UNR Joint Economics Working Paper Series
Working Paper No. 10-001

Latent Thresholds Analysis of Choice Data with Multiple Bids and Response Options

Mimako Kobayashi, Klaus Moeltner and Kimberly Rollins

Department of Resource Economics / MS 204
University of Nevada, Reno
Reno, NV 89557
(775) 784-6701 │ Fax (775) 784-1342
email: moeltner@cabnr.unr.edu

March, 2010

Abstract

In many stated preference settings stakeholders will be uncertain as to their exact willingness-to-pay for a proposed environmental amenity. To accommodate this possibility analysts have designed elicitation formats with multiple bids and response options that allow for the expression of uncertainty. We argue that the information content flowing from such elicitation has not yet been fully and efficiently exploited in existing contributions. We introduce a Latent Thresholds Estimator that focuses on the simultaneous identification of the full set of thresholds that delineate an individual's value space in accordance with observed response categories. Our framework provides a more complete picture of the underlying value distribution, the marginal effects of regressors, and the impact of bid designs on estimation efficiency. We show that the common practice of re-coding responses to derive point estimate of willingness-to-pay leaves useful information untapped and can produce misleading results if thresholds are highly correlated.

JEL Classification: C11, C15, C35, C52, Q51

Keywords: Stated Preference; Multiple Bounded Elicitation; Polychotomous Choice; Bayesian Estimation; Value Uncertainty

Latent Thresholds Analysis of Choice Data with Multiple Bids and Response Options

Abstract:

In many stated preference settings stakeholders will be uncertain as to their exact willingness-to-pay for a proposed environmental amenity. To accommodate this possibility analysts have designed elicitation formats with multiple bids and response options that allow for the expression of uncertainty. We argue that the information content flowing from such elicitation has not yet been fully and efficiently exploited in existing contributions. We introduce a Latent Thresholds Estimator that focuses on the simultaneous identification of the full set of thresholds that delineate an individual's value space in accordance with observed response categories. Our framework provides a more complete picture of the underlying value distribution, the marginal effects of regressors, and the impact of bid designs on estimation efficiency. We show that the common practice of re-coding responses to derive point estimate of willingness-to-pay leaves useful information untapped and can produce misleading results if thresholds are highly correlated.

Key Words: Stated Preference, Multiple Bounded Elicitation, Polychotomous Choice, Bayesian Estimation, Value Uncertainty

JEL Codes: C11, C15, C35, C52, Q51

I) Introduction

In recent decades elicitation formats for Stated Preference surveys have been extended to incorporate measures of respondent uncertainty for the payment of hypothetical contributions to obtain or preserve an environmental amenity.[1] For example, Li and Mattsson [3], Champ et al. [4], and Ekstrand and Loomis [5] capture payment uncertainty via numerical scales, while Ready et al. [6], Wang [7] , and Ready et al. [8] elicit uncertainty levels via a discrete set of qualitative response options ("definitely yes", "yes", "not sure", etc). Ready et al. [6] deem this latter approach the "polychotomous choice" (PC) format. All of these examples employ their uncertainty scale in combination with a single referendum question.

Welsh and Poe [9] extend this framework by adding additional bid amounts, each of which is paired with a polychotomous response choice. The authors label this approach the "Multiple Bounded Discrete Choice" format. It is also referred to as the "Multiple Bounded Uncertainty Choice format ([10]), and the "Multiple Bounded Polychotomous Choice" format ([11]) in subsequent applications. We adopt Evans et al's [10] acronym "MBUC" throughout this text.

The original rationale for allowing respondents to express uncertainty in contingent valuation studies was to lower the number of non-responses and "protest-zeros" that often plague generic Discrete Choice (DC) formats (e.g. [6]), and to gain further insights into the reliability and interpretation of responses when participants are constrained to choose between a simple "yes" and "no", as is the case in a basic DC setting (e.g. [4; 6; 8; 12]). Welsh and Poe's [9] primary motivation for extending the polychotomous format to an entire range of bids was to combine the benefits of the PC format with the efficiency gains that can be expected of multi-bounded elicitation (e.g. [13]). In addition, the authors isolate the two effects via a comparison of the MBUC method with a generic payment card approach (multiple bids *without* the polychotomous choice option) and a generic PC format (*single* bid with polychotomous choice).

In this study we argue that the information content flowing from an MBUC elicitation has not yet been fully and efficiently exploited in existing contributions. We propose a new estimator that focuses on the simultaneous estimation of multiple decision thresholds rather than a single point estimate of willingness-to-pay (*WTP*), or the mean of the underlying value distribution. We illustrate that this approach operates under less restrictive assumptions than existing estimators, and provides a more complete picture of the underlying value distribution and the marginal effects of regressors on the location and spread of this distribution. In addition, our estimator nests several existing models and thus allows for a rigorous examination of the underlying assumptions associated with these alternative specifications. We illustrate our framework using data from the first field implementation of the MBUC format in the 1994 Glen Canyon Pilot Study ([14]), and a recent valuation study on rangeland restoration in the Great Basin Region ([16]).

In the next section we discuss existing models of uncertain responses and highlight their strength and shortcomings. Section III introduces the econometric framework for the Latent Thresholds Estimator (LTE). This is followed by an empirical section that introduces the data and discusses estimation results. Section V concludes.

II) Modeling of Uncertain Responses

There are important differences in how existing studies have processed the additional information collected via numerical or qualitative uncertainty scales. All of them assume (implicitly or explicitly) that a given respondent's true value for the amenity in question is unknown to herself and the analyst, but that the respondent, in contrast to the analyst, knows its probability distribution. Let v_i be the respondent's uncertain value (or *WTP*) and $g_i\left(v_i\right)$ and $G_i\left(v_i\right)$ its continuous probability density function (*pdf*) and cumulative density function (*cdf*), respectively. Further denote the first two moments of v_i as μ_i and σ_i^2, respectively. As for the distribution, these moments are known to the subject, but not the analyst. Consider a given bid b_j and a response $r_i\left(b_j\right)$ chosen by *i* from a provided uncertainty scale. For a numerical scale, $r_i\left(b_j\right)$ might be "80% certain" or "certainty level 8 on a scale of 10", while for a polychotomous scale $r_i\left(b_j\right)$ will be a qualitative answer, such as "probably yes" or "definitely no".

Existing estimation strategies can be divided into two general groups: Those that interpret the uncertainty-response as an exact probability for the location of v_i relative to a given bid, i.e. as $G_i\left(b_j\right)$, and those that (implicitly or explicitly) interpret the response as a statement about the location of b_j relative to specific threshold amounts along the support of v_i that correspond to the discrete answer categories. We label the first set of approaches "probability-based estimators (PBEs)", and the second set as "threshold-based estimators (TBEs)".

Probability-based estimators

Li and Mattsson [3] were the first to propose a PBE for a DC framework with a percentage-based uncertainty scale (in 5% increments). They specify a normal distribution for v_i with common variance for all subjects, i.e.

$$v_i = \mu_i + \tau_i \qquad \tau_i \sim n\left(0,\sigma_\tau^2\right) \tag{1}$$

An observed response of $r_i\left(b_j\right)$ can then be directly interpreted as a probabilistic statement for the location of b_j within the v_i distribution, i.e.

$$r_i\left(b_j\right) = pr\left(v_i < b_j\right) = pr\left(\mu_i + \tau_i < b_j\right) = \Phi\left(\frac{b_j - \mu_i}{\sigma_\tau}\right) \tag{2}$$

where $\Phi(.)$ denotes the standard normal *cdf*. This, in turn leads to the definition of a standard normal variate that is a direct function of the observed probabilistic response, i.e

$$z_i = \frac{b_j - \mu_i}{\sigma_\tau} = \Phi^{-1}\left(\frac{b_j - \mu_i}{\sigma_\tau}\right) = \Phi^{-1}\left(r_i\left(b_j\right)\right) \tag{3}$$

The authors then combine this derived measure with a regression model for the expectation of v_i and estimate the model via Maximum Likelihood techniques.

Evans et al. [10] extend this approach to an MBUC setting with choice categories "definitely yes" (DY), "probably yes" (PY), "not sure" (NS) , "probably no" (PN), and "definitely no" (DN). With guidance from the psychological and psychometric literature they map these responses into probabilistic statements regarding the location of v_i relative to a given bid. For example, their initial mapping is given as

$$\begin{aligned} &r_i\left(b_j\right) = DY \rightarrow pr\left(v_i < b_j\right) = 0, \quad r_i\left(b_j\right) = PY \rightarrow pr\left(v_i < b_j\right) = 0.25, \quad r_i\left(b_j\right) = NS \rightarrow pr\left(v_i < b_j\right) = 0.5 \\ &r_i\left(b_j\right) = PN \rightarrow pr\left(v_i < b_j\right) = 0.85, \quad r_i\left(b_j\right) = DN \rightarrow pr\left(v_i < b_j\right) = 1 \end{aligned} \tag{4}$$

This, in turn, allows for the derivation of bounded probabilities for a sequence of bids in MBUC elicitation. For example, using the mapping in (4), observing $r_i\left(b_{j-1}\right) = NS$ and $r_i\left(b_j\right) = PN$ implies $P_{i;j-1,j} = pr\left(b_{j-1} < v_i < b_j\right) = 0.85 - 0.5 = 0.35$.

