceived degree of consequentiality, and the posterior places virtually all mass over positive values. Simulations from the posterior distribution were also used to show that receipt of the highlighted article increased the probability that an "average" college educated-female will be in the highest consequentiality groups (i.e., $c_i \in \{4, 5\}$) by about 4.5 percent. Thus our instrument, though not highly influential, does play some role in explaining variation in consequentiality perceptions. The only remaining predictor which plays some role in the consequentiality equation is the

college indicator, with the college educated more likely to believe that the survey is definitely consequential, and less likely to believe that it has no impact on policy decisions.

Interestingly, we also find evidence suggesting an important role for unobserved confounding in our application. Specifically, the posterior mean of the correlation between the errors in (1) and (2), denoted ρ_{cw}, was found to be 0.099, and $\Pr(\rho_{cw} > 0|y) = 0.925$. This suggests, with a reasonable degree of posterior certainty, that unobserved characteristics leading the agent to believe that a survey is likely to be consequential are positively related with unobserved factors contributing to WTP. Said differently, characteristics of people who believe that their actions (such as responses to a survey) can make a difference correlate positively with (unobserved) factors making a person willing to pay to see such improvements come to fruition. Though the magnitude of this correlation is small, the fact that the marginal posterior concentrates on positive values suggests the need to control for the potential endogeneity of perceived consequentiality.

A more formal investigation regarding the importance of unobserved confounding can be conducted by providing a test of the hypothesis $\rho_{\varepsilon u} = 0$. From a Bayesian point of view, this is conducted by calculating the *posterior odds ratio*, K_{12}:

$$K_{12} \equiv \frac{p(\mathcal{M}_1|y)}{p(\mathcal{M}_2|y)} = \frac{p(y|\mathcal{M}_1)}{p(y|\mathcal{M}_2)} \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)}, \tag{14}$$

where $\mathcal{M}_j$ denotes *model j*, $j = 1, 2$, $p(\mathcal{M}_j|y)$ denotes the *posterior probability* of model j, $p(y|\mathcal{M}_j) = \int_{\Theta_j} p(y|\theta_j, \mathcal{M}_j) p(\theta_j|\mathcal{M}_j) d\theta_j$ is the *marginal likelihood* associated with model j and $p(\mathcal{M}_j)$ is the *prior probability* associated with model j. A common convention is to assume that both models are equally probable *a priori* so that the prior odds ratio cancels in (14) and thus the posterior odds ratio K_{12} equals the ratio of marginal likelihoods (also known as the Bayes factor).

To apply this model comparison method to our particular problem, let $\mathcal{M}_1$ denote the restricted model with $\rho_{\varepsilon u} = 0$ and let $\mathcal{M}_2$ denote the unrestricted model in (3) and (4). For hypotheses like this one, provided parameters common to $\mathcal{M}_1$ and $\mathcal{M}_2$ are given the same priors in both models, the ratio of posterior odds in (14) reduces to (see, [13, p. 69])

$$K_{12} = \frac{p(\mathcal{M}_1|y)}{p(\mathcal{M}_2|y)} = \frac{p(\rho_{\varepsilon u} = 0|\mathcal{M}_2, y)}{p(\rho_{\varepsilon u} = 0|\mathcal{M}_2)},$$

the ratio of the marginal posterior and prior ordinate at zero under the unrestricted model. Calculating this ratio produces a posterior odds value equal to 0.35, suggesting that the unrestricted model with $\rho_{\varepsilon u} \neq 0$ is favored over the restricted version by a factor of approximately 2.84 to 1. This, again, provides evidence supporting the need to control for unobserved confounding and the endogeneity of consequentiality perceptions in our application.

