FLEXIBLE MIXTURE DISTRIBUTION MODELING OF DICHOTOMOUS CHOICE CONTINGENT VALUATION WITH HETEROGENITY

Jorge E. Araña[1] and Carmelo J. León [2 (*)]

[1] University of California, Berkeley
Department of Agriculture and Resource Economics
207 Giannini Hall # 3310, University of California
Berkeley CA 94729-3310, USA
e-mail: Jorgea@empresariales.ulpgc.es

[2] University of Las Palmas de Gran Canaria
Department of Applied Economic Analysis
Edificio de Económicas, Modulo D-3.16
35017 Las Palmas, Spain
e-mail: Carmelo@empresariales.ulpgc.es

[(*)] Corresponding author

JEL. C11, C14, C15, C25, Q29, Q26

FLEXIBLE MIXTURE DISTRIBUTION MODELING OF DICHOTOMOUS CHOICE CONTINGENT VALUATION WITH HETEROGENITY

Abstract

This paper considers the performance a model of mixture normal distributions for dichotomous choice contingent valuation data, which allows the researcher to consider unobserved heterogeneity across the sample. The model is flexible and approaches a semi-parametric model, since any empirical distribution can be represented by augmenting the number of mixture distributions. Bayesian inference allows for simple estimation of the model and is particularly appropriate for conducting inference with finite data sets. The proposed model is compared with other semi-parametric and parametric approaches using Monte Carlo simulation, under alternative assumption regarding heteroschedasticity and heterogeneity in sample observations. It is found that the mixture normal model reduces bias and improves performance with respect to an alternative semi-parametric model, particularly when the sample is characterized by heterogeneous preferences.

Keywords. Bayesian inference, Contingent Valuation, Flexible distributions, Heterogeneity, Normal mixture, Willingness to pay.

1. Introduction

The valuation of environmental goods requires appropriate methods to measure the benefits society receives from their preservation. Contingent valuation (CV) is one of such methods which is basically a survey approach to the study of individual preferences. The question format can be a binary request for a yes/no response to a given price randomly chosen from a set of prices and randomly distributed across the sample. This is also called the dichotomous choice method (DC), which was earlier proposed by Bishop and Heberlein (1979) and further developed by Hanemann (1984) and Cameron (1988). This format has also been supported by Arrow *et al.* (1993) as a protocol to measure the damages to passive use values.

DC data can be statistically modeled by fitting a probability function and then integrating for the computation of the welfare measures represented by the mean and the median willingness to pay (WTP). Parametric models have the problem that they can fail to represent the empirical distribution leading to bias and inconsistent results (Yatchew and Griliches, 1984). Li and Mattson (1995) show that when the individuals are uncertain about their preferences, logit and probit models overestimate the variance of the error term. In addition, the welfare measures, particularly the mean, can be sensitive to the distribution assumption, as shown by Carson *et al.* (1994). These basic problems have led to the consideration of non-parametric approaches, such as Ayer´s (Kriström, 1991) and Turnbull's (Carson *et al.*, 1994) estimators. The former estimator is very simple from a computational point of view but fails to allow for the incorporation of covariates in explaining individual WTP. The latter involves complex computations when covariates are included in the model.

Semiparametric models allow for more flexible assumptions regarding the true distribution and permit the incorporation of explanatory variables besides the bid price. In general, these approaches decompose the response probability function into a cumulative distribution and an index or link function (Hanemann and Kanninen, 1999). For instance, Coslett's (1983) approach estimates the distribution function non-parametrically and then maximizes the likelihood function numerically for the link parameters. This estimator has been applied by Li (1996) showing that it reduces bias with respect to parametric approaches. Similar results are found by other approaches that assume a known form for the distribution function and a more flexible assumption for the link function. This is the case of An's (2000) model based on the assumption of a Weibull distribution and a flexible link function. Similarly, Chen and Randall (1997) and Creel and Loomis (1997) considered the assumption of a logistic distribution and a Fourier flexible form for the index function, and Cooper (2002) proposes methods for the calculation of bounded willingness to pay measures with these approaches.

Even though semiparametric models are more flexible than the parametric approaches and result in better performance against real data, they still require estimation of the error distribution, as commented by Lewbel and McFadden (1997), and some assumptions for either the cumulative distribution or the link function. In addition, they cannot provide exact inference for the predictive probability conditional on a particular observation of the covariates. On the other hand, Horowitz (1993) finds that misspecification errors in the cumulative distribution could be more relevant when the assumptions of homoschedasticity and unimodality are not appropriate for the data.

In this paper we consider the performance of a mixture normal distribution approach to modeling DC-CV data, similar to Geweke and Keane (1999). The mixture model allows us to represent any empirical data, while the normality assumption can be removed by increasing the number of distributions in the mixture. Mixture modeling provides an appropriate representation of individual heterogeneity by mixing on both the mean and variance parameters. The model is estimated by Bayesian techniques involving Gibbs sampling and data augmentation, which enables us to simulate the posterior distribution and conduct exact inference. The performance of the model is evaluated by Monte Carlo simulation against the seminonparametric approach by Creel and Loomis (1997). The results indicate that the mixture normal approach represents DC data more accurately than competing models.