In contrast to [3] the authors leave the subject-specific value distribution $g_i\left(v_i\right)$ unspecified, but instead stipulate that the eventual *realization* of v_i, denoted here as v_i^*, follows a normal distribution across all subjects in the underlying population, i.e.

$$v_i^* = \beta + \varepsilon_i, \qquad \varepsilon_i \sim n\left(0, \sigma^2\right) \tag{5}$$

Under a paradigm of quadratic loss minimization the authors then specify an *expected* log-likelihood function. For an observed sequence of bids $b_1, b_2, \ldots, b_J$ by individual *i*, this term is given as

$$E_{v_i}\left(\ln l_i\left(\beta,\sigma\right)\right) = P_{i;-\infty,1} * \ln\left(\Phi\left(\frac{b_1-\beta}{\sigma}\right)\right) + \sum_{j=2}^{J} P_{i;j-1,j} * \ln\left(\Phi\left(\frac{b_j-\beta}{\sigma}\right) - \Phi\left(\frac{b_{j-1}-\beta}{\sigma}\right)\right) + P_{i;J,\infty} * \ln\left(1-\Phi\left(\frac{b_J-\beta}{\sigma}\right)\right) \quad (6)$$

Note that if *i*'s response to b_1 is "DY", $P_{i;-\infty,1} = pr\left(-\infty < v_i < b_1\right) = 0$ is zero and the first term in (6) drops out of the log-likelihood function. The same holds for the last term if the response to the final bid is "DN". Evans et al. [10] label this model the "Dual-Uncertainty Decision Estimator" (DUDE).

The main advantage of these PEB approaches is that they lead to estimates of expected *WTP* for the underlying population of stakeholders. This is an attractive feature as expected *WTP* is an important component in benefit-cost analysis, and an important value to feed into benefit-transfer applications. However, this gain comes at the cost of imposing very stringent and likely unrealistic model assumptions. Specifically, while it is reasonable from the analyst's perspective to let the *expectation* of v_i follow a pre-defined density such as the normal distribution, it is highly unlikely that every *individual's* uncertain valuation can be modeled by a common density, let alone a density with equal variance (equ. (1)). Given the numerous sources of value uncertainty that have been suggested in the literature (e.g. [6; 15; 16]) and the likely heterogeneous fashion in which these sources affect a given respondent's value assessment, we consider this assumption of "identical value distribution" for all subjects highly restrictive. Even if the homoskedasticity assumption is relaxed (a feasible extension of the Li and Mattsson [3] model), there is still the possibility that a given subject's value distribution might be skewed in either direction or even multi-modal.

Our main concern with Evans et al.'s [10] DUDE estimator is its interpretation and performance under "repeated infra-marginal answers" (RIAs). For example, it is quite possible (and frequently observed in practice) that a given subject issues the same non-certainty response for a series of sequential bids. In fact, this situation is virtually unavoidable whenever the number of bids (say *M*) exceeds the number of response categories (*J*). This is the case in most MBUC applications, with *J* ranging from 3 to 5, and *M* usually lying in the 10-20 range ([9; 16; 17]). For example, consider the following observed sequence of bids and responses:

$$b_1 \rightarrow DY, b_2 \rightarrow PY, b_3 \rightarrow PY, b_4 \rightarrow PY, b_5 \rightarrow NS, b_6 \rightarrow PN, b_7 \rightarrow DN \quad (7)$$

Using the mapping given in (4) this implies that

$P_{i;-\infty,1} = 0, P_{i;1,2} = 0.25, P_{i;2,3} = P_{i,3,4} = 0, P_{i;4,5} = 0.25, P_{i;5,6} = 0.35, P_{i;6,7} = 0.15, P_{i;7,\infty} = 0$. Nothing new is learned from the responses to bids three and four, and the corresponding terms in equation (6) will drop out of the likelihood function. While this makes intuitive sense mathematically, this observed choice sequence implies a rather unrealistic shape for the underlying value distribution. Specifically, $g_i(v_i)$ is forced to have zero density between bids two and four, while the density is well-defined in the immediate vicinity of these boundaries. Such "density gaps" are implied by any occurrence of RIAs in the sample. Therefore, the DUDE estimator is only well-defined under complete absence of RIAs. This is unlikely in most applications.

An additional limitation of the DUDE approach is its arbitrary mapping of polychotomous responses into probabilities. Even though Evans et al. [10] find their results to be fairly robust to different mappings, it is still unlikely that the same mapping applies to all respondents in the sample. In summary, we consider neither of these two PBE models as optimal approaches to process MBUC data. We will therefore direct our attention to threshold-based estimators.

Threshold-based estimators

The decision-threshold interpretation of polychotomous responses was originally proposed by Wang [7] for a 3-tiered choice scale ("yes", "don't know", "no"), and a single DC format. Alberini et al. [11] extend this approach to the MBUC format with the standard 5-category uncertainty scale. As mentioned above, in the TBE setting an individual's polychotomous response is interpreted as a statement about the location of the bid in question relative to specific threshold amounts. These thresholds are assumed to be fully known to the respondent, but unobserved by the researcher.

This notion is illustrated in Figure 1 for a standard 5-point decision scale. The survey participant will answer DY if the proposed bid lies to the left of the entire distribution of v_i, and DN if the payment amount exceeds the upper bound of the support of v_i. Thus, one can interpret thresholds t_{1i} and t_{4i} as marking the boundaries for the support of $g_i(v_i)$. Any bid between these thresholds will trigger an uncertain response. Specifically, the respondent will answer PY if the bid amount falls between t_{1i} and t_{2i}, NS if the bid lies between t_{2i} and t_{3i}, and PN if the bid is located between t_{3i} and t_{4i}. As implied by the "*i*"-subscript these thresholds will likely be individual-specific. [2]

However, Wang [7] and Alberini et al.'s [11] main focus lies not on the estimation of these thresholds, but rather the expectation of v_i. This is accomplished by re-defining the threshold values as limiting distances from $E(v_i)$, as depicted in Figure 1. The identification of the expectation requires imposing restrictions on at least one of these distances, such as symmetry (e.g. $c_i = -b_i$ or $d_i = -a_i$), or fixed ratio (e.g. $a_i = \alpha b_i$). The authors call this approach the "Random Valuation Model" (RVM).

The TBE framework also incorporates models that de facto "remove" respondent uncertainty by re-coding uncertain answers into boundary responses (i.e. DY or DN). This was the original approach taken in Welsh and Poe [9] in the MBUC context. This binary re-mapping is also adopted in other applications with discrete-numerical or polychotomous choices, such as Ready et al. [6], Champ et al. [4], and Ready et al. [8]. Once a binary mapping is achieved, standard estimation tools for single or double-bounded DC formats can be used to derive an estimate of

v_i for a given respondent.[3] Within our general TBE framework this re-mapping approach is equivalent to letting the distribution of v_i collapse to one of the four thresholds. For example, if all responses other than DY are interpreted as DN the estimate for v_i will center around the first threshold, t_{1i}. Conversely, if all responses other than DN are recoded to DY, the re-mapping approach will produce an estimate of t_{4i}. Estimates for infra-marginal thresholds can be derived in analogous fashion.

As for the PBE models discussed above, we consider the assumptions required for the Random Valuation Approach as overly restrictive. It imposes stringent constraints on the location of $\mu_i = E(v_i)$ relative to the decision thresholds. Moreover, these *same* restrictions must be assumed to hold for all individuals in the sample to identify the model. Consider, for example, the symmetry assumption imposed by Wang [7] and Alberini et al. [11] for some of their sub-models, i.e. $c_i = -b_i$. This implies $t_{2i} = \mu_i - b_i$, $t_{3i} = \mu_i + b_i$, or, equivalently, $b_i = (t_{3i} - t_{2i})/2$. This restricts μ_i to lie *precisely* between the second and third threshold, and always in the "NS" segment. This is a rather strong assumption in any application. It suggests that any difference in answer patterns (and thus threshold locations) across respondents is exclusively driven by differences in the shape of $g_i(v_i)$, and rules out any threshold heterogeneity due to other factors, such as differences in the respondent's interpretation of the qualitative response categories. Similar concerns arise for alternative restrictions of these limiting distances.

