A Mixture Model of Willingness to Pay Distributions [1]

YUYING AN and ROBERTO A. AYALA [2]

Duke University and *Central Bank, Ecuador*

This version: June 1996

Abstract

In this paper we propose a mixture model of willingness to pay distributions for contingent valuation studies. By allowing a point mass at zero, this model nests the conventional model as a special case. We discuss both parametric and nonparametric estimations of the mixture model. We consider estimation under two different data information settings for a double bounded dichotomous choice format. The implications of the mixture model in the estimation of the mean and the median of the distribution are presented.

Key Words: Mixture Model, Contingent Valuation, Zero Willingness to Pay.

JEL Classification: C25, C51, Q26.

[1] An earlier version of this paper was presented at Camp Resources 3, in August 1995 and circulated as An and Ayala (1995). We thank Timothy Haab, Glenn Harrison, Barbara Kanninen, John Loomis, Carol Mansfield, Kerry Smith and seminar participants at Cornell and Duke Universities for helpful comments. We are indebted to Barbara Harrow for her excellent editorial assistance. Financial support was provided by the Foundation Francisco Marroquín and Duke University Art and Science research Council. The usual disclaimer applies. An electronic version of the paper can be obtained in postscript format as "pub/man/papers/wtp.ps" on lewis.econ.duke.edu.

[2] Direct correspondence to Professor Mark An, Department of Economics, Duke University, Durham, NC 27708-0097. Telephone: (919) 660-1809. Email: man@econ.duke.edu.

1 Introduction

Contingent valuation (CV) is one of the most popular methods to assess the value of public goods. [3] It uses survey methods to elicit consumers' preferences by finding out how much consumers would be willing to pay for specified changes in the level of provision of a public good.

A stylized fact identified in previous CV studies is that the distribution of the willingness to pay (WTP) tends to be bimodal. Conventional models of WTP assume absolutely continuous distributions, therefore do not capture the bimodality. As a consequence, the estimation of the mean or median willingness to pay is often unbelievably high and severely imprecise.

In this paper we specify a mixture model that incorporates the possibility that a respondent's willingness to pay be actually zero. This situation can arise in one of two scenarios. First, as noticed by Kriström (1995), zero WTP represents the truncation point for those whose WTP is actually negative. Alternatively, it may well be the case that the proposed public good is so remote from the respondent's interest that he or she is completely indifferent to it. In this case, zero WTP represents honest responses. [4]

Modelling zero WTP is a continuing effort in the contingent valuation literature (McFadden 1994, Kriström's (1995), Werner 1995, Haab 1995). In this paper, we (1) specify the WTP distribution as a mixture of two distributions, one with a point mass at zero and the other with full support on the positive half of the real line; (2) concentrate on the statistical inference of such a mixture model using data from double bounded dichotomous choice surveys; and (3) discuss the added information from the follow-up question which asks a respondent whether or not she is willing to pay anything for the good in question.

[3] For a complete review of the contingent valuation method see Cummings *et al* (1986) and Mitchell and Carson (1989). For a critical view of the method see Hausman (1993) and McFadden (1994).

[4] In principle, true zero willingness to pay is different from what is called a "protest" vote, even though both may generate identical observations in the data. For a discussion of the two, see Mitchell and Carson (1989).

It is worth pointing out the relationship between our mixture model and the *spike* model due to Kriström (1995). Kriström assumes that the respondent's WTP might be either positive or negative. This is quite natural in some situations, such as city residents' valuation of a proposed local airport. People who travel frequently might have positive value due to the added convenience. People who do not travel at all might have negative value due to the added air noise and ground traffic. Therefore the WTP distribution should have support in the entire real line. Kriström argues that CV surveys often ask respondents' willingness to *pay* for the good (the proposed airport, say). These surveys generate data with a spike at zero which is the truncation at zero of the negative part of the distribution. Quantitatively, the height of the spike is the integration of the probability density from negative infinity to zero. Our mixture model includes this situation as a special case. In our setting, the zero WTP might be genuine indifference for a subpopulation. Quantitatively, the point mass in the mixture model is an extra free parameter. Since both the conventional model and the *spike* model can be interpreted as special cases of our mixture model, one can construct easy statistical tests for the validity of these models.

In the next section we present the way in which the current literature has conventionally modeled CV questions. In section 3 we discuss the estimation of the proposed model under two different information settings. We demonstrate how to test the validity of the conventional and the spike model of Kriström. We finish with empirical applications and a discussion of the implications that using this type of model will have on future CV studies.

2 The Conventional Model

The double bounded dichotomous choice (DBDC) question is the most frequently used elicitation method in contingent valuation studies. It was first proposed by Hanemann (1985) and Carson (1985), and first implemented by Carson, Hanemann and Mitchell (1985). A DBDC question presents each respondent a sequence of two bids and asks for a "yes" or "no" vote on whether the respondent's willingness to pay equals or exceeds each bid. The second bid is conditioned on the respondent's response to the first bid; lower if the first response is "no" and higher if it is "yes".

