Johannes Reichl and Sylvia Frühwirth-Schnatter

# A Censored Random Coefficients Model for the Detection of Zero Willingness to Pay

# A Censored Random Coefficients Model for the Detection of Zero Willingness to Pay†

by

Johannes Reichl[1]

Sylvia Frühwirth-Schnatter[2]

October 27, 2011

JEL Codes: C11, Q51

Keywords: Willingness to Pay, Censored Random Coefficients, Bayesian Inference

[1] Johannes Reichl is research associate at the Energy Institute at the Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria.

[2] Sylvia Frühwirth-Schnatter is full professor at the Institute for Statistics and Mathematics, Vienna University of Economics and Business, Augasse 2-6, 1090 Vienna, Austria

† Funding was provided by the Austrian research program KIRAS (grant number: 821746) and the non-profit research association Energy Institute at the Johannes Kepler University Linz.

**Abstract**

In this paper we address the problem of negative estimates of willingness to pay. We find that there exist a number of goods and services, especially in the fields of marketing and environmental valuation, for which only zero or positive WTP is meaningful. For the valuation of these goods an econometric model for the analysis of repeated dichotomous choice data is proposed. Our model restricts the domain of the estimates of WTP to strictly positive values, while also allowing for the detection of zero WTP. The model is tested on a simulated and a real data set.

# 1 Introduction

Willingness to pay (WTP) is an important concept, in particular, in marketing and environmental economics. In marketing, the willingness-to-pay refers to the maximum amount of money an individual is willing to pay in order to acquire a certain product and is the basis for designing a pricing strategy for that particular good. In environmental economics, willingness-to-pay may be described as the maximum amount a person is willing to pay in order to avoid something undesired, such as pollution. Using willingness-to-pay as a proxy for the value of non-market goods as in the contingent valuation method (CVM) has found broad acceptance for valuing non-market goods, since the NOAA Panel (Arrow et al., 1993) has judged this method as suitable for producing "estimates reliable enough to be the starting point of a judicial process of damage assessment, including passive-use values".

In any of these applications, even for a non-market good, an individual's WTP is measured in monetary units. The domain of meaningful values for WTP for a certain good is best discussed within the framework of Random Utility Modeling (RUM). There, the individual's decision between $J$ multi-attributed goods is based on the utility the individual assigns to each of the alternatives. Considering the standard representation of a RUM (see e.g. Sonnier et al., 2007; Train, 2003; Baltas and Doyle, 2001) and denoting the coefficient of the $k^{th}$ attribute $x_k$ of the valued good by $\tilde{\beta}_k$ and the coefficient of the negative price (i.e. of the reduction of the disposable income) by $\alpha$, McFadden (1996) derives WTP for attribute $k$ as

$$WTP_k = \frac{\tilde{\beta}_k}{\alpha}. \tag{1}$$

The sign of the WTP for an improvement of attribute $k$ is therefore a function of the marginal utility of $x_k$ and the marginal utility of income. When $\alpha$ is specified as a common parameter for all respondents, it has shown being robustly positive in numerous applications (see e.g. Layton and Levine, 2003; Layton and Brown, 2000; Adamowicz et al., 1997), while when specified on the unit-level there is a significant risk that at least some of the $\alpha_i$ are estimated with negative sign. Under positive $\alpha$, attributes imposing negative utility on the respondent will technically lead to negative estimates for WTP. There are goods for which a significant share of the respondents might gain utility and for which a significant share of the respondents might experience a loss in utility. For many goods relevant to environmental valuation and marketing, this is not a realistic assumption.

Hence, many authors demand that estimates of WTP need to be somewhere between zero and $\infty$ (see e.g. Bateman et al., 2002; Habb and McConnell, 1997; Carson et al., 1992). Other authors argue that negative values of WTP do exist for most goods and need to be incorporated in the study design as well as in the choice of the econometric model (see e.g. Hanley et al., 2008; Bohara et al., 2001; Clinch and Murphy, 2001). According to these authors, negative values represent resistance to a change in the status quo, and neglecting them biases WTP upwards.

Our paper contributes to this literature by proposing a discrete choice model for the analysis of consumer decisions with the purpose of avoiding negatives estimates of individual willingness to pay, while allowing for a mass build-up at zero. Our model is able to address (what we think) is an important problem that is common in applied work in both marketing and economics. It is reasonable to expect that some of the attributes are of no importance to a number of people, but when they are one can expect a positive effect on these peoples' utility. For such goods the utility $\tilde{\beta}_k$ of one additional unit of the evaluated good $x_k$ is expected to be zero or greater than zero, but not negative. The proposed econometric model is specified for the analysis of repeated dichotomous choice data and produces only positive estimates of WTP through a censored random coefficients specification. This is achieved by defining a strictly non-negative heterogeneity distribution for WTP.

When choices are modeled by specifying the latent utility as dependent variable, as it is done in a RUM, these heterogeneity distributions are specified for $\tilde{\boldsymbol{\beta}}$ and $\alpha$ separately, and hence the heterogeneity distribution of WTP can not be chosen directly. Re-parameterizing the choice model as proposed by Sonnier et al. (2007) allows us to define a heterogeneity distribution for WTP directly, and we overcome the problem of negative estimates of WTP by the choice of a censored distribution resulting in a domain of the estimated WTPs of $[0, \infty)$. The ability to detect zero WTP appears desirable, not only because of the resulting easily interpretable domain for WTP of $[0, \infty)$, but also since it allows us to detect resistance to abandoning the status-quo.

Our method is different from alternative approaches allowing for such a mass build-up at zero and seems to be preferable in the context of estimating WTP. Bayesian variable selection, for instance, allows to set coefficients close to zero to a small value by way of mixture priors, see e.g. George and McCulloch (1997). While this method found numerous applications in statistics (see e.g. Wagner and Duller, 2011) as well as in marketing research (see e.g. Fong and DeSarbo, 2007) it does not solve the problem of (unreasonable) negative WTP and, more importantly, it assumes indifference to a certain attribute for all individuals or for none. While it is not straightforward to introduce heterogeneity in variable selection, see e.g. Chandukala et al. (2011) for a recent application in marketing research and Frühwirth-Schnatter and Wagner (2011) for an application in random-effects modeling, our approach based on the censored random coefficients specification easily captures heterogeneity in the truncated effects.

An alternative approach would be modeling the heterogeneity distribution through a mixture of a strictly positive distribution for some of the $K$ coefficients and a mass at zero for the remaining coefficients. Mixture models are very popular in marketing research (see e.g. Allenby et al., 1998; Frühwirth-Schnatter et al., 2004) and have been introduced in zero WTP estimation for the special case $K = 1$ by Kristroem (1997), where they are referred to as *spike models*[1]. However, in the general case where $K > 1$ there are $2^K$

[1] For a review of spike models in contingent valuation see Hanemann and Kanninen (1999); applications

different subsets of variables where WTP could be 0, limiting the applicability of this approach. While each additional attribute of the good intended for valuation increases the number of model parameters by factor 2 in the mixture model approach, our approach based on the censored random coefficients specification does not suffer from this curse of dimensionality. Nevertheless, we show the preferences of the censored random coefficients approach compared to a mixture model approach exemplarily for our application in Section 4.2.

The remainder of this article is organized as follows: Section 2 introduces the censored random coefficients model specification, and parameter estimation by Markov Chain Monte Carlo (MCMC) sampling is outlined in Section 3. Section 4 demonstrates the properties of the estimation procedure on a simulated data set and on the real data set originally evaluated in Kuriyama et al. (1999). Based on these data four alternative models are estimated and compared to the censored random coefficients model presented in this article. Firstly, beside reporting the estimates given in Kuriyama et al. (1999) we demonstrate the need for restricting the domain of WTP by estimating an uncensored random coefficients model. Secondly, we apply a specification resulting in strictly positive WTPs by defining a lognormal heterogeneity distribution. In a final step we apply a mixture model approach also facilitating a domain for WTP of $[0, \infty)$, but with the strong restriction that WTP is zero or greater than zero for all attributes to overcome the curse of dimensionality discussed in the preceding paragraph. In Section 5 some concluding remarks about the advantages of the model and possible fields for its application are presented.

---

are found in Werner (1999) and Hackl and Pruckner (1999).

## 2 The Censored Random Coefficients Model

In this section an econometric model for the analysis of dichotomous repeated choice data is formulated. We account for the longitudinal structure of the repeated choices by the specification of random coefficients, as was done in numerous papers (see e.g. Greene et al., 2004; Layton and Moeltner, 2004; Rossi et al., 1996). To follow the approach of Sonnier et al. (2007) the utility $u_i$ (as defined in a RUM) is divided through the marginal utility of income $\alpha_i$ to yield respondent $i$'s surplus $s_i = \frac{u_i}{\alpha_i}$ as dependent variable. While this re-parametrization has no implications for the theoretical considerations we have made in Section 1, Sonnier et al. (2007) show that for the likely case that the data contains only few observations per respondent, the specification as a surplus model results in a distribution of WTP with more reasonable tail behaviour in real world applications.