For these reasons our proposed Latent Thresholds Estimator (LTE) abstracts from any attempt to estimate $E(v_i)$. Conceptually, it is most closely related to re-mapping approach of Welsh and Poe [9]. However, instead of the independent estimation of a single threshold, we aim for the simultaneous estimation of all four thresholds. Moreover, we ex ante allow these thresholds to be jointly distributed via common unobservables (from the researcher's perspective). Our approach does not require any arbitrary re-mapping of polychotomous answers, and operates under a minimal set of assumptions on the distribution of $g_i(v_i)$ or the relative location of the thresholds. As will be illustrated in more detail below, the LTE nests the re-mapping estimator and allows for a formal examination of threshold restrictions underlying the Random Valuation Estimator.

III) Econometric Framework

Consider an MBUC format with $j = 1 \ldots J$ bid levels and $m = 1 \ldots M$ response categories. This implies that there are $T = M - 1$ decision thresholds, one for each category transition. Let each threshold be a simple linear function of observables and a normally distributed additive error term, i.e.

$$t_{ti} = \mathbf{x}_{ti}'\boldsymbol{\beta}_t + \varepsilon_{ti}, \qquad \varepsilon_{ti} \sim n\left(0, \sigma_t^2\right), \quad i = 1 \ldots N, t = 1 \ldots T. \tag{8}$$

One advantage of our framework over Alberini et al's [11] Random Valuation Model is that in theory all threshold functions are fully identified by the exogenous bids, such that the contents of $\mathbf{x}_{it}$ can remain unchanged, and marginal effects can be allowed to vary across thresholds (thus the *t*-subscript for $\boldsymbol{\beta}_t$). At the individual (panel) level the full model with correlated thresholds can be written as

$$\mathbf{t}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$
$$\mathbf{t}_i = \begin{bmatrix} t_{1i} \\ t_{2i} \\ \vdots \\ t_{Ti} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{1i}' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2i}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_{Ti}' \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_T \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{1i} \\ \vdots \\ \varepsilon_{Ti} \end{bmatrix} \sim n(\mathbf{0}, \boldsymbol{\Sigma}) \tag{9}$$

If thresholds were observed, equation (9) would describe a basic Seemingly Unrelated Regression (SUR) model.

However, instead of the actual thresholds we only observe a series of bid / response combinations from each survey participant. Let y_{ij} be the observed response by individual *i* when confronted with bid *j*. If we sort the response options from most affirmative (usually "DY") to most disapproving (usually "DN") and code them as increasing integers we have $y_{ij} \in \{1, 2, \cdots, M\}$. We observe $y_{ij} \mid b_{ij} = m$ if the *cdf* of the underlying valuation v_i at bid b_{ij} falls between the *cdf* at the preceding and the following threshold, i.e. $G(t_{m-1,i}) < G(b_{ij}) < G(t_{m,i})$. Since we do not wish to explicitly specify the value distribution, we re-express this condition directly in terms of the location of b_{ij} vis-à-vis the two adjacent thresholds, i.e.

$$y_{ij} = m \text{ if } \quad t_{m-1,i} < b_{ij} < t_{mi} \tag{10}$$

For the most affirmative response, the left bound in (10) will generally be negative infinity. It could also be zero if a given sample includes only confirmed program supporters, i.e. individuals who would always answer DY to a bid of zero. This will be the case for our first empirical application. For $y_{ij} = M$ the upper bound of (10) will usually be infinity.

Using the entire series of *J* observed responses for person i, collected in vector $\mathbf{y_i}$, it is then straightforward to elicit a set of *T* location restrictions for the thresholds, i.e.

$$pr(\mathbf{y_i}) = pr\begin{bmatrix} b_{1i,l} < t_{1i} < b_{1i,u} \\ b_{2i,l} < t_{2i} < b_{2i,u} \\ \vdots \\ b_{Ti,l} < t_{Ti} < b_{Ti,u} \end{bmatrix} = \Phi(\mathbf{X_i}\boldsymbol{\beta}, \boldsymbol{\Sigma}; R_i) \tag{11}$$

where (with slight abuse of notation) $\Phi(.)$ denotes the *cdf* of the truncated multivariate normal density with mean $\mathbf{X_i}\boldsymbol{\beta}$, variance matrix $\boldsymbol{\Sigma}$, and truncation region R_i implicitly defined by the T boundary conditions. Vector $\boldsymbol{\beta}$ comprises all T sets of threshold coefficients. Bids $b_{ti,l}$ and $b_{ti,u}$ denote, respectively, the *relevant* lower and upper bounds for a given threshold. The relevant bounds are the bids closest to a given threshold. Due to the possibility of repeated infra-marginal answers (RIAs, see above), not all bids offered to a given respondents will be relevant. Furthermore, it is possible for several thresholds to share one or both bounds if the observed answer pattern for a given individual does not traverse the entire *M*-dimensional response space, i.e. if one or more response categories are skipped. For those cases we impose the general ranking restriction $t_{1i} < t_{2i} < \ldots < t_{Ti}$ in our estimation framework. Examples for the identification of relevant bounds under different response scenarios are given in Appendix A.

The likelihood contribution by person *i* is thus given as

$$p(\mathbf{y_i} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X_i}) = \Phi(\mathbf{X_i}\boldsymbol{\beta}, \boldsymbol{\Sigma}; R_i) I(t_{1i} < t_{2i} < \cdots < t_{Ti}) \tag{12}$$

In theory, the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ could be estimated via Maximum Likelihood techniques. We prefer a Bayesian approach primarily for the following two reasons: (i) The effective sample size for threshold identification in one of our applications is relatively small, preempting the interpretation of estimation results in the light of classical asymptotic theory, and (ii) while it would be computationally challenging to impose the simultaneous threshold boundary and ranking conditions in an MLE framework, our Bayesian Gibbs Sampler can handle these restrictions in a straightforward fashion.

A Bayesian approach requires the specification of priors for all model parameters. We choose the standard multivariate normal priors for $\boldsymbol{\beta}$ and an inverse Wishart (*IW*) prior for the elements of $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\beta} \sim mvn(\boldsymbol{\mu_0}, \mathbf{V}_0)$, $\boldsymbol{\Sigma} \sim IW(v_0, \mathbf{S_0})$, where v_0 and $\mathbf{S_0}$ are the degrees of freedom and scale matrix, respectively. The *IW* density is parameterized such that $E(\boldsymbol{\Sigma}) = (v_0 - k_r - 1)^{-1} \mathbf{S_0}$. When combined with the likelihood function, these priors yield tractable conditional posterior densities. We further improve the speed and efficiency of our posterior simulator (Gibbs Sampler) by augmenting the model with draws of the unknown thresholds. A general discussion of the merits of this technique of *data augmentation* is given in Tanner and Wong [18] and van Dyk and Meng [19]. The augmented posterior distribution will thus be proportional to the priors times the augmented likelihood, i.e.

$$p\left(\boldsymbol{\beta},\boldsymbol{\Sigma},\left\{\mathbf{t_i}\right\}_{i=1}^{N}\mid\mathbf{y},\mathbf{X}\right)\propto p(\boldsymbol{\beta})\,p(\boldsymbol{\Sigma})\,p\left(\left\{\mathbf{t_i}\right\}_{i=1}^{N}\mid\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{y},\mathbf{X}\right)p\left(\mathbf{y}\mid\left\{\mathbf{t_i}\right\}_{i=1}^{N}\right) \tag{13}$$

The Gibbs Sampler draws consecutively and repeatedly from the conditional posterior distributions $p\left(\boldsymbol{\beta}\mid\boldsymbol{\Sigma},\left\{\mathbf{t_i}\right\}_{i=1}^{N},\mathbf{y},\mathbf{X}\right)$, $p\left(\boldsymbol{\Sigma}\mid\boldsymbol{\beta},\left\{\mathbf{t_i}\right\}_{i=1}^{N},\mathbf{y},\mathbf{X}\right)$, and $p\left(\left\{\mathbf{t_i}\right\}_{i=1}^{N}\mid\boldsymbol{\beta},\boldsymbol{\Sigma},\mathbf{y},\mathbf{X}\right)$. Posterior inference is based on the marginals of the joint posterior distribution $p(\boldsymbol{\beta},\boldsymbol{\Sigma}\mid\mathbf{y},\mathbf{X})$. As an additional benefit, our algorithm also returns all threshold draws for each respondent. The detailed steps of the posterior simulator and the Matlab code to implement this model are available from the authors upon request.