Let $i = 1, ..., N$ be the index for each respondent in the sample. In addition, let B_i be the original bid, and B_{iH} and B_{iL} be the higher and lower bids respectively. The four different outcomes from a DBDC question are represented by the following indicator variables:

$D_{1i} = 1$ iff the outcome is "no-no" $(WTP_i < B_{iL})$;

$D_{2i} = 1$ iff the outcome is "no-yes" $(B_{iL} \leq WTP_i < B_i)$;

$D_{3i} = 1$ iff the outcome is "yes-no" $(B_i \leq WTP_i < B_{iH})$;

$D_{4i} = 1$ iff the outcome is "yes-yes" $(B_{iH} \leq WTP_i)$.

Let P_{1i}, P_{2i}, P_{3i}, and P_{4i} respectively denote the probability associated with each of these outcomes. If we adopt a parametric assumption for the cumulative distribution function of WTP, $F(x;\theta)$, with parameter(s) θ, we can express these probabilities as

$$\begin{cases} P_{1i} = F(B_{iL};\theta); \\ P_{2i} = F(B_i;\theta) - F(B_{iL};\theta); \\ P_{3i} = F(B_{iH};\theta) - F(B_i;\theta); \\ P_{4i} = 1 - F(B_{iH};\theta). \end{cases} \tag{1}$$

Then, the sample log-likelihood function is

$$l(\theta) = \sum_{i=1}^{N} \left\{ D_{1i} \log P_{1i} + D_{2i} \log P_{2i} + D_{3i} \log P_{3i} + D_{4i} \log P_{4i} \right\}. \tag{2}$$

The maximum likelihood estimator ($\hat{\theta}$) for θ is obtained by maximizing expression (2). The asymptotic variance-covariance matrix of $\hat{\theta}$ can be estimated by inverting the negative of the Hessian matrix,

$$\hat{V} = -\left[\frac{\partial^2 l(\hat{\theta})}{\partial\theta\partial\theta'}\right]^{-1}. \tag{3}$$

With the maximum likelihood estimator for θ, the mean (μ) and the p-th quantile (m_p) of the willingness to pay distribution can be easily estimated as

$$\hat{\mu} = \int_0^\infty [1 - F(x; \hat{\theta})]\, dx, \tag{4}$$

and

$$\hat{m}_p = F^{-1}(p; \hat{\theta}). \tag{5}$$

For example, for a Weibull distribution the formula for the mean is

$$\hat{\mu} = \hat{\gamma}^{-1/\hat{\alpha}} \Gamma\left(1 + \frac{1}{\hat{\alpha}}\right), \tag{6}$$

where $\Gamma(\bullet)$ is the gamma function. For the p-th quantile, the formula is

$$\hat{m_p} = \left[-\frac{1}{\hat{\gamma}} \log(1 - p)\right]^{1/\hat{\alpha}}. \tag{7}$$

If $p = 0.5$, expression (7) is the estimator for the median of the distribution.

3 The Mixture Model

The model in section 2 does not incorporate the possibility of zero WTP because the cumulative distribution function of WTP, $F(x; \theta)$, is usually assumed to be absolutely continuous. In this section we propose a mixture model that not only incorporates zero WTP, but also nests model (2) as a special case, so that specification tests can be easily conducted against the conventional model.

Let us assume the cumulative distribution function of the true WTP to have the following form:

$$G(x; \rho, \theta) = \begin{cases} 0, & \text{if} \quad x < 0 \\ \rho, & \text{if} \quad x = 0 \\ \rho + (1 - \rho)F(x; \theta), & \text{if} \quad x > 0 \end{cases}, \tag{8}$$

where $F(x; \theta)$ is an absolutely continuous cumulative distribution function such that $F(0; \theta) = 0$. As we can see from expression (8), $G(x; \rho, \theta)$ is not absolutely continuous. It has a point mass at $x = 0$, represented by the parameter ρ.

Model (8) is called mixture model because it can be interpreted as the mixture of two distributions. With probability ρ the willingness to pay is drawn from the first distribution which has a unit mass at $x = 0$. With probability $1 - \rho$, the willingness to pay is drawn from the second distribution $F(x;\theta)$. It is obvious that if $\rho = 0$, the mixture model specializes to the conventional model discussed in section 2.

To see the relationship between the mixture model and the spike model of Kriström, notice that for the spike model the support of the WTP distribution, $H(x;\theta)$, is the whole real line, $(-\infty, \infty)$. Therefore if we restrict

$$\begin{cases} \rho = \int_{-\infty}^{0} dH(x;\theta) = H(0;\theta) \\ \\ F(x;\theta) = \int_{0}^{x} (1-\rho)^{-1} dH(y;\theta) = (1-\rho)^{-1}[H(x;\theta) - H(0;\theta)], \end{cases} \tag{9}$$

the mixture model specializes to the spike model.

3.1 Estimation under Two Data Information Settings

In practice, the researcher may face one of two data settings, depending on whether or not a follow-up question –"are you willing to pay anything for the good?"– is asked to those respondents who answered "no-no" to the first two questions. In what follows we refer as case 1 or "partial information case" (PIC) to situations where the researcher does not have information from this follow-up question, and as case 2 or "full information case" (FIC) to situations where this information is available.