We start with the formulation of the surplus model with unconstrained random coefficients

$$y_{it} = \begin{cases} 1, & \text{if} \quad s_{it} > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

with

$$s_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta}_i + p_{it} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, \tau_i), \tag{3}$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\gamma}'\boldsymbol{z}_i + \boldsymbol{\nu}_i, \qquad \boldsymbol{\nu}_i \sim N(\mathbf{0}, \boldsymbol{\Delta}), \tag{4}$$

$$\text{and} \quad \tau_i \sim IG(r, \varrho). \tag{5}$$

In this specification the single choices $y_{it}$ are combined in a choice vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})'$; similarly, $\boldsymbol{s}_i$ is defined as a vector of length $T_i$ holding the surpluses $s_{i1}, \ldots, s_{iT_i}$ of respondent $i$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT_i})'$. $T_i$ refers to the number of scenario comparisons revealed to respondent $i$, such that $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$. $\boldsymbol{x}_{it}$ is a row vector of length $K$ with the value of attribute $k$ at the $k^{th}$ position. $\boldsymbol{\beta}_i$ holds the coefficients being identified in the surplus specification, such that $\boldsymbol{\beta}_i = \frac{\tilde{\boldsymbol{\beta}}_i}{\alpha_i}$, where the $k^{th}$ element of $\boldsymbol{\beta}_i$ is already the WTP for attribute $k$ in this specification as evident from (1). Referring to the error variance in the utility specification by $\sigma^2$, note that only $\tau_i = \sigma^2 \frac{1}{\alpha_i}$ is identified in the surplus specification; and $r$ and $\varrho$ are the parameters of its distribution. $p_{it}$ is the negative price of scenario $t$, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\nu}$ are iid errors, respectively. Considering $\boldsymbol{\gamma}$ as the coefficients matrix of the random coefficients, and $\boldsymbol{\Delta}$ as their covariance matrix, the distribution of the vector holding the surpluses is given by $\boldsymbol{s}_i \sim N(\boldsymbol{X}_i\boldsymbol{\gamma}'\boldsymbol{z}_i + \boldsymbol{p}_i, \boldsymbol{X}_i\boldsymbol{\Delta}\boldsymbol{X}_i' + \tau_i\boldsymbol{I}_{T_i})$, where $\boldsymbol{I}_{T_i}$ is the identity matrix with $T_i$ rows, $\boldsymbol{X}_i = \left(\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT_i}'\right)'$ and $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iT_i})'$. In the case where $\boldsymbol{z}_i$ is a single 1 for all respondents, one yields the specification of the random coefficients model as originally proposed by Swamy (1970), where $\boldsymbol{\gamma}$ is a $K \times 1$-vector holding the average of $\boldsymbol{\beta}_i$ across units. If $\boldsymbol{z}_i$ holds the $C$ characteristics of respondent $i$, then the model has a hierarchical structure and is able to identify the influence of these characteristics on the unit-level coefficients $\boldsymbol{\beta}_i$ (see e.g. Rossi et al., 2005). In this case $\boldsymbol{\gamma}$ is a $K \times C$ matrix.

The specification in (4) does not prevent negative estimates of WTP, (see e.g. Carlsson and Martinsson, 2008). To overcome the problem of negative estimates of WTP, we exploit the fact that for several goods, especially in the fields of environmental and resource economics, but also in health economics or marketing sciences, to name only a few, it can reasonably be assumed that some or even all attributes of an evaluated good do not affect any individual in a negative manner. The non-existence of a loss in surplus can be ensured by restricting $\boldsymbol{\beta}_i$ to $\Re_0^+$ in the econometric specification of the model. Various distributions for $\boldsymbol{\beta}_i$ satisfy this requirement, such as the lognormal, the Weibull or the truncated normal distribution. However, to allow observing that any of the $WTP_{ik} = 0$ we need to implement a distribution for $\boldsymbol{\beta}_i$ for which $\Pr(\beta_{ik} = 0) > 0$.

As in (4), we assign a $\beta_{ik}$ to each respondent conditional on its characteristics. But whenever the characteristics of the respondent and his or her individual preferences do not indicate a positive gain in surplus for attribute $k$, we assume that the respondent considers attribute $k$ as not relevant to his or her decision, and hence, $\beta_{ik}$ is zero. We bring this aspect into the model from (2) to (5) by introducing a latent random coefficients vector $\boldsymbol{\beta}_i^*$, which is then censored with respect to its sign to obtain the effective coefficients vector $\boldsymbol{\beta}_i$. The formal definition of the above is

$$\begin{aligned} \boldsymbol{\beta}_i^* &= \boldsymbol{\gamma}' \boldsymbol{z}_i + \boldsymbol{\nu}_i, \qquad \boldsymbol{\nu}_i \sim N(\mathbf{0}, \boldsymbol{\Delta}), \\ \boldsymbol{\beta}_i &= \max(\boldsymbol{\beta}_i^*, \mathbf{0}). \end{aligned} \tag{6}$$

This is the specification of a multivariate tobit model (for applications of the multivariate tobit model in environmental valuation see e.g. Moeltner and Layton, 2002; Cornick et al., 1994). The interpretation of the censoring aspect is quite straightforward: If an individual does not align a gain in surplus with an attribute, the decision is based on the remaining attributes only. Censoring the latent individual surpluses of attribute $k$ therefore presupposes that no one in the population experiences a loss in surplus from an increase in attribute $k$. The selection of the attributes for which censoring is applicable is therefore limited by the validity of exclusively non-negative effects of the candidate attributes on the surplus of the evaluated good or service.

# 3 Estimation

## 3.1 Estimating Coefficients by Bayesian Inference

The complexity of the likelihood (see Appendix A) impedes inference by frequentist methods. The hierarchical structure of the model suggests a Bayesian approach, in which the conditional likelihood $p(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \tau_i)$ is combined with the distribution of the random coefficients $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})$, and with the distribution of the scale parameters $p(\tau_i|r, \varrho)$, which are considered as first-stage priors in a Bayesian setting. $r$ is chosen as fixed parameter, such that $\varrho$, $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ remain as unknown hyperparameters and follow the second-stage prior distributions $p(\boldsymbol{\gamma}, \boldsymbol{\Delta}|\boldsymbol{q})$ and $p(\varrho|\boldsymbol{h})$, such that the joint posterior distribution is given by

$$\begin{aligned} & p(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_n, \tau_1, \dots, \tau_n, \boldsymbol{\gamma}, \boldsymbol{\Delta}, \varrho|\boldsymbol{y}_1, \dots, \boldsymbol{y}_n, \boldsymbol{q}, \boldsymbol{h}, r) \propto \\ & \qquad \prod_i p(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \tau_i) p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}) p(\tau_i|r, \varrho) p(\boldsymbol{\gamma}, \boldsymbol{\Delta}|\boldsymbol{q}) p(\varrho|\boldsymbol{h}), \end{aligned} \tag{7}$$

where $\boldsymbol{q}$ and $\boldsymbol{h}$ are the parameters of the corresponding prior distributions.

The prior on the random coefficients $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ comprises deterministic censoring of the underlying latent variables $\boldsymbol{\beta}_i^*$ (following (6)), such that the preferences of the prior distribution are best discussed for the latent $\boldsymbol{\beta}_i^*$. From (6) one can see that the prior parameters $\boldsymbol{\gamma}'\boldsymbol{z}_i$ and $\boldsymbol{\Delta}$ of the unit-level coefficients can have a significant influence on the outcome, as they can not be chosen to be arbitrarily vague. The estimator for the unit-level coefficients in this regard is a shrinkage estimator, as the least-squares estimator of the $\boldsymbol{\beta}_i^*$ is pulled towards the prior mean. For the estimation of individual WTP, we find this property appealing. The uncertainty of some individuals about their WTP for a non-market good or service may be high and result in meaninglessly high least-squares estimates for WTP (Beenstock et al., 1998). The shrinkage tendency of the estimator used in this work may avoid these meaninglessly high estimated WTPs.

We select a Gamma distribution as prior of $\varrho$, giving

$$\varrho \sim G(a_\varrho, A_\varrho), \tag{8}$$

where $a_\varrho$ and $A_\varrho$ are the corresponding prior parameters. The natural conjugate priors for the common parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ are

$$\begin{aligned} \boldsymbol{\Delta} &\sim IW(a_{\Delta}, \boldsymbol{A}_{\Delta}), \\ vec(\boldsymbol{\gamma})|\boldsymbol{\Delta} &\sim N(vec(\boldsymbol{a}_{\gamma}), \boldsymbol{\Delta} \otimes \boldsymbol{A}_{\gamma}^{-1}), \end{aligned} \tag{9}$$

where $a_{\Delta}$ refers to the degrees of freedom and $\boldsymbol{A}_{\Delta}$ to the scale matrix of the inverted Wishart distribution. $vec(\boldsymbol{\gamma})$ is the vectorisation of the coefficients matrix $\boldsymbol{\gamma}$. $\boldsymbol{A}_{\gamma}$ is the covariance matrix of the normal distribution and $vec(\boldsymbol{a}_{\gamma})$, which is the vectorisation of the prior mean matrix $\boldsymbol{a}_{\gamma}$.

### 3.2 Gibbs Sampling

Analytical and numerical calculation of the posterior in (7) is not feasible, but given the latent variable structure of the model applying Gibbs sampling (Gamerman and Lopes, 2006; Geweke, 2005) for posterior inference suggests itself, where for each parameter $\boldsymbol{\theta}$ $M$ draws are generated from $p(\boldsymbol{\theta}^{m+1}|\boldsymbol{\theta}^{m}, \boldsymbol{Y})$.