Empirical Application

Glen Canyon Pilot Study Data

Our first application uses data from the 1994 Glen Canyon (GC) Pilot Study. Details for this study are given in Welsh et al. [14].[4] The general aim of the GC non-market valuation project was to elicit stakeholders' *WTP* for reducing environmentally harmful fluctuations of water levels of the Colorado River due to varying discharges from Glen Canyon Dam. The pilot study used an MBUC format with 13 bid amounts ($0.1, $0.5, $1, $5, $10, $20, $30, $40, $50, $75, $100, $150, and $200) and a response scale of DY, PY, NS, PN, and DN.[5] It was implemented via several versions that differed in fluctuation scenario, target population, and the type of information provided to respondents. We combine all data associated with a "seasonally adjusted steady flow" scenario (versions 3, 5, 6, 8, and 9), yielding an original sample size of 384 individuals. After eliminating observations with missing bid-responses or other key variables we retain 370 observations, for a total of 370 x 13 = 4810 observed bid/response combinations. Following the original study, all of these individuals had been screened to support the environmental program at no cost. Thus our general lower bound for all thresholds is zero. This bound becomes active for cases where the first threshold is not bounded from below by one of the offered bids, i.e. where a respondent's answer to the second lowest bid is not DY.

A subset (28%) of the data stems from residents living in the market area served by GC-generated electrical power. The remaining observations flow from a nation-wide sample. It was hypothesized that market participants would exhibit lower *WTP* for the environmental benefits of reduced flow fluctuations, as they would be exposed to resulting higher energy prices. We capture this sub-segment with an indicator variable "*market*". Another subset of participants (20%) from the national sample received a truncated information brochure that omitted a listing of the economic drawbacks of flow regulation for some stakeholders. We identify this group via the binary indicator "*empathy*", taking a value of "1" if costs to others were dropped from the information pamphlet. Contrary to expectations, the original authors found that *WTP* decreased significantly for this sub-group. Our richer analysis sheds additional light on this issue, as shown below.

The remaining variables in our GC model are as follows: (i) a standardized knowledge score ("*know*") based on respondents' answers to a short quiz at the beginning of the main survey instrument that covered the contents of the information brochure, (ii) a standardized factor-analytical score summarizing participant's relative preference of economic security over environmental protection ("*econ*") , and (iii) annual household income in $1000s. Table 1 captures the salient features of our sample. From an econometric perspective, the most important characteristic of the GC data is that all thresholds and threshold covariances are fully identified via reasonably large sub-samples. For example, there are 276 cases (out of 370) that provide both DY and PY answers, thus identifying the first threshold and its variance. Similarly, there are 204 individuals that exhibit the full triplet of DY, PY, and NS answers, thus jointly identifying the first *two* thresholds and their covariance (last column of Table 1). This limits cases where thresholds share common bounds, and thus enhances the efficiency of our estimator. As depicted in the bottom half of Table 1, knowledge scores range from -3.53 (poor score on the introductory quiz) to 0.77 (close to perfect score). Economic security scores cover a range of -1.18 (strong preference for environmental health over economic security) to 2.47 (strong preference for economic security over environmental health). The average household income (in 1994 dollars) lies in the mid-50 thousands.

Rangeland Restoration Pilot Study Data

Our second application is a 2005 pilot study on Nevada residents' *WTP* to reduce the risk of wildfires via reseeding and other restoration efforts targeting Great Basin rangelands. Details are given in Kobayashi and Rollins [16]. We focus here on their "obtain gain" scenario, which was stipulated to reduce the wildfire risk by 50% throughout the State via a comprehensive vegetation management program. We further restrict our attention to the subset of respondents that received a 9-bid MBUC elicitation format (128 valid observations). After eliminating cases with deficiencies in key variables we obtain a final sample of 113 individuals (1017 observed bid-responses) suitable for this analysis. The MBUC bids are $0, $1, $12, $31, $52, $83, $114, $157, and $282. The wording of the MBUC question and the response categories are the same as for the GC study. Contrary to the GC applications, retained individuals had not been identified as conditional (zero-cost) program supporters. Therefore, it is possible for this sample to issue a response other than DY for the zero-bid. As a consequence, the general lower bound for all thresholds not bounded by an actual bid is negative infinity. This is an important deviation from the GC case. As our results will show, the distribution of all identified thresholds enter the negative domain for a substantial proportion of the underlying population.

The main characteristics of the Rangeland Restoration (RR) sample are given in Table 2. Contrary to the GC case, not all thresholds are identified for the RR data set. Specifically, nobody in the sample transitioned from NS to PY, or from PY to DN.[6] This forced us to combine the inframarginal response categories into a single group and settle for the estimation of the two outer thresholds. As can be seen from the Table, the identifying sample sizes for these two thresholds are relatively small (62 and 38 cases, respectively). The covariance is only identified by 31 observations. However, in our Bayesian estimation framework these relevant sample sizes are sufficient to induce substantial posterior learning for this application.

Due to the sparse structure of our data we limit the number of regressors in the threshold equations to three (in addition to a constant term). These explanatory variables are (i) a factor-analytically derived preference score for rangeland "*quality* " (higher score = stronger preferences for environmentally healthy rangelands), (ii) a 0/1 indicator for the sub-sample of respondents (close to 40%) that received a survey booklet with additional information on the detrimental impacts of wildfires and the causes of the intensifying wildfire cycle in the great Basin ("*info*"), and (iii) annual household income in units of one thousand 2005 dollars.

Relevant Bounds and Threshold Identification

Table 3 provides a closer look at the distribution of observations across relevant bounds for each threshold for the GC application. The first three columns list the lower bound, upper bound, and inter-bound range for each set of relevant bounds observed in the data. Each possible consecutive bid interval figures as relevant bound for the first three thresholds. For the fourth threshold (T4) there are no observations that falls within the lowest two bid segments. As has been standard practice in MBUC applications, inter-bid ranges increase substantially over the entire set of bids, here from $0.1 for the lowest bracket to $50 for the highest two brackets. As we will discuss in more detail below, this common practice, likely driven by the traditional focus on the first or "certainty" threshold (T1), can impede the efficient estimation of higher thresholds.

For each threshold and relevant bound the Table depicts the number of fully identified cases ("fi" column), the number of not fully identified cases ("nfi"), and the total number of observations for which a given pair of bids forms the relevant bound. The "nfi" cases include both observations for which a given bound is shared by two or more other thresholds for the same individual, and observations for which one or both bounds are infinite. For example, looking at the first row in the T1 triplet of columns, there are 22 individuals in our sample that respond with DY to a bid of $0 and PY to a bid of $0.1. Consequently, the $0 / $0.1 pair becomes the relevant set of bounds for T1 for these cases. For three individuals in this group other thresholds in addition to T1 also fall into the same relevant bracket. This implies that these individuals skipped the PY and perhaps additional higher response categories, i.e. they answered DY to $0, and NS , PN, or DN to $0.1. As a result, T1, T2 and perhaps even higher thresholds all fall within the $0 / $0.1 relevant bracket. The total number of fully identified cases for each thresholds also corresponds to the respective entry in the third-to-last column of Table 1.

Perhaps the most striking feature of Table 3 is the wide spread of observations across virtually the entire set of relevant bounds for all four thresholds. For example, focusing again on T1, a given individual might switch from DY to PY at virtually any point between $0.1, and $100, although the bulk of switches occur in the $1 - $30 range. Similar patterns can be observed for T2 through T4. This highlights the pronounced heterogeneity in the range (and likely shape) of the underlying value distribution across individuals. From an estimation perspective this both desirable, as it aids in the identification of marginal effects of regressors, and problematic, as the unobservable effects in the threshold equations (8) and therefore threshold variances can be expected to be large given our rather sparse set of observables.

The second most important feature of Table 2 is the relatively large share of observations that fall into the $200 – to – infinity category, especially for T3 and T4 (96 and 164 cases, respectively). As our results will show this both inflates threshold variances and decreases the posterior efficiency for parameters related to these thresholds. In retrospect, a few additional bids at the upper end of the spectrum would have likely been extremely beneficial for this application. As we will show below, our LTE estimator can also be used for guidance in bid design.

Table 4 captures analogous sample statistics for the RR application. In general, the distributional pattern of thresholds across relevant bid brackets is similar to the one observed for the GC application. As in the former case, a substantial proportion of individuals issued a more affirmative response than DY even at the highest bid. This leaves the upper bound for T4 undefined for 64 of 113 cases. It should be noted, however, that the two highest relevant bins with finite bounds are only sparsely populated. This suggests a strong dichotomy in this sample, with approximately 50% of participants switching to DN by $114, and the other half reluctant to declare certain rejection along the entire bid range. We will revisit these bid design issues in our discussion of results and in the concluding section.