Using the same notation as in section 2, the sample log-likelihood function for case 1 (PIC) takes the following form:

$$\begin{aligned} l_1(\rho,\theta) = \sum_{i=1}^{N} \{ & D_{1i} \log[\rho + (1-\rho)P_{1i}] + D_{2i} \log[(1-\rho)P_{2i}] \\ & + D_{3i} \log[(1-\rho)P_{3i}] + D_{4i} \log[(1-\rho)P_{4i}] \}, \end{aligned} \tag{10}$$

where P_{ji} for $j = 1, 2, 3, 4$ are defined in expression (1) above. Denote the maximum likelihood estimator by

$$(\hat{\rho}_1, \hat{\theta}_1) = \arg \max_{(\rho,\theta)} \; l_1(\rho, \theta).$$

In case 2 (FIC), we need to include an extra indicator variable. Let us call this new binary variable D_{0i}, which will equal 1 in the case of a "no-no-no" answer, and 0 otherwise. The log-likelihood function takes the following form:

$$l_2(\rho, \theta) = \sum_{i=1}^{N} \{D_{0i} \log \rho + (D_{1i} - D_{0i}) \log[(1-\rho)P_{1i}] + D_{2i} \log[(1-\rho)P_{2i}]$$
$$+ D_{3i} \log[(1-\rho)P_{3i}] + D_{4i} \log[(1-\rho)P_{4i}]\} . \tag{11}$$

Denote the maximum likelihood estimator under this case by

$$(\hat{\rho}_2, \hat{\theta}_2) = arg \max_{(\rho,\theta)} \; l_2(\rho, \theta).$$

3.2 Theoretical Implications

We now study some theoretical implications of using the mixture model under these two different settings. The first proposition states that the mixture model can be estimated even in the "partial information case" (PIC). In the "full information case" (FIC), we can obtain an analytical solution for the maximum likelihood estimator of ρ and the exact sampling distribution of $\hat{\rho}$, as stated in proposition 2. The third proposition provides a comparison of the variance-covariance matrices of the maximum likelihood estimators in the two information settings.

Proposition 1 *Under the PIC, the parameter ρ is identified and can be consistently estimated.*

Proof: From (10) it can be shown that

$$\frac{\partial l_1(\rho, \theta)}{\partial \rho} = \sum_{i=1}^{N} \left\{ \frac{D_{1i}[1 - P_{1i}]}{\rho + (1-\rho)P_{1i}} - \frac{1 - D_{1i}}{(1-\rho)} \right\}, \tag{12}$$

$$\frac{\partial^2 l_1(\rho,\theta)}{\partial \rho^2} = \sum_{i=1}^{N} \left\{ -\frac{D_{1i}[1-P_{1i}]^2}{[\rho+(1-\rho)P_{1i}]^2} - \frac{1-D_{1i}}{(1-\rho)^2} \right\}. \tag{13}$$

Since (12) always has a root, and (13) is always negative, ρ is therefore identified. Consistency follows from the well-known maximum likelihood estimation properties. Q.E.D.

Proposition 2 *Under the FIC:*

(i) $\hat{\rho}_2$ *is the sample proportion of "no-no-no" respondents;*

(ii) $\hat{\rho}_2$ *is uncorrelated with* $\hat{\theta}_2$*;*

(iii) $N\hat{\rho}_2$ *is Binomial with parameters* (N,ρ)*;*

(iv) $E\hat{\rho}_2 = \rho$*;* $Var(\hat{\rho}_2) = \frac{1}{N}\rho(1-\rho)$.

Proof: From expression (11),

$$\frac{\partial l_2(\rho,\theta)}{\partial \rho} = \sum_{i=1}^{N} \left\{ \frac{D_{0i}}{\rho} + \frac{(1-D_{0i})}{1-\rho} \right\}. \tag{14}$$

Let N_0 be the number of "no-no-no" responses. Then from expression (14),

$$\hat{\rho}_2 = \frac{N_0}{N}. \tag{15}$$

This proves (i). Notice that (14) does not depend on θ. The cross-partial derivative is zero. This shows that there is no correlation between $\hat{\rho}_2$ and $\hat{\theta}_2$, which is (ii). The number of "no-no-no" responses, N_0, can be expressed as the sum of N independent and identically distributed Bernoulli variables: $N_0 = \sum_{i=1}^{N} D_{0i}$. Therefore, $N\hat{\rho}_2$ is distributed as a binomial with parameters N and ρ. This is (iii). Part (iv) is a corollary of (iii). Q.E.D.

Consistency and asymptotic normality of the estimators $(\hat{\rho}_1, \hat{\theta}_1)$ and $(\hat{\rho}_2, \hat{\theta}_2)$ are a direct consequence of the maximum likelihood estimation. However, the next theoretical implication allows us to evaluate these estimators in terms of efficiency by comparing their variance-covariance matrices.

Proposition 3 *$V_1 - V_2$ is a positive semi-definite matrix, where V_1 and V_2 are the asymptotic variance-covariance matrices of $(\hat{\rho}_1, \hat{\theta}_1)$ and $(\hat{\rho}_2, \hat{\theta}_2)$ respectively.*

Proof: Let $\delta = (\rho, \theta)$ represent the vector of parameters. To make the exposition simpler, let us define $Q_{0i} = \rho$ and $Q_{ji} = (1-\rho)P_{ji}$ for $j = 1, 2, 3, 4$.