Sampling from the model introduced in (2), (3), (5) and (6) requires sampling from the following distributions:

1. $p(\boldsymbol{s}_i|\boldsymbol{\beta}_i, \tau_i, \boldsymbol{y}_i) \quad \forall\, i = 1, \ldots, n,$
2. $p(\boldsymbol{\beta}_i^*|\boldsymbol{s}_i, \tau_i, \boldsymbol{\beta}_i, \boldsymbol{\gamma}, \boldsymbol{\Delta}) \quad \forall\, i = 1, \ldots, n,$
3. $p(\boldsymbol{\beta}_i|\boldsymbol{\beta}_i^*) \quad \forall\, i = 1, \ldots, n,$
4. $p(\tau_i|\boldsymbol{s}_i, \boldsymbol{\beta}_i, \varrho) \quad \forall\, i = 1, \ldots, n,$
5. $p(\varrho|\tau_1, \ldots, \tau_n),$
6. $p(\boldsymbol{\gamma}|\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_n^*, \boldsymbol{\Delta}),$
7. $p(\boldsymbol{\Delta}|\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_n^*, \boldsymbol{\gamma}).$

Sampling from distributions 1., 4., 5., 6. and 7. is standard in Bayesian econometric literature, and the reader is therefore referred to Albert and Chib (1993) and Rossi et al. (2005), for instance. Since in 3. $\boldsymbol{\beta}_i$ is a deterministic function of $\boldsymbol{\beta}_i^*$, only sampling from distribution 2. needs further explanation.

To access the distribution of the unit-level coefficients $\boldsymbol{\beta}_i^*$, the equation of the latent surpluses $s_{it}$ in (3) is rewritten as

$$\begin{aligned} s_{it} &= \boldsymbol{x}_{it}\boldsymbol{\beta}_i + p_{it} + \epsilon_{it}, \\ &= \boldsymbol{x}_{it,u}\boldsymbol{\beta}_{i,u} + \boldsymbol{x}_{it,c}\boldsymbol{\beta}_{i,c} + p_{it} + \epsilon_{it}, \\ &= \boldsymbol{x}_{it,u}\boldsymbol{\beta}_{i,u} + p_{it} + \epsilon_{it}, \end{aligned} \tag{10}$$

where $\boldsymbol{\beta}_{i,u}$ refers to the subset of positive values in $\boldsymbol{\beta}_i$ with the corresponding explanatory variables $\boldsymbol{x}_{it,u}$, while $\boldsymbol{\beta}_{i,c}$ refers to the zero elements in $\boldsymbol{\beta}_i$. Regressing $s_{it} - p_{it}$ on $\boldsymbol{x}_{it,u}$ is not feasible without further considerations, since it is not known a priori which of the elements of $\boldsymbol{\beta}_i$ are censored and which elements are uncensored.

We will access this problem by rewriting equation (10) and replacing $\boldsymbol{\beta}_i$ with $\boldsymbol{\beta}_i^*$. Recall that $\boldsymbol{\beta}_{i,u} = \boldsymbol{\beta}_{i,u}^*$, while $\boldsymbol{\beta}_{i,c} \neq \boldsymbol{\beta}_{i,c}^*$. To replace $\boldsymbol{\beta}_i$ with $\boldsymbol{\beta}_i^*$ in equation (10) we introduce the "inactive surplus" $is_{it} = \boldsymbol{x}_{it,c}\boldsymbol{\beta}_{i,c}^*$. This is the surplus that would be generated by those elements of $\boldsymbol{\beta}_i^*$, which were negative in an uncensored model specification. Then

$$s_{it} + is_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta}_i^* + p_{it} + \epsilon_{it}. \tag{11}$$

The inactive surplus $is_{it}$, in contrast to the active surplus $s_{it}$, does not influence the choice of the decision maker in any way and therefore nothing can be said about its sign or its magnitude. For this reason drawing $\boldsymbol{\beta}_i^*$ from $p(\boldsymbol{\beta}_i^*|s_{it}, is_{it})$ and $is_{it}$ from $p(is_{it}|\boldsymbol{\beta}_i^*)$ alternately is not identified.

However, we can still use $is_{it}$ to consistently draw from $p(\boldsymbol{\beta}_i^*|s_{it}, is_{it})$. Conditionally on the information about the signs of the elements of $\boldsymbol{\beta}_i^*$, estimating their magnitudes is straightforward. Therefore, we split the estimation of $\boldsymbol{\beta}_i^*$ into two steps:

**Step I** yields the information which coefficients are censored and obtains a posterior draw of all uncensored coefficients. Conditionally on $\boldsymbol{s}_i + \boldsymbol{is}_i$ we obtain a draw of the signs of $\boldsymbol{\beta}_i^*$ by drawing from $\mathring{\boldsymbol{\beta}}_i^*$, the random coefficients from (11). This distribution is obtained by treating equation (11) as a "usual" random coefficients equation. Therefore, the previous draw of $\boldsymbol{\beta}_{i,u}^*$ is used to augment $\boldsymbol{s}_i$ from $\boldsymbol{s}_i = \boldsymbol{X}_{i,u}\boldsymbol{\beta}_{i,u}^* + \boldsymbol{p}_i + \boldsymbol{\epsilon_i}$, and likewise the previous draw of $\boldsymbol{\beta}_{i,c}^*$ is used to define $\boldsymbol{is}_i = \boldsymbol{X}_{i,c}\boldsymbol{\beta}_{i,c}^*$.

Conditionally on the $\boldsymbol{s}_i$ and $\boldsymbol{is}_i$ obtained in this way the distribution of $p(\mathring{\boldsymbol{\beta}}_i^*|\boldsymbol{s}_i, \boldsymbol{is}_i, \boldsymbol{\Delta}, \boldsymbol{\gamma})$ can be written as

$$\begin{aligned}
&\mathring{\boldsymbol{\beta}}_i^*|\boldsymbol{s}_i, \boldsymbol{is}_i, \boldsymbol{\Delta}, \boldsymbol{\gamma} \sim N(\tilde{\boldsymbol{b}}_i, \tilde{\boldsymbol{B}}_i)\\
&\tilde{\boldsymbol{B}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i + \boldsymbol{\Delta}^{-1})^{-1}\\
&\tilde{\boldsymbol{b}}_i = \tilde{\boldsymbol{B}}_i(\boldsymbol{X}_i'(\boldsymbol{s}_i + \boldsymbol{is}_i - \boldsymbol{p}_i) + \boldsymbol{\Delta}^{-1}\boldsymbol{\gamma}'\boldsymbol{z}_i).
\end{aligned} \tag{12}$$

With this procedure a draw of the vector $\mathring{\boldsymbol{\beta}}_i^*$ not only holds the information about the signs of $\boldsymbol{\beta}_i^*$, but its positive elements are already a consistent draw of $\boldsymbol{\beta}_{i,u}^* = \boldsymbol{\beta}_{i,u}$. On the other hand, the magnitude of the negative elements of $\mathring{\boldsymbol{\beta}}_i^*$ can not be used to achieve further identification of the sampler. Since the specification of the $\boldsymbol{\beta}_i$ is the specification of a multivariate tobit model, we apply exactly the same technique as in the tobit case in Step II.

Conditionally on the $\boldsymbol{\beta}_{i,u}^*$ from Step I, **Step II** therefore obtains a draw of $\boldsymbol{\beta}_{i,c}^*$ by data augmentation, as in Chib (1992). In this case the magnitude of $\boldsymbol{\beta}_{i,c}^*$ is estimated independently of $\boldsymbol{is}_i$, since in this data augmentation step we use only information on the prior parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$ and on the positive coefficients $\boldsymbol{\beta}_{i,u}^*$. We obtain the conditional distribution $p(\boldsymbol{\beta}_{i,c}^*|\boldsymbol{\beta}_{i,u}^*, \boldsymbol{\gamma}, \boldsymbol{\Delta})$ by conditioning the common distribution $p(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ – as defined in (4) – on the uncensored coefficients and truncating the resulting distribution above zero. This distribution is equal to a multivariate truncated normal distribution

$$\begin{aligned}
&\boldsymbol{\beta}^*_{i,c}|\boldsymbol{\beta}^*_{i,u},\boldsymbol{\gamma},\boldsymbol{\Delta} \sim TN_{[-\infty,0]}(\tilde{\boldsymbol{\beta}}^*_{i,c},\tilde{\boldsymbol{V}}^*_{i,c}),\\
&\tilde{\boldsymbol{V}}^*_{i,c} = \boldsymbol{\Gamma}^{-1}_{cc}\\
&\tilde{\boldsymbol{\beta}}^*_{i,c} = \boldsymbol{\gamma}'_c\boldsymbol{z}_{i,c} - \boldsymbol{\Gamma}^{-1}_{cc}\boldsymbol{\Gamma}_{cu}(\boldsymbol{\beta}^*_{i,u} - \boldsymbol{\gamma}'_u\boldsymbol{z}_{i,u})
\end{aligned} \tag{13}$$

where $\boldsymbol{\gamma}_c$ is the matrix constructed from the columns belonging to $\boldsymbol{\beta}_{i,c}$ and $\boldsymbol{z}_{i,c}$ is the corresponding vector of the characteristics of respondent $i$. Likewise $\boldsymbol{\gamma}_u$ is the matrix constructed from the columns belonging to $\boldsymbol{\beta}_{i,u}$ and $\boldsymbol{z}_{i,u}$ is the corresponding vector of the characteristics of respondent $i$. The matrices $\boldsymbol{\Gamma}_{cc}$ and $\boldsymbol{\Gamma}_{cu}$ are submatrices of the following matrix $\boldsymbol{\Gamma}_{u,c}$

$$\boldsymbol{\Gamma}_{u,c}= \begin{pmatrix} \boldsymbol{\Gamma}_{uu} & \boldsymbol{\Gamma}_{uc} \\ \boldsymbol{\Gamma}_{cu} & \boldsymbol{\Gamma}_{cc} \end{pmatrix},$$

where $\boldsymbol{\Gamma}^{-1}_{u,c}$ is the covariance matrix $\boldsymbol{\Delta}$ from (4) which has been reorganized in such a way that $\boldsymbol{\Delta}_{uu}$ is the covariance matrix of the $\boldsymbol{\beta}^*_{i,u}$ and $\boldsymbol{\Delta}_{cc}$ is the covariance matrix of the $\boldsymbol{\beta}^*_{i,c}$: $\boldsymbol{\Gamma}_{u,c} = \begin{pmatrix} \boldsymbol{\Delta}_{uu} & \boldsymbol{\Delta}_{uc} \\ \boldsymbol{\Delta}_{cu} & \boldsymbol{\Delta}_{cc} \end{pmatrix}^{-1}$.