Estimation Results

We estimate all models using the following vague but proper parameter settings for our priors: $\boldsymbol{\mu}_0 = 0, \mathbf{V}_0 = 10, v_0 = T + 2,$ and $\mathbf{S}_0 = \mathbf{I}_{\mathbf{k}_r}$.[7] For each application we also estimate a version with independent thresholds, where $\boldsymbol{\Sigma}$ is restricted to a diagnonal matrix. For these cases we specify inverse-gamma (ig) priors for the T variance terms with shape and scale parameters set to ½. We first test all models using simulated data to assure the accuracy of our computational algorithm. For all actual estimation runs we discard the first 2000 draws generated by the Gibbs Sampler as "burn-ins", and retain the following 1000 draws for posterior inference. We evaluate the performance of the posterior simulator using Geweke's [20] convergence diagnostics (CD), and inefficiency (IEF) scores as described in Chib [21]. The CD scores clearly indicate convergence for all our models.

The IEF scores, which convey the degree of (undesirable) autocorrelation in the series of posterior draws, range from the single digits (i.e. near-independence) for most slope coefficients and variance terms to 20-40 for the somewhat less clearly identified variances and covariances associated with higher thresholds.[8]

Table 5 summarizes results for the GC application. The left half of the Table shows posterior means and standard deviation for threshold coefficients, while the right half depicts posterior results for the elements of $\boldsymbol{\Sigma}$ in terms of standard deviations and correlations. Within each half the first two columns pertain to the fully correlated model and the last two columns to the model with independent error terms. The estimates for the first threshold (T1) flowing from the independent model can be interpreted as those that would be obtained from a binary estimation framework that treats all responses other than DY as DN.

We can immediately realize that the error terms in the full model exhibit close to perfect correlation for all six threshold pairs. Moreover, these correlation terms are estimated with very high precision as indicated by the negligible magnitudes of the respective posterior standard deviations. This casts serious doubt on the legitimacy of the independent model. To allow for a more rigorous comparison we compute the marginal likelihood for each case using the simulation method outlined in Chib [22]. These terms are given in log form toward the bottom of the right half of Table 5. The difference between the marginal likelihood of the full model and the independent model yields a logged Bayes Factor (BF) of 760.4 (last row of the Table). Using the interpretation thresholds for BFs given in Kass and Raftery [23] this result provides "decisive" evidence in favor of the full model. It is also evident from the Table that the independent model produces higher posterior standard deviations for virtually all slope coefficients. The full model, with its ability to exploit threshold linkages via unobservable effects, uses the data more efficiently. It is thus better suited to update model priors and reduce posterior uncertainty. We will henceforth focus on the full model in our examination of posterior results.

Not surprisingly given our discussion of bid ranges and relevant bounds above the posterior means of estimated threshold standard deviations increase from the lowest threshold (57.44) to the highest (244.08). The posterior standard deviations for threshold coefficients follow the same pattern, suggesting a general trend of decreasing posterior precision as one moves from the lowest to the highest threshold. However, some slope coefficients, such as those corresponding to income and the economic security score (*econ*) are estimated with relatively high precision across all four cases.

Since the set of regressors is held constant across all thresholds we can distinguish between variables that exert an even shift on the entire value distribution from those that have varying effects across thresholds. Logically, the former can be interpreted as regressors that primarily affect the *expectation* of the value distribution, while the latter are likely to affect both the expectation and variance. In our case virtually all explanatory variables exhibit noticeable changes in posterior means across thresholds. For example, the *empathy* indicator shifts the lower thresholds to the left and the highest threshold to the right by comparable magnitude. Thus, omitting reminders of costs to others in a survey version does not necessarily lead to lower expected *WTP* (the puzzling conclusion reached by the original authors), but rather increases the *spread* of the value distribution. This subtle but important difference in inference becomes only apparent when all four thresholds are estimated. A similar finding holds for the *market* indicator – participants from the market area served by GC power have lower estimates for T1 and T2 (i.e. switch from DY to PY and from PY to NS at lower bids), but are also more reluctant to enter the DN category. Thus, market participants exhibit a wider spread in underlying valuation than non-market respondents. In contrast, the direction of marginal effects remains unchanged across thresholds for the knowledge score (*know*), the economic security score (*econ*) and income. A better understanding of GC power generation and related environmental issues, as measured by *know*, shifts the entire value distribution to the right, with a slightly increasing trend across thresholds. On the other hand, individuals with high preference for economic security over environmental conservation, i.e. a high econ score, can be associated with both a strong leftward shift of their value distribution, and a tighter overall distribution. The latter insight stems from the fact that the leftward shift for the highest threshold (-17.22) is substantially larger than the leftward shift for the lowest threshold (-5.76). This reduces the overall range of the underlying valuation. As expected, income exhibits an efficiently estimated positive effect on all thresholds. Since its marginal effect also increases from lowest to highest threshold, we can infer that higher income also translates into higher variability of underlying valuation.

Table 6 summarizes estimation results for the RR application. In this case, the posterior mean for error correlation between T1 and T4 is essentially zero. This is not all that surprising given our recoding of all infra-marginal responses into a single "slush category", which likely weakened any existing correlation patterns. A decisive Bayes Factor of 32.7 lends formal support to the independent specification. Given the near-zero error correlation, the full and independent models generate very similar results for coefficient estimates. The income effect is perfectly analogous to the one discussed for the GC case. Interestingly, a high score on the preference ladder for rangeland environmental health, as measured by the *quality* indicator, translates into a dramatic rightward shift for T1, but has a comparatively smaller positive effect on T4. Thus, individuals with strong environmental preferences for the rangeland ecosystem have a much tighter value distribution and likely a much higher expected value than others. Perhaps the most important result for this application is the effect of providing more detailed information, measured by the *info* indicator. Contrary to expectations, the added information does not shift the entire value distribution to the right, but rather leads to an increased variability in values. This illustrates how the LTE framework can be employed to test if information provision or other design treatments have the desired or expected effect on the underlying value distribution. In this case, it appears that the added information either confused some respondents or, more likely, split them into separate camps with respect to their preferences for the proposed vegetation management plan.

Aside from marginal effects of explanatory variables, the predicted *location* of thresholds will be of central interest in most applications. We generate posterior predictive distributions (PPDs) for some combinations of regressors by combining the corresponding settings for $\mathbf{X}_i$ with parameter draws flowing from the Gibbs Sampler. For details on the derivation and interpretation of PPDs see for example [24; 25]. Table 7 captures the posterior means and standard deviations for these PPDs for both the full and independent model of the GC application.

The posterior means for T1 and T2 generated by the independent model correspond closely to the authors' original estimates for *WTP* under different recoding approaches ([14], p. C-11). The most important result captured in Table 7 is the pronounced difference in posterior means between the full and the independent model. Specifically, the independent model tends to over-predict the lowest threshold and severely under-predicts the highest threshold. The latter shortcoming could have serious implications in policy applications where a conservative estimate of welfare *losses* is sought, which would logically shift the inferential focus to the upper end of the value distribution.

Figure 2 depicts PPDs for both applications and all thresholds for baseline regressor settings, with preference scores and income held at the sample mean. The Figure largely confirms the numerical results discussed above: Threshold variances increase from lowest to highest threshold, and all threshold distributions cover a relatively wide range of the underlying value distribution. The predictive expectations (vertical lines) for the first and last threshold describe the expected range of *WTP* for an individual with these baseline characteristics. In the GC case, the value distribution for a prototypical stakeholder can be expected to lie between $50, the predictive expectation of the lowest threshold and an upper bound of approximately $220, the predictive mean of the highest threshold. Similarly, for the RR application the *WTP* distribution for a baseline consumer type is expected to lie between $40 and $300. For both applications it is clear from the PPD of threshold four that the upper end of the value distribution may be substantially higher for some baseline individuals, perhaps as high as $500 in the GC case, and $800-900 for the RR case. For the RR application, a substantial segment of baseline individuals may have pronounced negative *WTP* for the proposed rangeland improvements, as is evident from the lower tail of the PPD for threshold one.

A Test for Threshold Symmetry

Our modeling framework also allows for an examination of the maintained assumption in Alberini et al.'s [11] primary Random Valuation Model that both inner and outer thresholds are equidistant to the expectation of the value distribution. We can cast this assumption as a linear model restriction, i.e $E\left(\left(t_{4i}-t_{3i}\right)-\left(t_{2i}-t_{1i}\right)\right)=0$ or, equivalently, a linear parameter restriction, i.e. $\left(\boldsymbol{\beta}_4-\boldsymbol{\beta}_3\right)-\left(\boldsymbol{\beta}_2-\boldsymbol{\beta}_1\right)=\mathbf{0}$. Based on Figure 1, such dual symmetry seems unlikely. We also perform a formal model comparison, which produces a logged Bayes Factor of 104.7 in favor of the unrestricted specification.[9] Therefore, this dual-symmetry assumption is not supported by the observed data for our application. A Random Valuation Model based on such an assumption would be mis-specified.