As before, the i index refers to the individual observation, and P_{ji} for $j = 1, 2, 3, 4$ is defined by (1). Using this notation we can write the log-likelihood functions for case 1 (PIC) and case 2 (FIC) defined above, as follows:

$$l_1(\delta) = \sum_{i=1}^{N} \{D_{1i} \log[Q_{0i}(\delta) + Q_{1i}(\delta)] + D_{2i} \log Q_{2i}(\delta) + D_{3i} \log Q_{3i}(\delta) + D_{4i} \log Q_{4i}(\delta)\}. \quad (16)$$

$$l_2(\delta) = \sum_{i=1}^{N} \{D_{0i} \log Q_{0i}(\delta) + (D_{1i} - tD_{0i}) \log Q_{1i}(\delta) + D_{2i} \log Q_{2i}(\delta) + D_{3i} \log Q_{3i}(\delta) + D_{4i} \log Q_{4i}(\delta)\}. \quad (17)$$

The score vector and the information matrix for each case are given by the following expressions:

$$S_{1i}(\delta) = \frac{D_{1i}[Q'_{0i}(\delta) + Q'_{1i}(\delta)]}{Q_{0i}(\delta) + Q_{1i}(\delta)} + \sum_{j=2}^{4} \left\{ \frac{D_{ji} Q'_{ji}(\delta)}{Q_{ji}(\delta)} \right\}, \quad (18)$$

$$S_{2i}(\delta) = \frac{D_{0i} Q'_{0i}(\delta)}{Q_{0i}(\delta)} + \frac{(D_{1i} - D_{0i}) Q'_{1i}(\delta)}{Q_{1i}(\delta)} + \sum_{j=2}^{4} \left\{ \frac{D_{ji} Q'_{ji}(\delta)}{Q_{ji}(\delta)} \right\}, \quad (19)$$

$$I_{1i}(\delta) = E[S_{1i}(\delta) S_{1i}(\delta)^T] = \frac{[Q'_{0i}(\delta) + Q'_{1i}(\delta)][Q'_{0i}(\delta) + Q'_{1i}(\delta)]^T}{Q_{0i}(\delta) + Q_{1i}(\delta)} + \sum_{j=2}^{4} \frac{Q'_{ji}(\delta) Q'_{ji}(\delta)^T}{Q_{ji}(\delta)}, \quad (20)$$

$$I_{2i}(\delta) = E[S_{2i}(\delta) S_{2i}(\delta)^T] = \sum_{j=0}^{4} \frac{Q'_{ji}(\delta) Q'_{ji}(\delta)^T}{Q_{ji}(\delta)}, \quad (21)$$

where the apostrophe ($'$) represents the first derivative with respect to the vector of parameters δ, and the superscript T represents the transpose of the matrix. It can easily be shown by subtracting $I_{1i}(\delta)$ from $I_{2i}(\delta)$ that the difference is a matrix of the form

$$\Delta_i = I_{2i}(\delta) - I_{1i}(\delta) = \frac{[Q'_{0i}(\delta)Q_{1i}(\delta) - Q'_{1i}(\delta)Q_{0i}(\delta)][Q'_{0i}(\delta)Q_{1i}(\delta) - Q'_{1i}(\delta)Q_{0i}(\delta)]^T}{Q_{0i}(\delta)Q_{1i}(\delta)[Q_{0i}(\delta) + Q_{1i}(\delta)]}. \quad (22)$$

Δ_i is a positive semi-definite matrix since the numerator is the outer product of two identical vectors. This implies that $\sum_{i=1}^{N} I_{2i} - \sum_{i=1}^{N} I_{1i}$ is also positive semi-definite. Therefore, $V_1 - V_2 = \left[\sum_{i=1}^{N} I_{1i}\right]^{-1} - \left[\sum_{i=1}^{N} I_{2i}\right]^{-1}$ is positive semi-definite. Q.E.D.

3.3 Estimation of the Mean and the Median

An important objective of CV studies is to estimate the mean and/or median of the WTP distribution. We now examine the theoretical implications of the mixture model with respect to the estimation of these welfare measures. Let $\hat{\mu}_j$ be the maximum likelihood estimator for the mean and $\hat{m}_j$ be the estimator for the median. The j index takes a value of 0 for the conventional model and values of 1 and 2 for the mixture model under PIC and FIC respectively. Then, we can show the following:

Proposition 4 *If $\rho \neq 0$, then:*

(i) $\hat{\mu}_0$ ($\hat{m}_0$) is inconsistent;

(ii) Both $\hat{\mu}_1$ ($\hat{m}_1$) and $\hat{\mu}_2$ ($\hat{m}_2$) are consistent;

(iii) $\hat{\mu}_2$ ($\hat{m}_2$) is asymptotically more efficient than $\hat{\mu}_1$ ($\hat{m}_1$).

Proof: Under model (8) the true mean WTP can be written as $\mu = (1 - \rho)\int_0^\infty [1 - F(x;\theta)]\,dx$. If $\rho \neq 0$, the conventional model is mis-specified; hence, $\hat{\mu}$ is inconsistent. Since for $j = 1, 2$, $\hat{\mu}_j = f(\hat{\delta}_j)$, consistency of $\hat{\delta}_j$ implies that $\hat{\mu}_j$ is also consistent. The relative efficiency of $\hat{\mu}_2$ over $\hat{\mu}_1$ is a direct consequence of proposition 3. The proof for the median ($\hat{m}_j$) case follows a similar argument. Q.E.D.