## 4 Applications of the Model

In this section, the model is applied to a simulated data set in Section 4.1 and to the Tokyo Bay data set (Kuriyama et al., 1999) in Section 4.2. The simulation study examines the properties of the censored coefficients model specification and compares the estimated parameters to those retrieved under the uncensored, but otherwise identical, model specification. On the base of the re-estimation of the Tokyo Bay data set, we discuss the value of the additional information on the respondents preferences obtained through the censored coefficients model specification by comparing it to different selected

**Table 1:** Prior settings for the analysis of the examples in this chapter.

| Prior | Simulation study | Tokyo Bay data set |
|---|---|---|
| $r$ | . | 1/2 |
| $a_\varrho$ | . | 1 |
| $A_\varrho$ | . | 0.01 |
| $\boldsymbol{a}_\gamma$ | $\mathbf{0}_{C\times K}$ | $\mathbf{0}_{C\times K}$ |
| $\boldsymbol{A}_\gamma$ | $0.01 \times \boldsymbol{I}_K$ | $0.01 \times \boldsymbol{I}_K$ |
| $a_\Delta$ | $K+3$ | $K+\max(T_i)/2$ |
| $\boldsymbol{A}_\Delta$ | $a_\Delta \cdot 0.1\boldsymbol{I}_K$ | $a_\Delta \cdot \boldsymbol{I}_K$ |

models.

For the simulation study the parameters of the prior distributions are chosen to be very diffuse, following the setting in Rossi et al. (2005) for their (uncensored) random coefficients model. In the case of real data, however, we experienced convergence problems under diffuse priors. The problem arises from the fact that information on the shape of the likelihood is contained in the positive elements of $\boldsymbol{\beta}_i^*$ only, and with an increasing number of negative elements of $\boldsymbol{\beta}_i^*$, this information decreases. If the number of negative elements of $\boldsymbol{\beta}_i^*$ is large and the information in the data on the positive $\boldsymbol{\beta}_i$ is uncertain, it is likely that during tens of thousands of draws of $\boldsymbol{\beta}_i^*$ the fraction of zeros of a specific draw will approach 1. In the case of diffuse priors this can lead to extremely large draws for $\boldsymbol{\gamma}$ and $\boldsymbol{\Delta}$, and the sampler will lose itself in unreasonable areas of the likelihood. To avoid this problem informative priors are chosen that prevent drawing unreasonably large values. Prior settings for the simulation study and the Tokyo Bay data set are given in Table 1.

## 4.1 A Simulation Study

In this simulation study the emphasis is on examining the effects the censoring of the random coefficients has on the precision of the coefficients estimated. To achieve this we compare the censored coefficients specification of our model with the uncensored random coefficients model. Since we want to relate the difference in precision of the two

**Table 2:** Results from the simulation study; true values of the parameter matrix $\boldsymbol{\gamma}$ in the first three columns and corresponding posterior means of the censored and the uncensored model in the subsequent three columns, respectively.

| $\gamma_{.1}$ | $\gamma_{.2}$ | $\gamma_{.3}$ | $\hat{\gamma}^c_{.1}$ | $\hat{\gamma}^c_{.2}$ | $\hat{\gamma}^c_{.3}$ | $\hat{\gamma}^u_{.1}$ | $\hat{\gamma}^u_{.2}$ | $\hat{\gamma}^u_{.3}$ |
|---|---|---|---|---|---|---|---|---|
| 1.00 | -0.25 | 0.50 | 0.96 | -0.24 | 0.62 | 0.94 | -0.37 | 0.48 |
| -0.25 | 0.50 | -0.50 | -0.17 | 0.61 | -0.50 | -0.20 | 0.64 | -0.50 |
| -0.25 | 0.25 | -0.50 | -0.19 | 0.26 | -0.44 | -0.20 | 0.28 | -0.49 |

**Table 3:** Results from the simulation study; true values of the covariance matrix $\boldsymbol{\Delta}$ in the first three columns and corresponding posterior means of the censored and the uncensored model in the subsequent three columns, respectively.

| $\boldsymbol{\Delta}_{.1}$ | $\boldsymbol{\Delta}_{.2}$ | $\boldsymbol{\Delta}_{.3}$ | $\hat{\boldsymbol{\Delta}}^c_{.1}$ | $\hat{\boldsymbol{\Delta}}^c_{.2}$ | $\hat{\boldsymbol{\Delta}}^c_{.3}$ | $\hat{\boldsymbol{\Delta}}^u_{.1}$ | $\hat{\boldsymbol{\Delta}}^u_{.2}$ | $\hat{\boldsymbol{\Delta}}^u_{.3}$ |
|---|---|---|---|---|---|---|---|---|
| 1.0 | 0.5 | 0.8 | 0.84 | 0.47 | 0.64 | 0.99 | 0.46 | 0.74 |
| 0.5 | 1.5 | 0.4 | 0.47 | 1.27 | 0.59 | 0.46 | 1.44 | 0.47 |
| 0.8 | 0.4 | 0.8 | 0.64 | 0.59 | 0.66 | 0.74 | 0.47 | 0.89 |

specifications to the censoring of the coefficients only, we consider the latent surpluses as observable in this simulation study. In addition, the price term $p_{it}$ is removed and the scale is set to $\tau_i = 1,\ \forall\, i = 1, \ldots, n$. Thus, we specify the uncensored random coefficients model as

$$\begin{aligned} s^u_{it} &= \boldsymbol{x}_{it}\boldsymbol{\beta}^u_i + \epsilon_{it}, \qquad \epsilon_{it} \sim N(0,1), \\ \boldsymbol{\beta}^u_i &= \boldsymbol{\gamma}'\boldsymbol{z}_i + \boldsymbol{\nu}_i, \qquad \boldsymbol{\nu}_i \sim N(\mathbf{0}, \boldsymbol{\Delta}), \end{aligned} \tag{14}$$

where the superscript $.^u$ denotes the uncensored specification. Likewise, we use the superscript $.^c$ to refer to the censored specification in subsequent paragraphs.

The parameter matrix $\boldsymbol{\gamma}$ and the covariance matrix $\boldsymbol{\Delta}$ of the latent random coefficients $\boldsymbol{\beta}^*_i$ used in this simulation are given in Table 2 and Table 3.

The subject characteristics $\boldsymbol{z}_i$ are chosen to have a 1 in the first position, are uniformly distributed $Unif(0,2)$ in the second position and follow a normal distribution $N(0,2)$ in the third position, hence $\boldsymbol{z}_i$ has dimension $3 \times 1$ in the simulation. The size of the

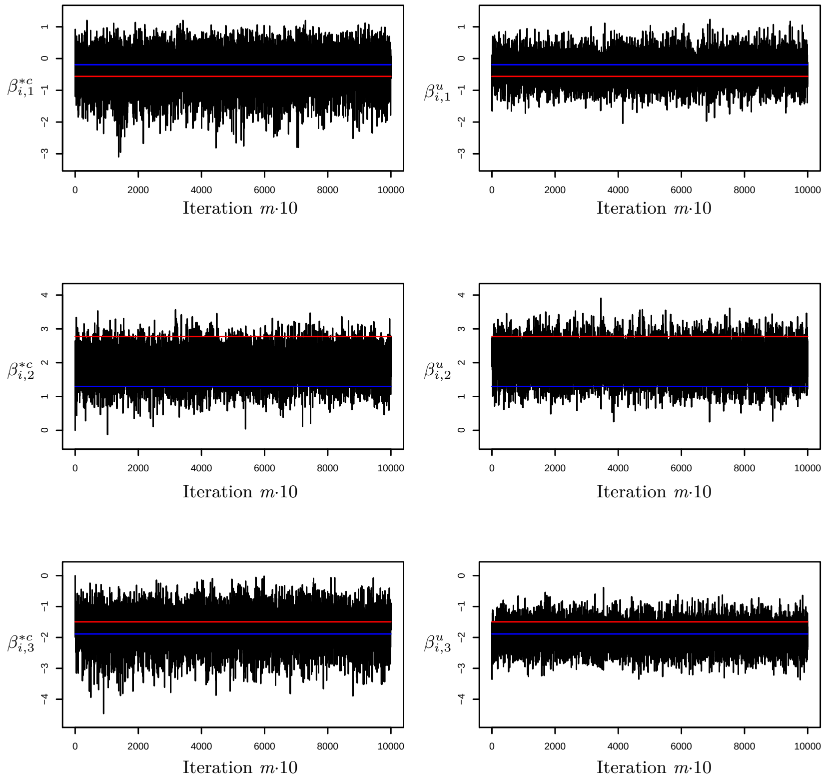


**Figure 1:** Results from the simulation study: MCMC sampling pathes of one latent censored random coefficients vector $\boldsymbol{\beta}_i^{*c}$ (left-hand side) and of the corresponding effective uncensored coefficients vector $\boldsymbol{\beta}_i^{u}$ (right-hand side) with prior means (blue lines) and actual coefficients (red lines).

sample is set to $n = 350$ and $T = 15$ for all units. From these values the latent random coefficients vectors $\boldsymbol{\beta}_i^*$ are drawn and censored according to $\boldsymbol{\beta}_i^c = \max(0, \boldsymbol{\beta}_i^*)$, while the uncensored coefficients equal these latent coefficients, such that $\boldsymbol{\beta}_i^u = \boldsymbol{\beta}_i^*$. This setting yields a fraction of censored coefficients of $\pi(\beta_{i,1}^c = 0) = 26\%$, $\pi(\beta_{i,2}^c = 0) = 42\%$ and $\pi(\beta_{i,3}^c = 0) = 53\%$.