5.3. Do the data support a knife-edge effect of consequentiality?

The results of Table 2 clearly indicate that the perceived policy consequentiality of the survey does impact the individual's response to the CV referendum, with those believing the survey to be completely inconsequential reporting a WTP that is significantly lower than the baseline group (i.e., $c_i = 2$). A natural question to ask in this setting is whether the effect is a knife-edge one. That is, are the WTP distributions, and specific features of those distributions, equal for individuals with varying, but positive, perceptions regarding the survey's policy consequentiality? Recall that the theoretical work of Carson and Groves [2] supports the notion of a knife-edge result when strong consequentiality is considered. It is less clear whether a similar result will hold in the context of policy consequentiality alone.
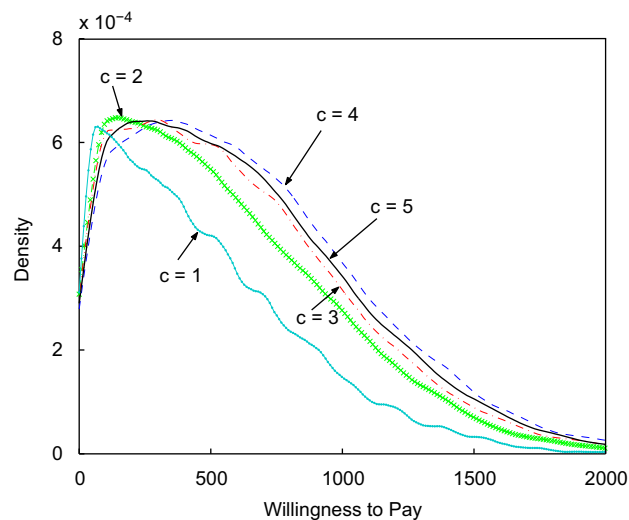

Fig. 2. Willingness to pay posterior predictive distributions for five consequentiality groups.

Table 3
Posterior probabilities and posterior predictive WTP statistics across different levels of consequentiality.

Variable	Consequentiality response				
	1	2	3	4	5
$\Pr(\delta_{c1} > \cdot \,|y)$	–	–	–	–	–
$\Pr(\delta_{c2} > \cdot \,|y)$	0.97	–	–	–	–
$\Pr(\delta_{c3} > \cdot \,|y)$	0.96	0.44	–	–	–
$\Pr(\delta_{c4} > \cdot \,|y)$	0.94	0.51	0.64	–	–
$\Pr(\delta_{c5} > \cdot \,|y)$	0.73	0.29	0.20	0.08	–
Median WTP	0	$171.06	$264.9	$355.0	$306.2
Mean WTP	$218.44	$349.42	$404.67	$468.12	$434.22

Our first step in the investigation of this question is to plot the posterior predictive willingness to pay distributions for the various consequentiality groups, using the methodology described in Section 3.1 of the paper. These plots are presented in Fig. 2, and are calculated, for simplicity, for a college-educated female of average age and income who received the article touting the importance of the survey.

What we take away from Fig. 2 is some evidence that is broadly consistent with the knife-edge result. The implied WTP distributions for the group believing the survey to be completely inconsequential ($c = 1$) seems rather different from those of the other groups, and the remaining densities seem mostly similar. The location of the mass of the densities might also seem somewhat surprising, as non-negligible mass is placed over seemingly large dollar amounts. In this regard, we remind the reader of results discussed in our diagnostic checks of Section 5.1: in the raw data, 34.5 percent of individuals are willing to pay the maximum assigned bid value of $600. As such, the posterior predictives *must* place considerable mass to the right of this largest bid point. The problem here is that we do not observe any outcomes to the right of the maximum bid of $600 to inform the shape of this distribution over that region; instead, its shape is determined by estimating a mean, a variance and other statistics purely from a sequence of binary responses, which are then used (together with our parametric assumptions) to characterize the entire WTP predictive. Although one might suspect that these predictives places too much mass over large WTP values, there is simply no information available from the data to either support or refute these suspicions.

Importantly, the predictive WTP distributions in the figure unite two conceptually different sources of information. First, they account for the direct "structural" impacts of consequentiality perceptions on willingness to pay through the $\boldsymbol{\delta_c}$ coefficients. Second, they also account for the role of unobservables and the fact that individuals believing the survey to be very consequential have, on average, unobserved characteristics that also make them more likely to pay for a given environmental improvement.