3.4 Tests of the Conventional Model

To test the validity of the conventional model, we need to test $H_o : \rho = 0$ in the mixture model. The only complication is that under the null hypothesis, the true parameter ρ is on the boundary of the parameter space $[0, 1]$, and therefore, a one-sided t-test would be required.

3.5 Tests of the Spike Model

In order to estimate the spike model, we need to assume a WTP distribution function with support in the whole real line $(-\infty, \infty)$. Following Kriström's example, consider the following logistic distribution for the willingness to pay:

$$H(x; \theta) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}, \tag{23}$$

with $H(0, \theta) = e^{\alpha}/(1 + e^{\alpha})$. We can test this version of the spike model making use of the restrictions provided by expression (9). For that, we estimate the mixture model of the following form:

$$G(x; \rho, \theta) = \begin{cases} 0, & \text{if} \quad x < 0 \\ \rho, & \text{if} \quad x = 0 \\ \rho + (1 - \rho)\left[(1 + e^{\alpha})H(x; \alpha, \beta) - e^{\alpha}\right], & \text{if} \quad x > 0 \end{cases} . \tag{24}$$

Then, the validity of the "spike" model can be easily verified by testing the simple nonlinear restriction given by

$$H_o : \ \rho = \frac{e^{\alpha}}{1 + e^{\alpha}}. \tag{25}$$

3.6 Model with Covariates

So far we have discussed the model as if there were no covariates. This is not the case in most economic situations. Economists strongly believe that individuals' preferences are influenced by personal and demographic characteristics such as income, education, age, sex, etc. When conducting CV surveys, the researcher usually tries to obtain data on attitudes and demographic characteristics of respondents. The link between this type of data and the stated WTP amount is often used as a means of checking consistency in responses and validity of the survey.

In the case where covariates are to be modeled, we can incorporate them by letting

$$\rho_i = \frac{e^{Z_i'\beta}}{1 + e^{Z_i'\beta}}, \tag{26}$$

where Z_i is the vector of covariates for respondent i and β is the vector of deep parameters to be estimated. Covariates can also be introduced directly in the cumulative distribution function $F(x|W_i'\beta;\theta)$, where the vectors Z_i and W_i may or may not coincide.

4 Empirical Applications

For the data sets we are going to use in this section we will assume that the WTP follows a Weibull distribution. The cumulative distribution function for this case is

$$F(x;\gamma,\alpha) = 1 - e^{-\gamma x^{\alpha}}. \tag{27}$$

According to the conventional model, the log-likelihood function is

$$l(\gamma,\alpha) = \sum_{i=1}^{N} \Big\{ D_{1i}\log[1 - e^{-\gamma B_{iL}{}^{\alpha}}] + D_{2i}\log[e^{-\gamma B_{iL}{}^{\alpha}} - e^{-\gamma B_i{}^{\alpha}}]$$

$$+D_{3i}\log[e^{-\gamma B_i{}^{\alpha}} - e^{-\gamma B_{iH}{}^{\alpha}}] + D_{4i}\log[e^{-\gamma B_{iH}{}^{\alpha}}]\Big\}. \tag{28}$$

The corresponding formulas for the mean and the median of the WTP when using a Weibull distribution are given by expressions (6) and (7) in section 2 above.

On the other hand, the mixture model, given by expressions (10) and (11), takes the following form:

$$l_1(\gamma,\alpha) = \sum_{i=1}^{N} \Big\{ D_{1i}\log[\rho + (1-\rho)(1 - e^{-\gamma B_{iL}{}^{\alpha}pha})]$$

$$+D_{2i}\log[(1-\rho)(e^{-\gamma B_{iL}{}^{\alpha}} - e^{-\gamma B_i{}^{\alpha}})]$$

$$+D_{3i}\log[(1-\rho)(e^{-\gamma B_i{}^\alpha}-e^{-\gamma B_{iH}{}^\alpha})]$$
$$+D_{4i}\log[(1-\rho)(e^{-\gamma B_{iH}{}^\alpha})]\Big\}, \tag{29}$$

$$l_2(\gamma,\alpha)=\sum_{i=1}^{N}\Big\{D_{0i}\log\rho+D_{1i}\log[(1-\rho)(1-e^{-\gamma B_{iL}{}^\alpha})]$$
$$+D_{2i}\log[(1-\rho)(e^{-\gamma B_{iL}{}^\alpha}-e^{-\gamma B_i{}^\alpha})]$$
$$+D_{3i}\log[(1-\rho)(e^{-\gamma B_i{}^\alpha}-e^{-\gamma B_{iH}{}^\alpha})]$$
$$+D_{4i}\log[(1-\rho)(e^{-\gamma B_{iH}{}^\alpha})]\Big\}. \tag{30}$$