In Figure 1 the sampling paths of an arbitrarily chosen $\boldsymbol{\beta}_i^*$ under the censored and under the uncensored model specification are displayed. From these plots several propositions can be derived. Firstly, we see that the loss in precision for an element of $\boldsymbol{\beta}_i^{*c}$ with most probability mass on the positive space (here $\beta_{i,2}^{*c}$) in the censored model specification is very low, compared to the uncensored model specification. Secondly, we see that the

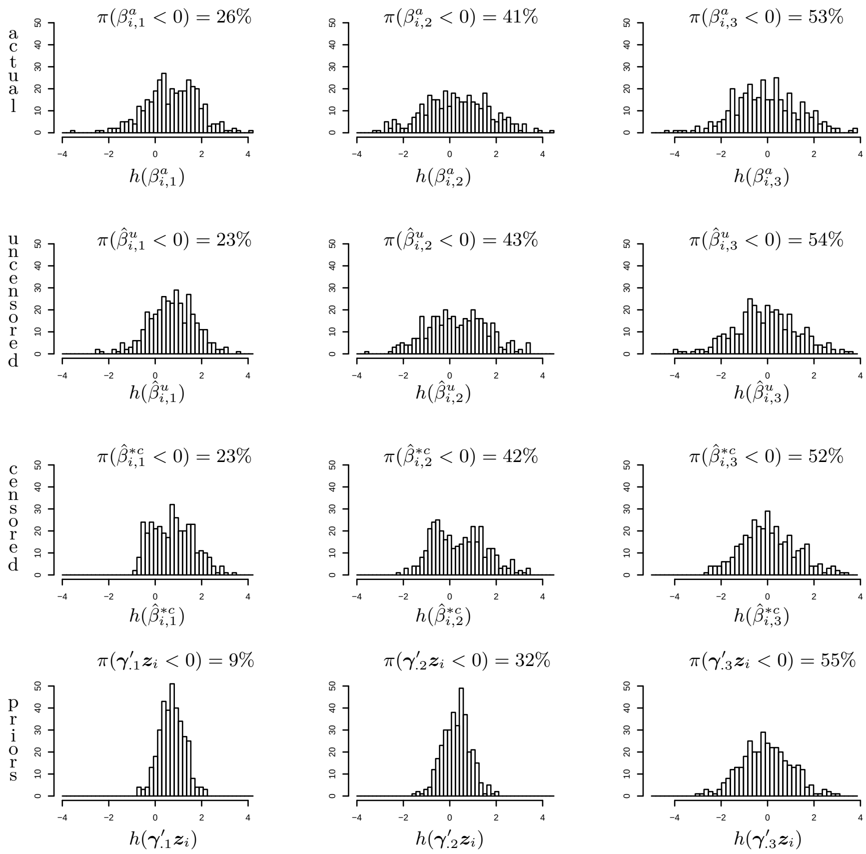


**Figure 2:** Results from the simulation study: histograms of the actual simulated random coefficients $\boldsymbol{\beta}^a_i$, the estimated coefficients from the uncensored specification $\hat{\boldsymbol{\beta}}^u_i$, the estimated latent coefficients from the censored specification $\hat{\boldsymbol{\beta}}^{*c}_i$ and the actual first-stage priors $\hat{\boldsymbol{\gamma}}'\boldsymbol{z}_i$.

variance of the estimator of the elements of $\boldsymbol{\beta}^*_i$ with substantial probability mass on the negative space is higher in the censored than in the uncensored model specification. This finding is plausible, since the utilities do not contain any information about the negative coefficients in the censored model specification, so that these coefficients need to be estimated from the positive elements of $\boldsymbol{\beta}^*_i$ in combination with the first-stage prior distribution only. As a third result we see that the medians of the posterior distribution of the coefficients do not differ much between the specifications. It should be noted that this last finding is valid for coefficient vectors $\boldsymbol{\beta}^{*c}_i$ with at least one positive element only. Otherwise the posterior distribution of $\boldsymbol{\beta}^{*c}_i$ is identical to the first-stage prior distribution and the coefficients $\boldsymbol{\beta}^{*c}_i$ are estimated independently of the set of surpluses of unit $i$.

In Figure 2 the histograms of the estimated $\hat{\boldsymbol{\beta}}_i^{*c}$ and $\hat{\boldsymbol{\beta}}_i^{u}$ are shown, and compared with the actual coefficients $\boldsymbol{\beta}_i^a$ and the actual first-stage prior locations $\boldsymbol{\gamma}'\boldsymbol{z}_i$. We see that the shrinkage of the latent random coefficients in the censored specification is asymmetric. Naturally, the influence of the first-stage prior distribution is much stronger on the negative space than on the positive space. This is plausible, since the positive coefficients are "regressed" on the utilities, while the negative coefficients are drawn from the prior distribution (conditional on the positive elements of $\boldsymbol{\beta}_i$). In addition, a detailed examination of the random coefficients reveals that – due to their high positive correlation – about 11% of the actual random coefficients are negative not only for one or two elements of $\boldsymbol{\beta}_i$, but for all three elements. In these cases the only information available for these coefficients is the respective prior distribution, from which they are then drawn. Thus, a stronger dependence on the prior distribution for the negative space of the coefficients comes naturally with the censored coefficients specification.

Despite the strong influence of the prior on shape and skewness of the distribution of the $\hat{\boldsymbol{\beta}}_i^{*c}$, Figure 2 reveals that the estimated ratio of $\pi(\hat{\beta}_{i,\cdot}^{*c} < 0)$ of the latent censored coefficients is similar to the estimated ratio under the uncensored coefficients specification $\pi(\hat{\beta}_{i,\cdot}^{u} < 0)$, and both ratios reproduce the actual $\pi(\beta_{i,\cdot}^{a} < 0)$ relatively well.

In Tables 2 and 3 moments of the posterior distributions of both models are reported. The parameter matrix $\boldsymbol{\gamma}$ is estimated with only moderate deviations from its actual values under both model specifications and the deviations give no evidence of any systematic error. For the covariance matrix $\boldsymbol{\Delta}$ the estimators of both model specifications tend to underestimate the actual values, while the censored model shows this preference more markedly. Again, the increased shrinkage of the censored coefficients specification is responsible for this effect.

## 4.2 The Tokyo Bay Data Set

In typical problems addressed by CVM the analyzed goods are often policy programs for improvements of more than one environmental variable. An example of the valuation of such a multi-attributed policy program is given in Kuriyama et al. (1999), where a program to protect Tokyo Bay from oil spills is evaluated. With an area of 960 $km^2$ Tokyo Bay is traversed by 600-900 ships per day. Beside several smaller oil spills, in 1997 the supertanker Diamond Grace leaked about 1,500 tons of crude oil off Tokyo Bay. Only 6 months earlier Japan had been hit by another major oil spill: the Russian tanker Nakhodka leaked about 5,200 tons of crude oil and threatened prized shellfish beds. In the light of this a survey of WTP for preserving the coastal ecosystem by minimizing the extent of four potential kinds of damage was investigated: (1) the percentage of the recreational sites polluted, (2) the number of people harmed by the oil, (3) the percentage of the tidal flat area contaminated, and (4) the percentage of the commercial fishing ports affected. As the specific configuration of a policy program can place emphasis on the improvement of one or other of these variables, WTP for the mitigation of each of these types of loss represents important information for decision makers to find the optimal program configuration.

In their survey each of 128 randomly selected residents was confronted with eight different scenarios of environmental improvement compared to the precisely defined status quo; see Table 4 for an exemplary choice set used in this survey.

The data set thus consists of 1,024 dichotomous choices with 8 choices from each respondent. For each respondent information is available on the values of the four environmental attributes and on the costs he or she was confronted with. No information regarding the characteristics of the respondents was collected, so the vectors $\boldsymbol{z}_i$ become a scalar with value 1 for all respondents.

To study the properties of the censored random coefficients model in an application to

**Table 4:** Tokyo Bay data set; choice sets used in Kuriyama et al. (1999).

| | Alternative A | Alternative B (status quo) |
|---|---|---|
| **recreation site:** | 7% protected | 7% protected |
| **harmed people:** | 10,000 affected | 10,000 affected |
| **tidal flat:** | 87% protected | 13% protected |
| **fishery:** | 100% protected | 66% protected |
| **yearly costs:** | 90,000 ¥ | 0 ¥ |

real data, we compare it to four alternative model specifications. As a starting point we report the results in Kuriyama et al. (1999), where choices are modeled as a RUM with fixed coefficients $\boldsymbol{\beta}$ and $\alpha$. The utility of choosing alternative A instead of alternative B, is therefore given as

$$u_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + p_{it}\alpha + \epsilon_{it}. \tag{15}$$

As a result of this constant coefficients vector only average WTP instead of the unit-level WTPs can be calculated. Equation (15) is estimated by means of the usual logit model, thus we denote their model by $\mathcal{M}_{logit}$.