Conclusion

We propose a new estimator for MBUC data that utilizes all observed response patterns for the simultaneous estimation of the full set of underlying decision thresholds. Our Latent Threshold Estimator has several advantages over existing approaches that process MBUC data, or - more generally - data from value elicitation with uncertain response options. It does not require stringent ex ante restrictions on value distributions or threshold locations, and handles repeated infra-marginal responses in a straightforward manner. By allowing thresholds to be fully correlated our framework exploits linkages across thresholds via unobservable effects. This can produce measurable efficiency gains, as evidenced by our GC application.

More importantly, the LTE framework provides insights into the marginal effect of regressors that would remain undiscovered using other estimation strategies. It clearly highlights if a given regressor primarily shifts the entire value distribution (and thus its expectation), or if it also affects the spread of the distribution. This can be exploited to examine the impact of changes in survey format, such as the provision of additional information, or altering the scope or scale of a proposed policy intervention.

Furthermore, by returning the full distribution of the outer thresholds, the LTE approach provides clear and explicit guidance as to the variability in the range of underlying values that can be expected for a given stakeholder population. This can be very helpful in devising efficient bid designs for final survey versions. Importantly, it opens the door for MBUC designs aimed at the accurate estimation of the *upper* decision threshold. This, in turn, broadens the applicability of the MBUC format to environmental policy scenarios with a primary focus on loss prevention or damage assessment.

Naturally, important caveats remain. Our modeling framework is fully anchored in the assumption that individuals are truly *unable* to assign a point value estimate to a given non-market amenity or service due to latent and potentially permanent uncertain factors. Thus, we rule out the possibility that respondents are *unwilling* to exert sufficient effort to zoom in on a single value, a concern raised by Alberini et al. [11]. Neither do we address the issue of bid ordering effects on measured values examined in that study, although our LTE framework could potentially be useful in identifying and controlling for such effects.

In general, our modeling assumptions appear reasonable and are certainly less stringent than those required for alternative estimation strategies. We believe that the LTE approach is a natural and to date overlooked extension of the MBUC framework. It has the potential to substantially broaden the applicability of this elicitation approach. We also conclude that the common practice of using MBUC data solely to derive an estimate for a single threshold, interpreted as point estimate of *WTP*, leaves useful information untapped and can produce misleading results if thresholds are highly correlated.

Appendix A: Relevant and shared bounds in the LTE framework

Case 1: Full threshold identification

This marks the ideal scenario in which all possible answer categories are observed for a given respondent. It leads to finite and unique lower and upper bounds for each threshold. For example, consider again the response pattern used previously in the discussion of [10]'s DUDE estimator:

$$b_1 \rightarrow DY, b_2 \rightarrow PY, b_3 \rightarrow PY, b_4 \rightarrow PY, b_5 \rightarrow NS, b_6 \rightarrow PN, b_7 \rightarrow DN \quad \text{(A1)}$$

This implies the following boundary conditions for thresholds:

$$b_1 < t_1 < b_2, \quad b_4 < t_2 < b_5, \quad b_5 < t_3 < b_6, \quad b_6 < t_4 < b_7 \quad \text{(A2)}$$

Bid b_3 contributes no new information to the identification of threshold locations and becomes irrelevant. Also, the general ranking condition $t_{1i} < t_{2i} < t_{3i} < t_{4i}$ is automatically assured through the increasing bid amounts.

Case 2: Partial threshold identification

Now consider the following response pattern:

$$b_1 \rightarrow DY, b_2 \rightarrow PY, b_3 \rightarrow PY, b_4 \rightarrow PY, b_5 \rightarrow PN, b_6 \rightarrow PN, b_7 \rightarrow DN \quad \text{(A3)}$$

Compared to the previous case, the NS category has been skipped. The resulting boundary conditions are:

$$b_1 < t_1 < b_2, \quad b_4 < t_2 < b_5, \quad b_4 < t_3 < b_5, \quad b_6 < t_4 < b_7 \quad \text{(A4)}$$

Thresholds two and three share the same upper and lower bounds. In our computational algorithm we handle this case by drawing such partially identified thresholds simultaneously from their shared interval, and imposing the relevant ranking condition ex post.

Notes:

[1] A general discussions of why such uncertainty may arise is given i.a. in [1] and [2].

[2] In this study we follow [7], [10], and [11] and interpret NS as a response that describes an underlying value segment that is wedged between the segments for PY and PN.

[3] In the standard case, the re-mapping is imposed uniformly for the entire sample (e.g. *all* PYs are recoded to DYs, etc). More recent contributions aim at a subject-specific recoding (e.g. [12], [16]).

[4] We thank Mary Evans and V. Kerry Smith for providing this data set and all accompanying documentation.

[5] The exact wording of the MBUC question was: "How would you vote on this proposal if passage of the proposal would cost your household these amounts **every year** for the foreseeable future?"

[6] The main reason for this phenomenon is probably that the response table listed the NS category in the last column as opposed to a column wedged between the PY and PN options. Once a respondent "jumped" to that column, she did not return to the other categories in the Table. A detailed discussion of this and other possible MBUC formatting effects is beyond the scope of this study.

[7] "Proper" prior distributions are those that integrate to one over their entire range. This characteristic is required for the derivation of Bayes Factors for model comparison, an important consideration in our case. "Vague" refers to the fact that the distribution has a relatively large variance, which preempts substantial prior density mass for any specific segment of the distribution range. This reflects the absence of any existing information to aid in the construction of priors.

[8] Detailed performance scores for all models are available from the authors upon request.

[9] We can derive this Bayes Factor of via the Savage-Dickey Density Ration (SDDR) . Details of these computations are available upon request.

Acknowledgements:

We thank Douglass Shaw and J. Scott Shonkwiler for helpful comments.

References:

[1] S.L. Shaikh, L. Sun, and G.C.v. Kooten, Treating respondent uncertainty in contingent valuation: A comparison of empirical treatments. Ecological Economics 62 (2007) 115-125.

[2] S. Akter, R. Brouwer, L. Brander, and P.v. Beukering, Respondent uncertainty in a contingent market for carbon offsets. Ecological Economics 68 (2009) 1858-1863.

[3] C.-Z. Li, and L. Mattsson, Discrete choice under preference uncertainty: An improved structural model for contingent valuation. Journal of Environmental Economics and Management 28 (1995) 256-269.

[4] P.A. Champ, R.C. Bishop, T.C. Brown, and D.W. McCollum, Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management 33 (1997) 151-162.

[5] E.R. Ekstrand, and J. Loomis, Incorporating respondent uncertainty when estimating willingness to pay for protecting critical habitat for threatened and endangered fish. Water Resources Research 34 (1998) 3149-3155.

[6] R.C. Ready, J.C. Whitehead, and G. Blomquist, Contingent valuation when respondents are ambivalent. Journal of Environmental Economics and Management 29 (1995) 181-196.

[7] H. Wang, Treatment of "don't know" responses in contingent valuation surveys: A random valuation model. Journal of Environmental Economics and Management 32 (1997) 219-232.

[8] R.C. Ready, S. Navrud, and W.R. Dubourg, How do respondents with uncertain willingness to pay answer contingent valuation questions? Land Economics 77 (2001).

[9] M.P. Welsh, and G.L. Poe, Elicitation Effects in contingent valuation: comparison to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management 36 (1998) 170-185.

[10] M.F. Evans, N.E. Flores, and K.J. Boyle, Multiple-bounded uncertainty choice data as probabilistic intentions. Land Economics 79 (2003) 549-560.

[11] A. Alberini, K. Boyle, and M. Welsh, Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management 45 (2003) 40-62.

[12] S.B. Caudill, and P.A. Groothuis, Modeling hidden alternatives in Random Utility Models: An application to "don't know" responses in contingent valuation. Land Economics 81 (2005) 445-454.

[13] W.M. Hanemann, J. Loomis, and B. Kanninen, Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics 73 (1991) 1255-1263.

[14] M.P. Welsh, R.C. Bishop, M.L. Phillips, and R.M. Baumgartner, GCES non-use value study, Report prepared for the Glen Canyon Environmental Studies Non-Use Value Committee, U. S. Bureau of Reclamation, 1995.

[15] S. Akter, J. Bennett, and S. Akhter, Preference uncertainty in contingent valuation. Ecological Economics 67 (2008) 345-351.

[16] M. Kobayashi, and K. Rollins, Sensitivity of *WTP* estimates to definition of "yes": Reinterpreting expressed response intensity. Agricultural and Resource Economics Review (forthcoming).

[17] T.A. Cameron, G.L. Poe, R.G. Ethier, and W.D. Schulze, Alternative non-market value-elicitation methods: Are the underlying preferences the same? Journal of Environmental Economics and Management 44 (2002) 391-425.