For calculating the mean and the median WTP, the corresponding formulas are

$$\mu=E(WTP)=(1-\rho)\gamma^{-1/\alpha}\Gamma\left(1+\frac{1}{\alpha}\right), \tag{31}$$

$$m=\begin{cases}\left[-\frac{1}{\gamma}\log\frac{0.5}{1-\rho}\right]^{1/\alpha}, & \text{if } \rho\le 0.5\\ 0, & \text{if } \rho>0.5.\end{cases} \tag{32}$$

4.1 The Data

The data we use comes from two different contingent valuation studies. The first one was conducted by Imber, Stevenson, and Wilks (1991) to calculate the WTP to prevent possible environmental damage from mining in the Conservation Zone located within the boundaries of the Kakadu National Park in Australia. The second one was conducted by Hanemann, Loomis, and Kanninen (1991) to elicit the WTP for protecting wetland habitats and wildlife in California's San Joaquin Valley. [5]

The Kakadu survey was conducted door to door with one person interviewed in each selected household. The interview process began on September 8th, 1990 and was completed by September 26th. Due to the fact that the environmental damage from the proposed mine was still in dispute at the time the survey took place, two different impact scenarios were developed.

[5] In fact, Hanemann *et al* evaluated five different environmental programs in their survey. We focus only on the first of these programs: protection of wetlands and wildlife.

One half of the sample was presented with a description of the mine that suggested little risk of damage (minor impact scenario). The other half received information that argued that the risks of damage were significant and that the impact could be very substantial (major impact scenario).

The San Joaquin survey, on the other hand, was a combination of mail and telephone media. Households were chosen based on random digit sampling. They were asked if they were willing to participate in a survey. If they accepted, a mail questionnaire was sent and a certain time was arranged for the interviewer to call the household back. The interview process was conducted in May 1989 in three geographical areas: (1) the San Joaquin Valley, (2) the rest of the state of California, and (3) the states of Oregon, Washington and Nevada as representatives of the "out-of-state" population. Of the 1960 households originally contacted, 1239 (63.1%) agreed to participate. From these, only 1004 completed interviews were collected: 227 from the San Joaquin Valley, 576 from the rest of California, and 201 from "out-of-state". In this empirical application we used the same data that Hanemann *et al* used for their paper: the 576 interviews from the rest of the California sample. [6]

Both surveys used a double bounded dichotomous choice format. [7] However, the San Joaquin study included a feature that was not exploited. In particular, for people who gave a "no-no" response, a third question was asked: "would you vote for such a program at any cost?". The answers to this last question allow us to estimate the mixture model for the "full information case" (FIC) given by expression (30).

4.2 Results

All the results of this paper were obtained by using the maximum likelihood application of GAUSS 386i. The results presented in Table 1 correspond to

[6] The valid number of observations for the program we used in this application is 569. We dropped 7 "don't know-don't know-don't know" observations which did not provide any information.

[7] The sets of bids used in the Kakadu Study are: (5,2,20), (20,5,50), (50,20,100), and (100,50,550), where the first element of each set corresponds to the original bid, the second element to the lower bid, and the third element to the higher bid. In the San Joaquin study, the sets of bids are: (40,25,80), (50,25,110), (65,30,125), (80,40,125), and (110,55,170).

the estimation of the conventional model and mixture model for the partial information case (PIC) from the Kakadu study. [8] All the estimated parameters in both scenarios (major and minor) are statistically significant. The results reject the hypothesis that ρ is equal to zero. In fact, the estimated ρ is 19% and 26% for the major and minor impact scenarios respectively. This indicates that a substantial proportion of the population is indifferent to the argued environmental damage. In their original study, Imber *et al* (1991) noticed the bimodality of the WTP distribution. This feature of the data is better captured by the mixture model than by the conventional one.

The improvement in terms of welfare measures is very significant. The values for the mean WTP obtained from the conventional model are not reliable as point estimators of WTP. This is a consequence of fitting a bimodal distribution with a unimodal Weibull. The mean values from the mixture model, on the other hand, are reasonable. The standard deviations for the mean and median are drastically reduced when the mixture model is fit to the data.

Table 2 allows a direct comparison of the results obtained from the conventional model with the two versions of the mixture model for the San Joaquin study. Notice that the estimator for ρ in both cases of the mixture model is significant. As expected, this estimator for the FIC is the sample average of "no-no-no" responses. [9] The decrease in the standard deviations of the mean and the median, as we move from the conventional model to the

mixture model (PIC) to the mixture model (FIC), shows an improvement in terms of efficiency of the welfare measures.

Comparing the second and third columns of table 2, we can notice the reduction in the standard deviations of the estimated parameters. For γ, there is an improvement by a factor of 3.7; for α, by a factor of 1.5; for ρ, of 1.9. The elements of the diagonal of the variance-covariance matrix for the FIC are lower than the ones for the PIC, as proposition 3 above suggests.

Table 3 presents the results of estimating the mixture model (FIC) with

[8] As mentioned before, data on "no-no-no" responses was not available for this study.