Beside $\mathcal{M}_{logit}$, all other specifications used for the model comparisons differ from the censored random coefficients model in (2), (3), (5) and (6) only in the distribution of the $\boldsymbol{\beta}_i$. Estimations are carried out by the Gibbs sampler from Section 3.2, where the step of drawing the $\boldsymbol{\beta}_i$ is adapted to the respective model specification. For all models estimated by Gibbs sampling, prior parameters are taken from Table 1.

The first unit-level model specifies $\boldsymbol{\beta}_i$ to be normally distributed as in (4), which does not constrain the domain of the WTPs in any way. In this regard, this specification reveals whether one or more of the valued attributes are estimated negatively for a significant number of respondents when no constraints are placed over the coefficients, and thus whether restricting the coefficients on the positive space is worth the effort. We denote this model specification by $\mathcal{M}_{uncensN}$.

The second unit-level specification defines a lognormal distribution for the random coefficients $\boldsymbol{\beta}_i$, such that their distribution is given by $\boldsymbol{\beta}_i \sim logN(\boldsymbol{\gamma}'\boldsymbol{z}_i, \boldsymbol{\Delta})$. The lognormal distribution ensures exclusively positive estimates of WTP, but resistance against a change in status quo, such as zero WTP, is not identified under this distributional assumption, which we denote by $\mathcal{M}_{logN}$. Estimation by Gibbs sampling requires a Metropolis-Hastings step to draw from the posterior distribution of $\boldsymbol{\beta}_i$, which is given by

$$p(\boldsymbol{\beta}_i|\tau_i, \boldsymbol{\gamma}, \boldsymbol{\Delta}) \propto N(\boldsymbol{s}_i|\boldsymbol{x}_i\boldsymbol{\beta}_i + \boldsymbol{p}_i, \tau_i I_K)\ logN(\boldsymbol{\beta}_i|\boldsymbol{\gamma}'\boldsymbol{z}_i, \boldsymbol{\Delta}). \tag{16}$$

Acceptance rates in the MH step lie between 0.15 and 0.6 $\forall\ i = 1, \dots, n$ under the proposal density $\boldsymbol{\beta}_i^{cand} \sim logN(\log(\boldsymbol{\beta}_i^m), 0.1\ (\boldsymbol{\Delta}^m)^{-1})$, where $\boldsymbol{\beta}_i^m$ and $\boldsymbol{\Delta}^m$ are the recent draws of these parameters.

For the last comparison we specify a mixture distribution for $\boldsymbol{\beta}_i$. The censored random coefficients model allows that each respondent has a WTP of zero for an arbitrary number of the $K$ valued attributes, while having a positive WTP for the remaining attributes. It is impractical to apply a heterogeneity distribution accounting for all possible combinations of signs of the elements of $\boldsymbol{\beta}_i$ by specifying $2^K = 16$ groups. Therefore we estimate only 2 groups, one with all elements of $\boldsymbol{\beta}_i$ being greater than zero and one with all elements being zero. This implies that the ratio of zero WTP is equal for all four attributes, which surely represents a highly restrictive assumption. Thereof we see the limitations of the mixture model approach when the number of valued attributes $K$ increases.

For the group with only positive coefficients $\boldsymbol{\beta}_i$, we again specify a lognormal distribution, such that the resulting mixture distribution is given by

$$\boldsymbol{\beta}_i \sim \eta_1 p(\boldsymbol{\beta}_i|\tau_i, \boldsymbol{\gamma}, \boldsymbol{\Delta}) + \eta_2 \delta_0(\beta_{i1}) \dots \delta_0(\beta_{iK}), \tag{17}$$

where $\delta_0(.)$ is Dirac‘s point mass, $p(\boldsymbol{\beta}_i|\tau_i, \boldsymbol{\gamma}, \boldsymbol{\Delta})$ is the density defined in (16) and drawing from it is carried out likewise. As prior for the relative group sizes we specify a uniform distribution and denote the resulting mixture model by$\mathcal{M}_{logNMix}$.

In contrast to this, we consider the censored random coefficients as specified in Section 2 and refer to it by $\mathcal{M}_{censN}$. Under the prior settings from Table 1 convergence of the Gibbs sampler in all specifications is reached after only a relatively small number of draws. Nevertheless, 50,000 draws are sampled, respectively, of which the last 40,000 enter the calculation of the posterior moments.

Because the scales differ between the utility specification in $\mathcal{M}_{logit}$ and the unit-level models, the magnitude of their coefficients can not be compared directly. Nevertheless, since our primary aim in making the comparison is estimating WTP under the different model specifications, we compare the respective ratios of two coefficients, and the different scales are canceled out. In Figure 3 the histograms of the WTPs estimated under $\mathcal{M}_{censN}$ for the different attributes of the environmental preventive measures are displayed. In Table 5 some key figures of the models are shown.

As a first finding we see that under the unconstrained model $\mathcal{M}_{uncensN}$ negative WTP is estimated for a significant number of respondents, therefore we conclude that applying a model restricting the domain of WTP is worth the effort. Only the censored random coefficients model $\mathcal{M}_{censN}$ and the heterogeneity model $\mathcal{M}_{logNMix}$ indicate how many of the respondents have zero WTP. While the percentage of respondents with zero WTP is equally 29% for all attributes in $\mathcal{M}_{logNMix}$, these percentages differ significantly in $\mathcal{M}_{censN}$. There, the largest number of zero WTPs is estimated for the attribute REC (protection of recreational facilities). For this environmental good 57 respondents, i.e. about 45%, are not willing to pay for the protection of these facilities. 45 persons or about 35% do not have a positive WTP for the protection of tidelands and 41 persons or about 32% are not willing to pay for protection from health problems. The smallest resistance to paying for the protection of the coastal ecosystem is found for the protection of fishing ports; only 31 persons or about 24% are not willing to pay for their protection.

**Table 5:** Tokyo Bay data set; selected key figures from the five estimated model specifications for REC (percentage of the polluted recreational sites), TIDE (percentage of the contaminated tidal flat area), HAR (number of people harmed by the oil), and FISH (percentage of the polluted recreational fishing sites).

| Attribute | | $\mathcal{M}^1_{censN}$ | $\mathcal{M}^1_{logit}$ | $\mathcal{M}^1_{uncensN}$ | $\mathcal{M}^1_{logN}$ | $\mathcal{M}^1_{logNMix}$ |
|---|---|---|---|---|---|---|
| REC | $\overline{WTP}$ | 55 | 88 | 55 | 64 | 31 |
| | | $(15, 117)$ | $(-32, 209)$ | $(-31, 107)$ | $(31, 120)$ | $(2, 85)$ |
| | $\pi(WTP = 0)$ | 47% | . | 0% | 0% | 29% |
| | $\pi(WTP < 0)$ | 0% | . | 23% | 0% | 0% |
| HAR | $\overline{WTP}$ | 205 | 177 | 175 | 243 | 184 |
| | | $(132, 253)$ | $(78, 271)$ | $(78, 208)$ | $(137, 358)$ | $(100, 250)$ |
| | $\pi(WTP = 0)$ | 35% | . | 0% | 0% | 29% |
| | $\pi(WTP < 0)$ | 0% | . | 23% | 0% | 0% |
| TIDE | $\overline{WTP}$ | 302 | 157 | 258 | 171 | 220 |
| | | $(186, 381)$ | $(-9, 309)$ | $(93, 310)$ | $(70, 286)$ | $(118, 330)$ |
| | $\pi(WTP = 0)$ | 32% | . | 0% | 0% | 29% |
| | $\pi(WTP < 0)$ | 0% | . | 29% | 0% | 0% |
| FISH | $\overline{WTP}$ | 452 | 674 | 472 | 683 | 530 |
| | | $(248, 617)$ | $(429, 943)$ | $(262, 614)$ | $(314, 877)$ | $(265, 704)$ |
| | $\pi(WTP = 0)$ | 24% | . | 0% | 0% | 29% |
| | $\pi(WTP < 0)$ | 0% | . | 0% | 0% | 0% |
| In-sample fit[2] | | -605.07 | -641.27 | -608.37 | -629.57 | -583.99 |
| Out-of-sample fit 1a[4] | | 0.417 | 0.418 | 0.423 | 0.410 | 0.375 |
| Out-of-sample fit 2a[5] | | 0.728 | 0.696 | 0.718 | 0.705 | 0.723 |
| Out-of-sample fit 1b[4] | | 0.252 | 0.409 | 0.240 | 0.236 | 0.255 |
| Out-of-sample fit 2b[5] | | 0.828 | 0.679 | 0.820 | 0.828 | 0.849 |

[1] 95% credibility interval in parentheses estimated from the MCMC output.
[2] In sample model fit is measured by the log marginal density calculated by importance sampling with an importance density as described in Congdon (2003).
[4] Out-of-sample fit 1. is measured by mean absolute deviation of estimated choice probability and actual choice.
[5] Out-of-sample fit 2. is measured by the fraction of correctly predicted actual choices.