[18] M.A. Tanner, and W.H. Wong, The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82 (1987) 528-550.

[19] D.A.v. Dyk, and X.L. Meng, The art of data augmentation. Journal of Computational and Graphical Statistics 10 (2001) 1-50.

[20] J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in "Bayesian Statistics 4" (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Eds.), Oxford University Press, Oxford, UK (1992).

[21] S. Chib, Markov Chain Monte Carlo methods: Computation and inference, in "Handbook of Econometrics" (J.J. Heckman and E. Leamer, Eds.) , Elsevier (2001).

[22] S. Chib, Marginal likelihood from the Gibbs Sampler. Journal of the American Statistical Association 90 (1995) 1313-1321.

[23] R.E. Kass, and A.E. Raftery, Bayes Factors. Journal of the American Statistical Association 90 (1995) 773-795.

[24] K. Moeltner, K. Boyle, and R. Paterson, Meta-analysis and benefit transfer for resource valuation - Addressing classical challenges with Bayesian modeling. Journal of Environmental Economics and Management 53 (2007) 250 - 269.

[25] K. Moeltner, and R.S. Rosenberger, Meta-regression and benefit transfer: Data space, model space, and the quest for 'optimal scope'. B.E. Journal of Economic Analysis & Policy, Contributions, 2008.

Table 1: Descriptive Statistics for the GC Application

Response Statistics

response category	obs	share of total	thresholds	identification obs	covariances	
DY	2349	0.49	t1	276	t1/t2	204
PY	638	0.13	t2	204	t1/t3	302
NS	469	0.10	t3	173	t1/t4	304
PN	458	0.10	t4	152	t2/t3	147
DN	896	0.19			t2/t4	254
					t3/t4	122
Total	4810	1				

Sample Statistics

	mean	std	min	max	% "1"	obs.
empathy indicator	-	-	-	-	20.00%	370
market indicator	-	-	-	-	28.11%	370
knowledge score	0	1	-3.53	0.77		370
econ security score	0	0.79	-1.18	2.47		370
income ($1000)	55.02	31.31	10	150		370

Table 2: Descriptive Statistics for the RR Application

Response Statistics

response category	obs.	share of total	thresholds	identification obs.	covariances	obs.
DY	367	0.36	t1	62	t1 & t4	31
PY/NS/PN	458	0.45	t4	38		
DN	192	0.19				
Total	1017	1				

Sample Statistics

	mean	std	min	max	% "1"	obs.
quality score	0	0.94	-3.5	0.77		113
information version	-	-	-	-	39.82%	113
income ($1000)	73.38	60.55	8	250		113

Table 3: Threshold Identification for the GC Application

relevant bounds			T1(DY->PY)			T2(PY->NS)		
lower	upper	range	fi	nfi	total	fi	nfi	total
0	0.1	0.1	19	3	22	0	3	3
0.1	0.5	0.4	10	0	10	1	0	1
0.5	1	0.5	10	1	11	4	2	6
1	5	4	26	11	37	11	17	28
5	10	5	28	4	32	12	6	18
10	20	10	42	11	53	19	16	35
20	30	10	34	11	45	29	18	47
30	40	10	32	3	35	15	7	22
40	50	10	13	3	16	28	6	34
50	75	25	25	9	34	32	14	46
75	100	25	15	3	18	21	10	31
100	150	50	19	8	27	22	10	32
150	200	50	3	0	3	10	2	12
200	Inf	Inf	0	27	27	0	55	55
column total			276	94	370	204	166	370
% of sample			75%	25%	100%	55%	45%	100%

relevant bounds			T3(NS->PN)			T4(PN->DN)		
lower	upper		fi	nfi	total	fi	nfi	total
0	0.1	0.1	0	1	1	0	0	0
0.1	0.5	0.4	1	0	1	0	0	0
0.5	1	0.5	0	1	1	0	1	1
1	5	4	2	13	15	0	9	9
5	10	5	3	5	8	2	4	6
10	20	10	15	13	28	4	8	12
20	30	10	22	14	36	9	5	14
30	40	10	13	7	20	22	4	26
40	50	10	16	7	23	14	3	17
50	75	25	26	13	39	20	5	25
75	100	25	35	10	45	31	4	35
100	150	50	25	12	37	32	8	40
150	200	50	15	5	20	18	3	21
200	Inf	Inf	0	96	96	0	164	164
column total			173	197	370	152	218	370
% of sample			47%	53%	100%	41%	59%	100%

fi = fully identified (no other threshold shares same bounds for a given individual)
nfi = not fully identified (threshold shares bounds with other thresholds for a given individual or bounds are not finite)

Table 4: Threshold Identification for the RR Application

relevant bounds			T1(DY->PY/NS/PN)			T4(PY/NS/PN->DN)		
lower	upper	range	fi	nfi	total	fi	nfi	total
-Inf	0	Inf	0	34	34	0	0	0
0	1	1	4	0	4	0	0	0
1	12	11	6	1	7	0	1	1
12	31	19	11	3	14	5	3	8
31	52	21	16	2	18	6	2	8
52	83	31	8	2	10	9	2	11
83	114	31	9	3	12	10	3	13
114	157	43	5	0	5	6	0	6
157	282	125	3	0	3	2	0	2
282	Inf	Inf	0	6	6	0	64	64
column total			62	51	113	38	75	113
% of sample			55%	45%	100%	34%	66%	100%

fi = fully identified (no other threshold shares same bounds for a given individual)
nfi = not fully identified (threshold shares bounds with other thresholds for a given individual or bounds are not finite)

Table 5: Estimation Results for the GC Application

	Full		Independent	
coeff.	post. mean	post. std	post. mean	post. std
T1:				
constant	0.24	3.27	20.42	4.80
empathy	-6.13	3.70	-11.78	5.70
market	-2.67	3.46	-7.56	5.40
know	2.26	2.01	6.03	2.73
econ	-5.76	2.30	-17.89	3.42
income	0.76	0.06	0.53	0.07
T2:				
constant	3.06	4.23	33.82	5.57
empathy	-8.18	4.48	-10.09	6.87
market	-2.57	4.22	-4.66	6.44
know	4.73	2.94	8.26	3.38
econ	-8.69	3.23	-24.10	4.14
income	1.32	0.10	0.74	0.08
T3:				
constant	6.11	5.54	45.95	6.43
empathy	-1.76	5.69	-5.08	7.56
market	-0.81	5.56	-2.59	7.07
know	4.42	4.26	5.86	4.17
econ	-10.92	4.61	-27.33	5.05
income	2.10	0.17	1.06	0.10
T4:				
constant	16.88	8.32	46.94	8.31
empathy	5.54	8.42	2.20	8.64
market	4.02	7.99	5.62	8.28
know	6.60	6.65	4.14	5.96
econ	-17.22	7.29	-28.20	6.77
income	3.45	0.32	2.01	0.17

	Full		Independent	
standard deviations / correlations	post. mean	post. std	post. mean	post. std
T1	57.44	2.40	52.45	1.99
T1/T2	0.96	0.01	-	-
T2	96.64	4.70	66.34	2.59
T1/T3	0.91	0.01	-	-
T2/T3	0.97	0.01	-	-
T3	151.08	9.50	83.41	3.59
T1/T4	0.87	0.02	-	-
T2/T4	0.95	0.01	-	-
T3/T4	0.98	0.00	-	-
T4	244.08	20.89	125.79	6.98

Model comparison	Full	Independent
log mL	3373.19	4133.58
log BF	760.39	

post. mean = posterior mean / post. std. = posterior standard deviation / coeff. = coefficients
log mL = logged marginal likelihood / log BF = logged Bayes Factor

Table 6: Estimation Results for the RR Application

	Full		Independent	
coeff.	post. mean	post. std	post. mean	post. std
T1:				
constant	-5.94	8.64	-5.00	7.97
quality	13.06	6.92	13.43	6.95
info	-1.22	8.45	-1.57	8.38
income	0.62	0.12	0.60	0.11
T4:				
constant	11.93	9.84	11.99	9.80
quality	3.10	9.35	2.42	10.00
info	2.06	9.75	1.75	9.98
income	3.19	0.44	3.30	0.46

	Full		Independent	
standard deviations / correlations	post. mean	post. std	post. mean	post. std
T1	87.09	8.08	89.01	8.29
T1/T4	0.05	0.15	-	-
T4	272.23	32.53	288.26	37.48

Model comparison

	Full	Independent
log mL	-529.22	-496.55
logBF		32.67

post. mean = posterior mean / post. std. = posterior standard deviation / coeff. = coefficients
log mL = logged marginal likelihood / log BF = logged Bayes Factor