[9] The number of "no-no-no" responses is 57 and the sample size is 569.

covariates. The second column illustrates the case in which the regressors are only introduced directly in the cumulative distribution function. We define $\gamma = e^{Z_i'\beta}$, where Z_i is the vector of regressors. In the third column, we incorporate the covariates in the parameter ρ by using expression (26). Notice that the signs of the deep parameters are as expected. However, most of them are statistically insignificant. The parameter ρ in the first column equals the proportion of "no-no-no" respondents. [10]

Table 4 compares the estimation results from the spike model and the mixture model. For this empirical application, we cannot reject the hypothesis that $\rho = (1 + e^{-\alpha})^{-1}$. The estimated parameters as well as the welfare estimates are almost identical for both models. Even the standard deviations of the mean and median take similar values. Therefore, for this particular case, it is equivalent to fit the data using the mixture model with a truncated logistic or the spike model with a logistic distribution.

5 Conclusion and Possible Extensions

In this paper we proposed a mixture model which generalizes the conventional way to model willingness to pay responses, by allowing the distribution to have a possible point mass at zero. The mixture model is better than the conventional model in capturing the common bimodality feature of the willingness to pay distribution. The mixture model nests both the conventional model and the "spike" model due to Kriström as special cases. Statistical tests of such restrictions can be easily done.

We discussed the estimation of the mixture model under two different data information settings: a "full information case", when data on "no-no-no" responses are available, and a "partial information case", when this type of data does not exist. One of the main advantages

of the mixture model is that even in the PIC, it performs better than the conventional model. The probability of zero willingness to pay, represented by the parameter ρ, is separately identified and can be consistently estimated.

[10] The number of "no-no-no" respondents equals 55. The number of observations is reduced to 543 because respondents with unreported incomes were not included.

In the case when data on "no-no-no" responses are available, an improvement in terms of efficiency of the parameters and of the welfare measures is achieved when moving from the mixture model (PIC) to the mixture model (FIC). Although this is a natural consequence of having more data, it shows a potential trade-off that CV practitioners face. Given a base design of DBDC surveys, two ways of improving the efficiency of the estimations are either to increase the sample size, or to ask an additional question only to the "no-no" respondents. Due to the usually high costs of increasing the sample size, asking a follow-up question seems to be an inexpensive way of achieving this objective.

Welfare measures have become the most important goal of CV studies. Therefore, efficiency gains in the estimation of the mean and the median are highly important not only for theoretical reasons, but also for the increasing use of CV surveys for natural resource damage assessments. In this regard, when the null hypothesis that $\rho = 0$ is not true, the mixture model provides consistent estimates for the mean and the median while the conventional model does not.

Even though this paper refers to an application of the mixture model to a double bounded dichotomous choice setting, the model is flexible enough to be applied to any of the elicitation methods used in contingent valuation studies. For example if the the single bounded dichotomous choice questions are used, with or without the follow-up question asking whether the respondent is willing to pay anything, the estimation and testing procedure will be similar. [11]

In this paper we have concentrated on parametric estimation of the mixture model. Nonparametric estimation of the mixture model is the subject of An and Ayala (1996) where we propose a self-consistent algorithm which generalizes the algorithm of Turnbull (1974, 1976). Unlike the parametric setting discussed here, in the non-parametric setting, the identification of the mixture model rely crucially on the availability of the follow-up question.

[11] Single bounded dichotomous choice questions were first implemented in CV studies by Bishop and Heberlein (1979).

References

An, Yuying and Roberto Ayala (1995): "A Mixture Model of Willingness to Pay Distributions," Department of Economics Working Paper #95-21, Duke University.

An, Mark Y. and Roberto Ayala (1996): "Nonparametric Estimation of a Survivor Function with Across-Interval-Censored Data," Department of Economics Working Paper #96-??, Duke University.

Bishop, Richard C. and Thomas A. Heberlein (1979): "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?". *American Journal of Agricultural Economics*, Vol.**61**, pp. 926-930.

Carson, Richard T. (1985): "Three Essays on Contingent Valuation". Ph.D. thesis, University of California, Berkeley.

Carson, Richard T., W. Michael Hanemann, and Robert C. Mitchell (1985): "Determining the Demand for Public Goods by Simulating Referendums at Different Tax Prices". Department of Economics Working Paper, University of California, San Diego.

Cummings, R.G., D.S. Brookshire, and W.D. Schulze (1986): *"Valuing Environmental Goods: An Assessment of the Contingent Valuation Method"*, Totowa NJ, Rowman & Allanheld.

Haab, Timothy (1995): "The Distributional Impacts of Indifference on Dichotomous Choice Contingent Valuation". Working paper, East Carolina University, November 1995.

Hanemann, W. Michael (1985): "Some Issues in Continuous- and Discrete-Response Contingent Valuation Studies". *Northeastern Journal of Agricultural and Resource Economics*, pp. 5-13.

Hanemann, W. Michael, John Loomis, and Barbara Kanninen (1991): "Statistical Efficiency of Double-Bounded Dichotomous Choice Contingent Valuation". *American Journal of Agricultural Economics*, Vol.**73**, pp. 1255-1262.

Hausman, Jerry (1993): *“Contingent Valuation: a Critical Assessment”*. North-Holland, Amsterdam, New York.

Imber, David, Gay Stevenson, and Leanne Wilks (1991): “A Contingent Valuation Survey of the Kakadu Conservation Zone”. Resource Assessment Commission Research Paper No. 3.

Kriström, Bengt, (1995): “Spike Models in Contingent Valuation: Theory and Illustrations”. Invited paper to the First Toulouse Conference on Environmental and Resource Economics, Toulouse, France, March 30-31, 1995.