The plausibility of the censored model specification is compared to the other model specifications by reporting the log marginal density (LMD) of all models (Table 5). Evidence on the predictive capabilities of the models is investigated by holdout validation. For applications in environmental economics the key issue is how good the estimated WTP in the sample can be transferred to the rest of the population. Therefore we reestimate all models applying all observations from the first 100 respondents, and their ability to predict the choices of the remaining 28 respondents is explored. Table 5 reports the mean absolute deviation of estimated choice probability and actual choice in

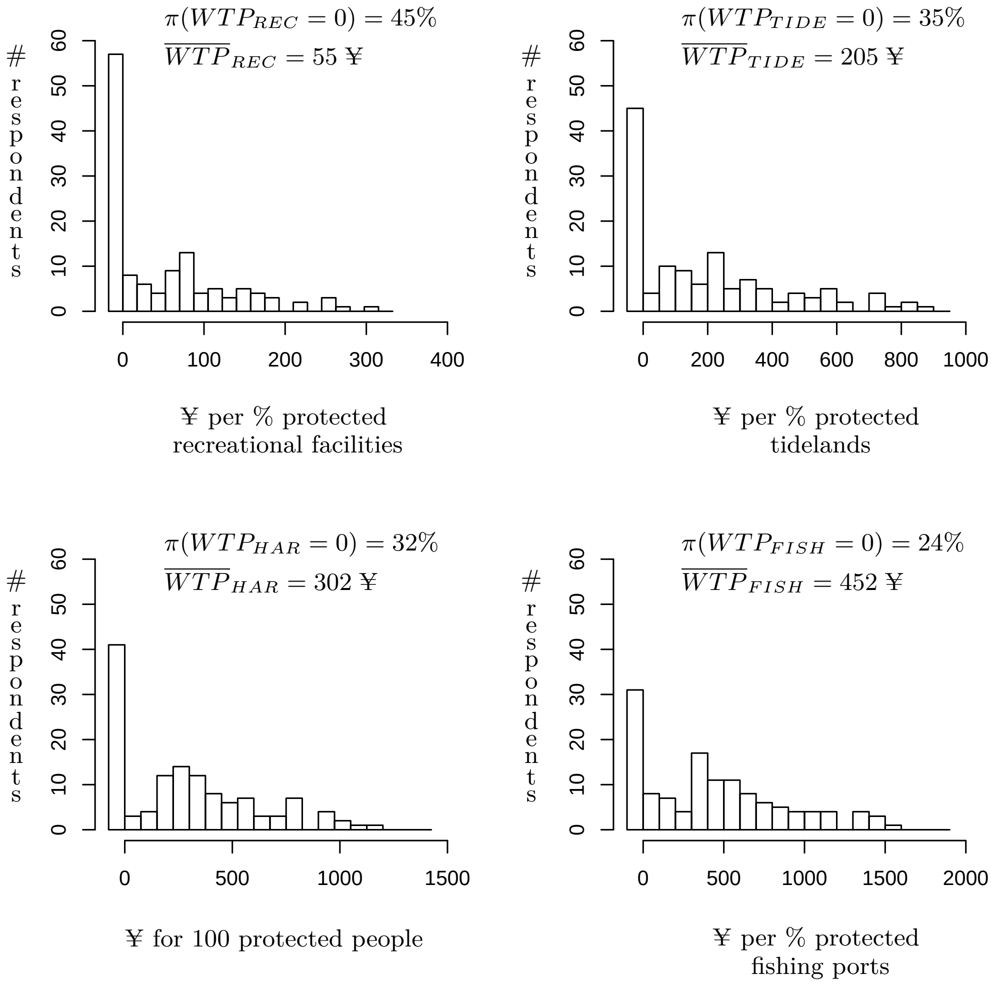


**Figure 3:** Results from the Tokyo Bay data set: histograms of estimated WTP under $\mathcal{M}_{censN}$ for REC (percentage of the polluted recreational sites), TIDE (percentage of the contaminated tidal flat area), HAR (number of people harmed by the oil), and FISH (percentage of the polluted recreational fishing sites).

figure *Out-of-sample fit 1a* (OSF 1a), and the ratio of the correctly predicted choices in figure *Out-of-sample fit 2a* (OSF 2a). Additionally, we report the same measures *Out-of-sample fit 1b* (OSF 1b) and *Out-of-sample fit 2b* (OSF 2b) when a randomly chosen observation of each respondent is predicted based on the re-estimated models analyzing the remaining observations of the respondents.

Considering these fit statistics in Table 5, we see that the assumption of exclusively positive WTPs of all respondents in $\mathcal{M}_{logN}$ is hardly supported by the data. Furthermore, we see that both models allowing for a mass build-up at zero, namely $\mathcal{M}_{censN}$ and $\mathcal{M}_{logNMix}$, outperform the other models. Even though $\mathcal{M}_{logNMix}$ shows remarkable performance in most parameters, its highly restrictive assumptions constrain its general applicability, especially in the politically sensitive field of environmental valuation in our opinion. In contrast, the censored coefficients model $\mathcal{M}_{censN}$ shows a good predictive performance, both when predictions can be based on respondent specific information (OSF .b) and when predictions are based on the prior information (OSF .a) only. Furthermore, as discussed in the introductory Section 1, the censored model is the only one producing estimates of WTP exploiting its theoretical domain $[0, \infty)$ separately for each variable. Thus we conclude that the censored coefficients approach may be of great avail when resistance against abandoning the status-quo is an important information to the decision makers, and the level of this resistance can not be assumed as being equal for all variables.

# 5 Discussion

In this article an econometric model for estimating consumer willingness to pay from repeated dichotomous choice data has been proposed. Based on random utility modeling a specification was developed in which the random coefficients (usually unrestricted) are censored below zero. This restriction on the random coefficients yields a domain for the estimated WTP of $[0, \infty)$, while unrestricted specifications usually result in negative estimates for WTP for at least some respondents and/or goods.

In the authors' opinion this specification has the advantages that first, the resulting domain of WTP is easily interpretable and therefore possibly welcome to those who need to share their results outside the scientific community, and second, it allows us to detect respondents with resistance to changing the status quo, as required by some authors (see e.g. Hanley et al., 2008; Bohara et al., 2001; Clinch and Murphy, 2001).

The application of the model to a simulated data set in Section 4 showed the properties of the estimation procedure and compared the censored coefficients specification with the usual uncensored coefficients specification, to study the influence censoring has on the precision of the outcome. Comparisons indicate that the estimation procedure is well suited to estimating the ratio of zero WTP with sufficient precision, and that the loss of accuracy for the estimated parameters is moderate, considering the high proportion of censored coefficients in the simulation study.

The application to the Tokyo Bay data set compared the censored random coefficients model to four different model specifications. Some key figures estimated under the different model specifications were given in Table 5 and show that the censored random coefficients model was the only one being flexibel enough to avoid negative estimates of WTP and revealing a differently pronounced resistance against abandoning the status-quo for each of the four valued attributes separately. The confidence intervals reported in Kuriyama et al. (1999) partly overlap the negative space, while credibility intervals of the model proposed in this article do not, and give the additional information which proportion of the respondents has a WTP of exactly zero. In this respect we conclude that the proposed model has some real advantages and may be suitable for further applications, especially in the fields of environmental valuation.

The specification of the model presented in Section 2 restricts the survey designer to revealing only two different alternatives to the respondents at one time. Since the take-it-or-leave-it format is very popular in stated-preference surveys, the model presented in this article is applicable to a large number of WTP studies. Nevertheless, restricting a survey to dichotomous choices only is not always desirable. In particular, in analysing revealed – not stated – preferences the number of alternatives is not (fully) under the control of the researcher, but is rather a matter of the environment and the circumstances of the study. For this reason an extension of the proposed model to allow the respondent to chose from an unlimited number of alternatives appears useful to the authors. Estimation of such a multinomial probit model is straightforward in the Gibbs sampling approach presented in Section 3. Only sampling the latent active surpluses $s_{it}$ in step 1. of the six sampling steps on page 11 needs to be carried out otherwise than in the dichotomous probit case. An overview of methods for sampling $s_{it}$ in the multinomial case, where it is a vector rather than a scalar, is given in Imai and van Dyk (2005) who also present their own method with some real advantages.

## Appendix A – Likelihood of the common parameters

The specification of the censored observations $y_{it}$ in (2) is that of the probit model. Hence, the probability of observing $\boldsymbol{y}_i$ conditional on parameters $\boldsymbol{\beta}_i$ and $\tau_i$ is given by

$$p(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \tau_i) = \prod_{t=1}^{T_i} \int_{A_{it}} \phi_i(y)dy,$$

where

$$A_{it} = \begin{cases} (-\infty, \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + p_{it}), & \text{if } y_{it} = 1, \\ [\boldsymbol{x}_{it}'\boldsymbol{\beta}_i + p_{it}, \infty), & \text{if } y_{it} = 0, \end{cases} \tag{18}$$

and $\phi_i(.)$ is the density of the univariate normal distribution with zero mean and variance $\tau_i$. The contribution of respondent $i$ to the unconditional likelihood of the common parameters is then given by

$$\ell_i = \int_0^\infty \dots \int_0^\infty p(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \tau_i)\; p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})\; p(\tau_i|r, \varrho) d\boldsymbol{\beta}_i d\tau_i. \tag{19}$$

The density of the unit-level scales $\tau_i$ is given in (5), while the density of the censored random coefficients $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ is obtained from the density of the latent random coefficients $p(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ as defined in (6). The main issue in the derivation of $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ is to take into account that it is a priori unknown which of the elements of $\boldsymbol{\beta}_i$ are 0. We denote the indicator holding the information which of the elements of $\boldsymbol{\beta}_i > 0$ by $\mathrm{I}_i$. For each $\boldsymbol{\beta}_i$ $M = 2^K$ permutation vectors $\mathrm{I}_{im}$ of length $K$ exist such that $\mathrm{I}_{imk} = 1$ if $\beta_{ik} > 0$ in the $m^{th}$ permutation. $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta})$ is then obtained by averaging the conditional density $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im})$ with respect to the probabilities $w_{im}$ of the permutations $\mathrm{I}_{im}$, such that

$$p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}) = \sum_{m=1}^{M} w_{im}\; p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im}). \tag{20}$$

To access the conditional density of $p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im})$, we start by writing down the conditional density of the latent $\boldsymbol{\beta}_i^*$, which is given by

$$p(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im}) = \frac{\phi(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}'\boldsymbol{z}_i, \boldsymbol{\Delta})}{\int\limits_{B_{im1}} \dots \int\limits_{B_{imK}} \phi(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}'\boldsymbol{z}_i, \boldsymbol{\Delta}) d\boldsymbol{\beta}_i^*} \tag{21}$$

where

$$B_{imk} = \begin{cases} (0, \infty), & \text{if } \mathrm{I}_{imk} = 1, \\ (-\infty, 0), & \text{if } \mathrm{I}_{imk} = 0, \end{cases}$$

and $\phi(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}'\boldsymbol{z}_i, \boldsymbol{\Delta})$ is the density of a $K$-variate normal distribution with mean vector $\boldsymbol{\gamma}'\boldsymbol{z}_i$ and covariance matrix $\boldsymbol{\Delta}$.