As a test of the knife-edge result, what seems to be of primary interest is a comparison of the structural coefficients. If differences in the WTP distributions for those with $c \geq 2$ arise only because of differences in unobservables across groups, then the findings of our analysis are still consistent with the knife-edge result. That is, any differences in willingness to pay can be interpreted as arising only from the observational nature of the survey itself, since, in the survey area, we are not able to randomly assign individuals to groups where all factors will be, on average, balanced across the groups. Conversely, if the structural parameters are significantly different, this suggests persistent WTP differences that cannot be explained through differences in unobservables, and could potentially be argued to be at odds with a knife-edge hypothesis.

To compare these structural coefficients more formally, we report a variety of quantities of interest in Table 3. For the first set of entries of the table, we calculate posterior probabilities of the form $\Pr(\delta_{cj} > \delta_{ck}|\mathbf{y}) \; \forall j, k \in \{1, \ldots, 5\}$.[23] These parameters, and the corresponding probabilities mentioned above, are what researchers might typically look to first as a "test" of the theory. As is evident from the table, the subgroup who perceives the survey as being completely irrelevant for policy purposes clearly has lower willingness to pay than those who believe the survey to be somewhat consequential, and also a lower (though this statement is made with far less certainty) WTP than those regarding the survey as completely consequential. This result is consistent with the CG theoretical predictions, wherein the behavior of individuals believing that the survey is irrelevant cannot be characterized; our results suggest that these individuals have different WTP reports than the remaining groups, supporting the notion that these agents and their reports can be excluded in practice.

There is also reasonably strong evidence that, among those believing the survey to be at least minimally consequential, the structural impacts across groups are not obviously distinguishable. This is evident from the fact, apart from the comparison between groups 4 and 5, most of the probabilities found in the table do not seem to provide strong evidence of parameter differences.

To investigate this issue more formally, we revisit Eq. (14) and calculate Bayes factors associated with the individual restrictions $\delta_{cj} = 0$, for $j \in \{1, 3, 4, 5\}$. We would regard our analysis as providing support for the theoretical predictions if the Bayes factor associated with δ_{c1} favored keeping the indicator in the model, while those associated with δ_{c3}, δ_{c4} and δ_{c5}

---

[23] For example, the table can be used to show that $\Pr(\delta_{c2} > \delta_{c1}|\mathbf{y}) = 0.97$, and similarly for the remaining entries.

favored excluding the indicators from the model. When performing these calculations, we find Bayes factors in favor of the zero restriction equal to 0.36, 2.93, 1.74 and 1.20 for $j = 1, 3, 4, 5$, respectively. Thus, by a factor of approximately 2.8 to 1, we support keeping the indicator for the inconsequential group in the model, and also support (to varying degrees), dropping the remaining consequentiality indicators.[24] Again, these results provide support for the knife-edge theoretical predictions; similar WTP distributions emerge for those believing the survey to be minimally consequential, while those believing the survey to be completely inconsequential have different WTP distributions, even after accounting for unobserved differences across the groups.

6. Conclusion

In this paper, we investigate the role of policy consequentiality in context of dichotomous choice CV referenda. Specifically, we have tested the hypothesis of whether willingness to pay distributions are equal for those individuals who believe the survey has at least some potential for shaping policy decisions. Using a treatment-response model that controls for unobserved confounding and exploits a survey design in which a subsample of individuals are randomly provided supporting material documenting the importance of the study, we find evidence that is broadly consistent with a knife-edge impact for policy consequentiality. That is, we find support for the equality of WTP distributions among those believing the survey is at least minimally consequential, while those believing the survey will have no effect on policy have statistically different distributions associated with WTP.[25] Our methodology also makes use of a new Bayesian posterior simulator for fitting the nonlinear treatment-response model, which we hope will appeal to other researchers seeking to estimate models with a similar structure.