McFadden, Daniel (1994): “Contingent Valuation and Social Choice”. *American Journal of Agricultural Economics*, Vol.**76**, pp. 689-708.

Mitchell, Robert C. and Richard T. Carson (1989): *“Using Surveys to Value Public Goods: The Contingent Valuation Method”*. Washington DC, Resources for the Future.

Turnbull, Bruce (1974): “Nonparametric Estimation of a Survivorship Function with Doubly Censored Data,” *Journal of the American Statistical Association*, Vol. **69**, pp. 169-173.

Turnbull, Bruce (1976): “The Empirical Distribution Function with Arbitrary Grouped, Censored, and Truncated Data,” *Journal of the Royal Statistical Society*, Series B, pp. 290-295.

Werner, Megan (1995): “Allowing for Indifference in Dichotomous Choice Contingent Valuation Models”. Paper presented at the 1996 Annual Convention of the Allied Social Science Associations, San Francisco, CA, January 5-7, 1996.

Table 1: Estimation Results for the Kakadu Data Using a Weibull Distribution

	Major Scenario		Minor Scenario	
	Conventional Model (1)	**Mixture Model (PIC) (2)**	**Conventional Model (3)**	**Mixture Model (PIC) (4)**
γ	**0.13467**** (0.01407)	**0.01042**** (0.00390)	**0.25415**** (0.02098)	**0.03097**** (0.01244)
α	**0.32349**** (0.02192)	**0.76607**** (0.07067)	**0.22468**** (0.01713)	**0.55702**** (0.07243)
ρ		**0.19845**** (0.01748)		**0.26302**** (0.02454)
Mean WTP	**3311.46** (1418.68)	**362.67** (65.26)	**21498.16** (15669.52)	**629.89** (221.49)
Median WTP	**158.33** (25.17)	**200.31** (14.15)	**86.97** (17.71)	**93.52** (13.81)
Log-likelihood	-1086.37	-1063.59	-1092.33	-1081.27
N	1018	1018	1011	1011

- Standard deviations are in parenthesis
- (**) significant at 1% level

Table 2: Estimation Results for the San Joaquin Data Using a Weibull Distribution

	Conventional Model (1)	**Mixture Model (PIC) (2)**	**Mixture Model (FIC) (3)**
γ	**0.002073****	**0.000374**	**0.000101***
	(0.000755)	(0.000333)	(0.000056)
α	**1.202548****	**1.536905****	**1.793688****
	(0.076429)	(0.178628)	(0.115243)
ρ		**0.06364****	**0.100176****
		(0.02427)	(0.01257)
Mean WTP	**160.18**	**143.16**	**135.26**
	(10.71)	(9.51)	(6.47)
Median WTP	**125.62**	**125.40**	**125.65**
	(6.39)	(5.48)	(5.04)
Log-likelihood	-678.99	-673.82	-726.56
N	569	569	569

- Standard deviations are in parenthesis
- (*) significant at 5% level
- (**) significant at 1% level

Table 3: Estimation Results for the San Joaquin Data Using a Weibull Distribution with Covariates

	Mixture Model: **FIC (1)**	**Mixture Model: FIC (2)**	**Mixture Model: FIC (3)**
α	**1.7824****	**1.7824****	**1.7824****
	(0.1196)	(0.1196)	(0.1196)
γ :	–	**0.000102***	–
		(0.00006)	
INTERCEPT	**-8.6020****		**-8.6020****
	(0.8625)		(0.8625)
YEARCA	**-0.655**		**-0.655**
	(0.458)		(0.458)
EDUCATION	**-0.0064**		**-0.0064**
	(0.0452)		(0.0452)
INCOME	**-0.671****		**-0.671****
	(0.259)		(0.259)
ρ :	**0.1014****		
	(0.0473)		
INTERCEPT		**-0.8993**	**-0.8993**
		(0.9697)	(0.9697)
YEARCA		**-0.0108**	**-0.0108**
		(0.0094)	(0.0094)
EDUCATION		**-0.0589**	**-0.0589**
		(0.0686)	(0.0686)
INCOME		**-0.0060**	**-0.0060**
		(0.0054)	(0.0054)
Log-likelihood	-670.89	-673.99	-668.98
No. of Obs.	543	543	543

- Standard deviations are in parenthesis
- (*) significant at 5% level, (**) significant at 1% level
- Yearca in 10 years, income in $ 10,000

Table 4: Testing the Spike Model for the San Joaquin Data Using a Logistic Distribution

	Spike Model	Mixture Model: FIC
α	**-2.2307**** (0.1279)	**-2.3836**** (0.2870)
β	**0.0177**** (0.0010)	**0.0187**** (0.0019)
ρ	**0.0970**** (0.0112)	**0.1001**** (0.0126)
Mean WTP	**131.79**** (4.84)	**129.92**** (5.45)
Median WTP	**126.02**** (4.82)	**125.60**** (4.69)
Log-likelihood	-732.08	-731.90
No. of Obs.	569	569

• Standard deviations are in parenthesis
• (**) significant at 1% level
$\chi^2_{(1)} = 0.36$. Hence, it is insignificant