In a second step those elements of $\boldsymbol{\beta}_i^*$ are integrated out in (21) for which $\mathrm{I}_{im}$ indicates a negative sign to obtain the marginal distribution of the positive elements of $\boldsymbol{\beta}_i^*$. Denoting the positive elements of $\boldsymbol{\beta}_i^*$ in the $m^{th}$ permutation by $\boldsymbol{\beta}_{i,mu}^*$ and the negative elements by $\boldsymbol{\beta}_{i,mc}^*$, we get

$$p(\boldsymbol{\beta}_{i,mu}^*|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im}) = \int\limits_{-\infty}^{0} \dots \int\limits_{-\infty}^{0} p(\boldsymbol{\beta}_i^*|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im}) d\boldsymbol{\beta}_{i,mc}^*. \tag{22}$$

Finally, the conditional distribution of the effective $\boldsymbol{\beta}_i$ is obtained by the product of the marginal distribution $p(\boldsymbol{\beta}_{i,mu}^*|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im})$ evaluated at the values of the equivalent non-zero elements of $\boldsymbol{\beta}_i$, and the product of Dirac's point masses $\delta_0(.)$ evaluated for the $R$ elements of $\boldsymbol{\beta}_{i,mc}$, such that

$$p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im}) = p(\boldsymbol{\beta}_{i,mu}^* = \boldsymbol{\beta}_{i,mu}|\boldsymbol{\gamma}, \boldsymbol{\Delta}, \mathrm{I}_{im})\ \delta_0(\beta_{i,mc1}) \dots \delta_0(\beta_{i,mcR}). \tag{23}$$

Since the probabilities $w_{im}$ in (20) are equal to the integral in (21), the final unconditional distribution of $\boldsymbol{\beta}_i$ is given by

$$p(\boldsymbol{\beta}_i|\boldsymbol{\gamma}, \boldsymbol{\Delta}) = \sum_{m=1}^{M} \int_{-\infty}^{0} \dots \int_{-\infty}^{0} \phi(\boldsymbol{\beta}_i|\boldsymbol{\gamma}^{'} \boldsymbol{z}_i, \boldsymbol{\Delta}) d\boldsymbol{\beta}_{i,mc}\, \delta_0(\beta_{i,mc1}) \dots \delta_0(\beta_{i,mcR}). \tag{24}$$

Due to the point mass operators $\delta_0(.)$ in (24) all summands become zero, except for that $m$ for which the signs of $\boldsymbol{\beta}_i$ correspond to $\mathrm{I}_{im}$.

## Literature

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of Environmental Economics and Management 32*(1), 65–84.

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*(422), 669–679.

Allenby, G. M., N. Arora, and J. L. Ginter (1998). On the heterogeneity of demand. *Journal of Marketing Research 35*, 384–389.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman (1993). Report of the NOAA panel on contingent valuation. Technical report, National Oceanic and Athmospheric Administration.

Baltas, G. and P. Doyle (2001). Random utility models in marketing research: a survey. *Journal of Business Research 51*(2), 115 – 125.

Bateman, I. J., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Leeand, G. Loomes, S. Mourato, E. Özdemiroglu, D. W. P. OBE, R. Sugden, and J. Swanson (2002). *Economic Valuation with Stated Preference Techniques: A Manual.* Cheltenhan, UK: Edward Elgar Publishing Ltd.

Beenstock, M., E. Goldin, and Y. Haitovsky (1998). Response bias in a conjoint analysis of power outages. *Energy Economics 20*(2), 135–156.

Bohara, A. K., J. Kerkvliet, and R. P. Berrens (2001). Addressing negative willingness to pay in dichotomous choice contingent valuation: A Monte Carlo simulation approach. *Environmental and Resource Economics 20*(3), 173–195.

Carlsson, F. and P. Martinsson (2008). Does it matter when a power outage occurs? – A choice experiment study on the willingness to pay to avoid power outages. *Energy Economics 30*(3), 1232–1245.

Carson, R. T., R. C. Mitchell, W. M. Hanemann, R. J. Kopp, S. Presser, and P. A. Rudd (1992). A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. Report to the attorney general of the State of Alaska, Natural Damage Assessment Inc., San Diego.

Chandukala, S. R., Y. D. Edwards, and G. M. Allenby (2011). Identifying unmet demand. *Marketing Science 30*, 61 – 73.

Chib, S. (1992). Bayes inference in the tobit censored regression model. *Journal of Econometrics 51*(1), 79–99.

Clinch, J. P. and A. Murphy (2001). Modelling winners and losers in contingent valuation of public goods. *The Economic Journal 111*(470), 420–443.

Congdon, P. (2003). *Applied Bayesian Modelling*. Wiley Series in Probability and Statistics. Chichester, UK: John Wiley & Sons.

Cornick, J., T. L. Cox, and B. W. Gould (1994). Fluid milk purchases: A multivariate tobit analysis. *American Journal of Agricultural Economics 76*(1), 74–82.

Fong, D. K. H. and W. S. DeSarbo (2007). A Bayesian methodology for simultaneously detecting and estimating regime change points and variable selection in multiple regression models for marketing research. *Quantitative Marketing and Economics 5*, 427–453.

Frühwirth-Schnatter, S., R. Tüchler, and T. Otter (2004). Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics 22*, 2–15.

Frühwirth-Schnatter, S. and H. Wagner (2011). Bayesian variable selection for random intercept modeling of Gaussian and non-Gaussian data. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 9*, pp. 165–200. Oxford (UK): Oxford University Press.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC.

George, E. I. and R. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*, 339–373.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. New York: Wiley.

Greene, W. H., D. A. Hensher, and J. M. Rose (2004). Using classical simulation-based estimators to estimate individual WTP values. In A. Alberini and R. Scarpa (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 17–33. New York: Springer.

Habb, T. C. and K. E. McConnell (1997). Referendum models and negative willingness to pay: Alternative solutions. *Journal of Environmental Economics and Management 32*(2), 251–270.

Hackl, F. and G. J. Pruckner (1999). On the gap between open-ended and closed-ended CVM-answers. *Applied Economics 31*(6), 733–742.

Hanemann, W. M. and B. Kanninen (1999). The statistical analysis of discrete-response CV data. In I. Bateman and K. Willis (Eds.), *Valuing Environmental Preferences*, pp. 302–441. New York: Oxford University Press.

Hanley, N., S. Colombo, B. Kriström, and F. Watson (2008). Accounting for negative, zero and positive willingness to pay for landscape change in a national park. *Journal of Agricultural Economics 60*(1), 1–16.

Imai, K. and D. A. van Dyk (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics 124*(2), 311–334.

Kristroem, B. (1997). Spike models in contingent valuation. *American Journal of Agricultural Economics 79*(3), 1013–1023.

Kuriyama, K., K. Takeuchi, and T. Washida (1999). Repeated dichotomous choice contigent valuation for multi-attributed environmental resources: A comparisaon between contingent valuation and a choice experiment format. Environmental economics working paper #9902, School of Political Science and Economics, Waseda University.

Layton, D. F. and G. Brown (2000). Heterogenous preferences regarding global climate change. *The Review of Economics and Statistics 82*(4), 616–624.

Layton, D. F. and R. Levine (2003). How much does the far future matter? A hierarchical Bayesian analysis of the public's willingness to mitigate ecological impacts of climate change. *Journal of the American Statistical Association 98*(3), 533–544.

Layton, D. F. and K. Moeltner (2004). The cost of power outages to heterogenous households – an application of the mixed gamma-lognormal distribution. In A. Alberini and R. Scarpa (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 35–54. New York: Springer.

McFadden, D. (1996). Computing willingness-to-pay in random utility models. Working Papers 011, University of California at Berkeley, Econometrics Laboratory Software Archive.

Moeltner, K. and D. F. Layton (2002). A censored random coefficients model for pooled survey data with application to the estimation of power outage costs. *The Review of Economics and Statistics 84*(3), 552–561.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing.* Chichester: John Wiley and Sons, Ltd.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15*(4), 321–340.

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

Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica 38*(2), 311–323.

Train, K. E. (2003). *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

Wagner, H. and C. Duller (2011). Bayesian model selection for logistic regression models with random intercept. *Computational Statistics and Data Analysis.*

Werner, M. (1999). Allowing for zeros in dichotomous-choice contingent-valuation models. *Journal of Business and Economic Statistics 17*(4), 479–485.