Acknowledgments

References

[1] K. Arrow, R. Solow, E. Leamer, P. Portney, R. Radner, H. Schuman, Natural resource damage assessments under the Oil Pollution Act of 1990, Federal Register 58 (1993) January 15.
[2] R.T. Carson, T. Groves, Incentive and informational properties of preference questions, Environmental and Resource Economics 37 (2007) 181–210.
[3] M. Cowles, Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models, Statistics and Computing 6 (1996) 101–111.
[4] B. Nandram, M.-H. Chen, Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence, Journal of Statistical Computation and Simulation 54 (1996) 129–144.
[5] E. Bulte, S. Gerking, J.A. List, A. de Zeeuw, The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study, Journal of Environmental Economics and Management 49 (2005) 330–342.
[6] G. Harrison, J. List, Field experiments, Journal of Economic Literature XLII (December) (2004) 1009–1055.
[7] R.G. Cummings, E.G. Harrison, J. Murphy, Are Hypothetical referenda incentive compatible?, Journal of Political Economy 105 (1997) 609–621.
[8] R.G. Cummings, L.O. Taylor, Does realism matter in contingent valuation surveys?, Land Economics 74 (2) (1998) 203–215.
[9] J. Shogren, L. Tadevosyan, Bidding behavior in a consequential second price auction, Working paper, Department of Economics and Finance, University of Wyoming, 2007.
[10] C.E. Landry, J.A. List, Using ex ante approaches to obtain credible signals for value in contingent markets: evidence from the field, American Journal of Agricultural Economics 89 (2) (2007) 420–429.
[11] R.T. Carson, T. Groves, J.A. List, M.J. Machina, Probabilistic influence and supplemental benefits: a field test of the two key assumptions underlying stated preferences, Working Paper, University of California, San Diego, 2004.
[12] M. Tanner, M. Wong, The calculation of posterior distributions by data augmentation, Journal of the American Statistical Association 62 (1987) 810–818.
[13] G. Koop, D.J. Poirier, J.L. Tobias, Bayesian Econometric Methods, Cambridge University Press, Cambridge, 2007.
[14] J.H. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, Journal of American Statistical Association 88 (1993) 669–679.
[15] K. Li, Bayesian inference in a simultaneous equation model with limited dependent variables, Journal of Econometrics 85 (1998) 387–400.
[16] P. Deb, M. Munkin, P. Trivedi, Bayesian analysis of the two-part model with endogeneity: application to health care expenditure 21 (7) (2006) 1081–1099.

---

[24] It is seemingly important here to mention Bartlett's paradox, where the restricted model receives larger and larger support as the prior becomes increasingly non-informative. More non-informative priors (those with prior standard deviations for the δ parameters greater than 200) will only lead to stronger evidence in favor of the restrictions, i.e., more support for the equality of WTP distributions for those believing the survey to be at least somewhat consequential.

[25] Again, we emphasize that our analysis is focussed on the role of policy consequentiality, not strong consequentiality. As such, it is not a direct test of Carson and Groves [2] theoretical prediction of the knife-edge result for strong consequentiality *unless* survey respondents believe that the policy consequentiality carries with it an implication of "payment" consequentiality as well. While we believe this to be the case, the reader will have to judge for themselves as to whether or not this is a reasonable belief.

[17] P. Damien, J. Wakefield, S. Walker, Gibbs sampling for Bayesian nonconjugate and hierarchical models using auxiliary variables, Journal of the Royal Statistical Society Series B 61 (1999) 331–344.

[18] B.C. Arnold, R.J. Beaver, An alternative construction of skewed multivariate distributions, Technical Report 270, Department of Statistics, University of California-Riverside, 2003.

[19] B.C. Arnold, R.J. Beaver, Elliptical models subject to hidden truncation or selective sampling, in: M.C. Genton (Ed.), Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality, 2004, pp. 101–112 (Chapter 6).

[20] T. Lancaster, An Introduction to Modern Bayesian Econometrics, Blackwell, Oxford, 2004.

[21] A. Gelman, J.B. Carlin, H. Stern, D. Rubin, Bayesian Data Analysis, second ed., Chapman & Hall, London, 2004.