Table 7: Posterior Predictive Results for the GC Application

			full model		independent model	
type	description		mean	std	mean	std
1	empathy, no market, mean know, mean econ, mean income	T1	48.16	31.80	50.79	34.87
		T2	84.81	49.97	67.14	46.45
		T3	130.90	79.43	88.75	58.05
		T4	215.91	134.02	119.70	85.28
2	empathy, **market**, mean know, mean econ, mean income	T1	47.77	32.40	49.18	35.12
		T2	86.36	54.02	64.55	44.74
		T3	137.88	86.23	88.59	60.41
		T4	230.45	140.49	121.32	83.98
3	**no empathy**, no market, mean know, mean econ, mean income	T1	43.91	31.13	46.23	34.19
		T2	79.05	50.97	63.01	43.94
		T3	131.55	81.43	88.06	58.60
		T4	226.74	136.98	120.56	84.95
4	empathy, no market, mean know, **high econ**, mean income	T1	43.88	30.29	47.02	33.56
		T2	79.68	49.59	60.73	45.39
		T3	126.41	79.28	78.24	56.79
		T4	207.60	133.07	116.37	82.13

Figure 1: Decision Thresholds for Polychotomous Responses and the Random Valuation Model

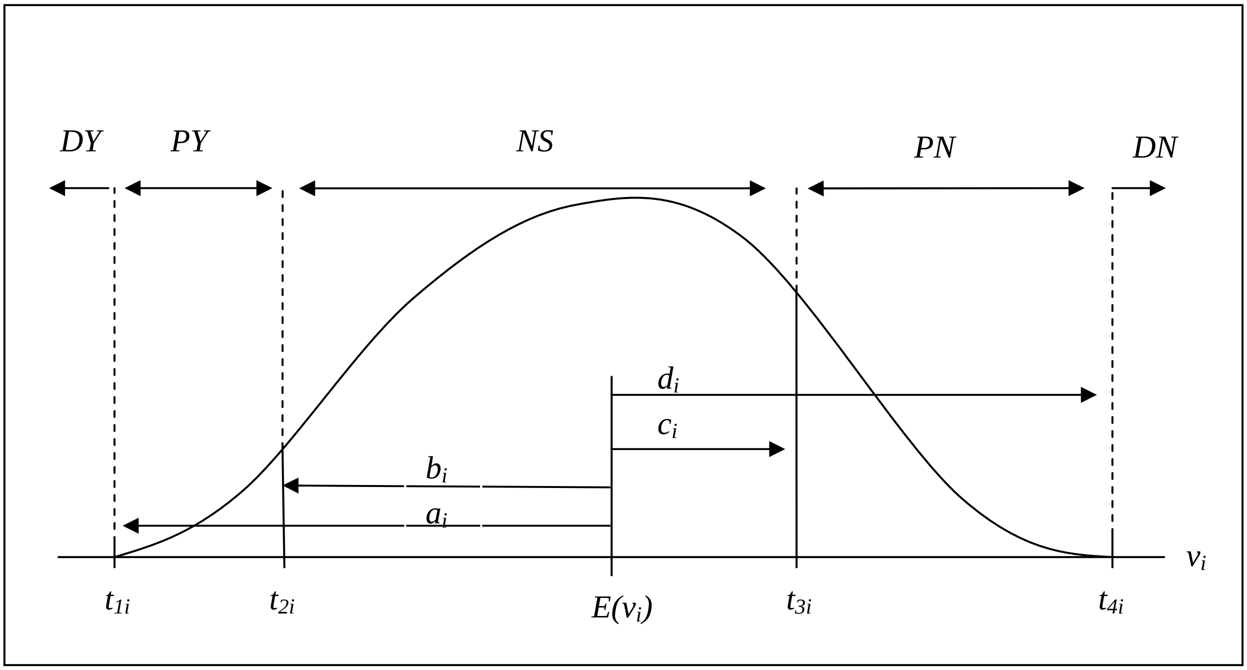
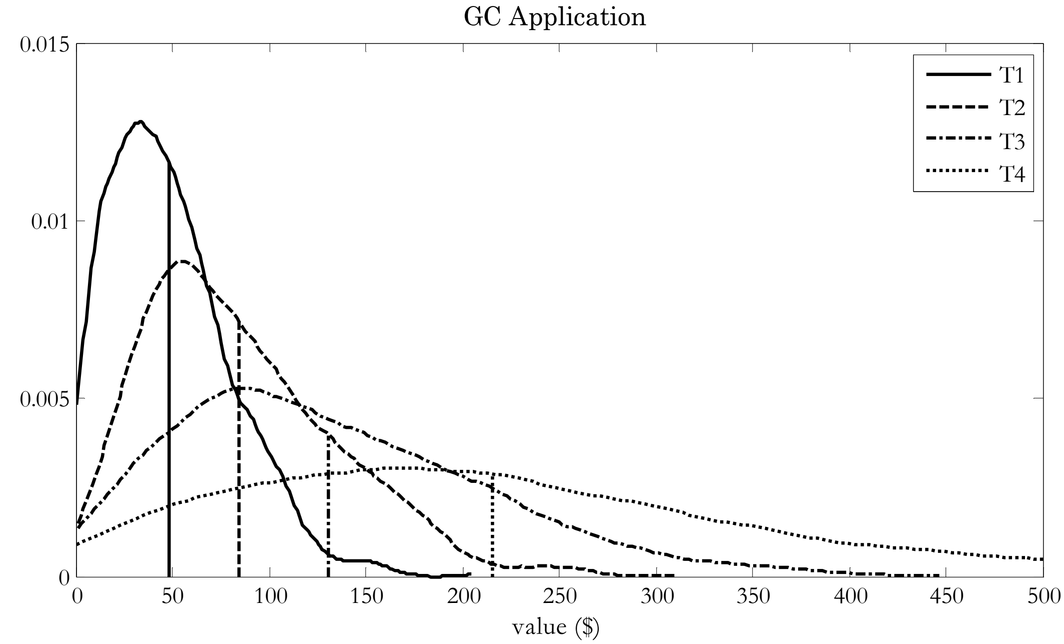
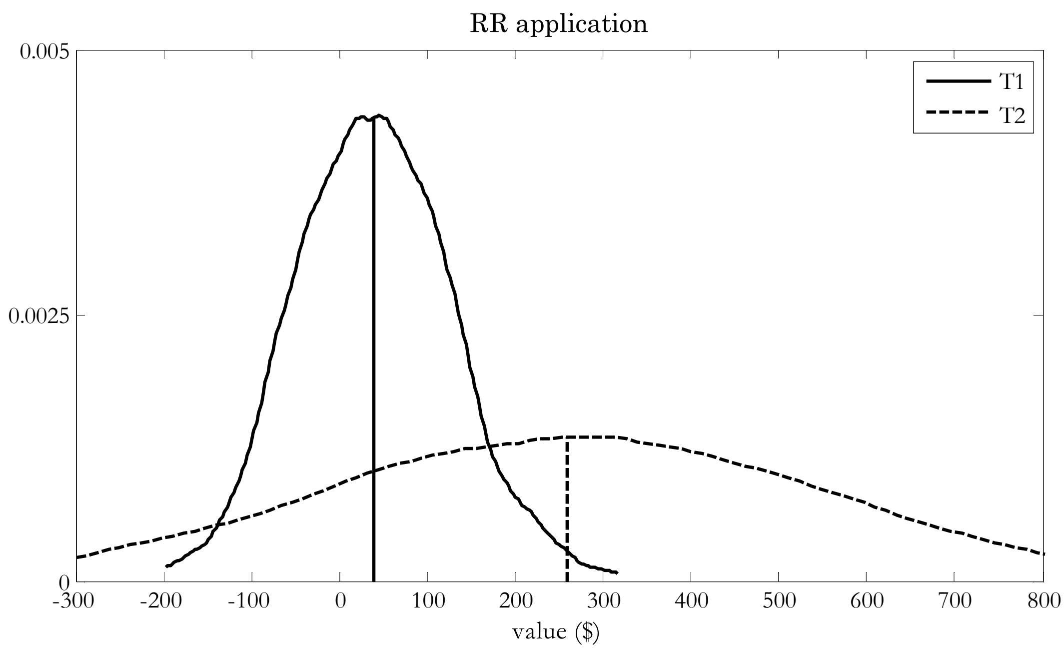


Figure 2: Posterior Predictive Densities for Baseline Types





Thresholds means are indicated with vertical lines.

Regressor settings for GC application:

empathy = 0 (standard survey version highlighting costs to others); *market* = 0 (national sample); mean *know* score (approx. 0), mean *econ* score (approx. 0), mean income (approx. $55,000, 1994 dollars)

Regressor settings for RR application:

mean *qualituy* score (approx. 0), no extra *info*, mean income (approx. $73,500, 2005 dollars)