The mixture model approach allows us to model unobserved heterogeneity as arising from the empirical data. The number of mixture distributions is endogenously determined and reflects preference variation across different groups of individuals in the sample. There can be individuals who favor the environmental good in question, while others might oppose it or be indifferent. In the literature on CV methods, there are several *ad hoc* procedures which have attempted to model sample heterogeneity. For instance, McFadden (1994) and Kriström (1997) model zero responses with a probability mass at zero, or “spike”, while Habb (1999) consider a participation model. Huttala (2000) also consider a controversial policy where there are different opinions in favor and against the policy.

These approaches have in common the inclusion of additional questions in the market construct regarding individual attitudes, which could lead to strategic or starting point biases in the responses. That is, if the discriminating question comes after the valuation question, it could be influenced by the price offered, whereas if it comes before the valuation question, the subject might find it easier to reject the valuation scenario, opting to be out of the market for any positive price. The approach proposed in this paper does not rely on additional questions to model unobserved heterogeneity. This arises from the modeling process by considering a mixture of distributions which accurately reflects the various patterns of preferences showed by the empirical data.

The paper is organized as follows. Section 2 presents the modeling approach based on Bayesian inference for mixture normal distributions. Section 3 compares the proposed model against alternative parametric and non-parametric models using Monte Carlo simulation. We consider alternative assumptions of the error distribution regarding heteroschedasticity and unobserved heterogeneity. Finally, section 4 summarizes the main conclusions.

2. Model of mixture normal distributions

Let us assume that the individual faces a change in a non-market or environmental good from q^0 to q^1 (with $q^1>q^0$). If the indirect utility function V after the change becomes higher then $V^1(M,z,q^1)+\varepsilon^1 > V^0(M,z,q^0)+\varepsilon^0$, where M is income and z is a vector of socioeconomic variables and prices, and ε^{i}, i=1,2 are error terms which reflect the change in unobserved non deterministic preferences.

If the subject faces price B_i for the change in q, the answer would be "yes" if $V_i^1(M_i - B_i, z_i, q^1) + \varepsilon_i^1 \geq V_i^0(M_i, z_i, q^0) + \varepsilon_i^0$, and no in the opposite case. Thus,

$$\Pr ob\{\text{"yes"}\} = \Pr ob\left\{V_i^1(M_i - B_i, z_i, q^1) + \varepsilon_i^1 \geq V_i^0(M_i, z_i, q^0) + \varepsilon_i^0\right\} \tag{1}$$

A monetary measure of the welfare change is given by the compensating variation, i.e.

$$V_i^0(M_i, z_i, q^0) + \varepsilon_i^0 = V_i^1(M_i - WTP_i, z_i, q^1) + \varepsilon_i^1 \tag{2}$$

where *WTP* is the maximum willingness to pay that the subject gives to the change in q, and can be written as a function $WTP_i(M_i, z_i, q^0, q^1, \varepsilon_i^0, \varepsilon_i^1)$. Let us denote the deterministic part $\Delta V_i(M_i, z_i, B_i, q^0, q^1) = V_i^1(M_i - B_i, z_i, q^1) - V_i^0(M_i, z_i, q^0)$ and the random part as $\varepsilon_i = \varepsilon_i^1 - \varepsilon_i^0$. The probability of an affirmative answer is given by:

$$\text{Prob}\{\text{"yes"}\} = \text{Prob}\{WTP_i(M_i, z_i, q^0, q^1, \varepsilon_i) > B_i\} = F_\varepsilon(\Delta V) \tag{3}$$

where F_ε is the cumulative distribution function. The error distribution can be specified as some parametric form, and the model estimated by maximum likelihood (Hanemann and Kanninen, 1999). However, flexible forms and non-parametric methods provide better representations of the empirical data, reducing bias (Creel and Loomis, 1997). A way of introducing flexibility in the distribution is to consider a mixture of m normal distributions. This assumption eliminates the normality assumption, approaching a semi-parametric model. By increasing the number of normal distributions in the mixture, any distribution can be approximated (Titterington et al., 1985). The stochastic terms adopt the following form:

$$\varepsilon_i = \sum_{j=1}^{m} e_{ij}(\alpha_j + h_j^{-1/2}\eta_i) \qquad (4)$$

where α_j and h_j represent the mean and the precision of each of the normal forms in the mixture[1], and $\alpha^{'} = (\alpha_1,...,\alpha_m) \in R^m; h^{'} = (h_1,...,h_m) \in R_+^m$ and $\eta_i | X \sim$ N(0, 1). The random vectors $e^{'} = (e_1,...,e_m)$ are i.i.d. with multinomial probability distribution of parameters $p_j = \text{Prob}(e_{ij} = 1)\ \ (\text{j} = 1,...,\text{m})$.

The welfare measures are given by the mean and the median. The mean is

$$E[WTP] = \int_0^{\infty} F_\varepsilon\left[\Delta V\left(M, z, B, q^0, q^1\right)\right] dB = \int_0^{\infty}\left\{\sum_{j=1}^{m} p_j \Phi\left[h_j^{1/2}\left(B_i - \alpha_j - \beta^{'} X_i\right)\right]\right\} dB \qquad (5)$$

where vector X contains the variables which influence utility (z, M, q^0, q^1), β is a k x 1vector of parameters and Φ(.) is the standard normal distribution. The median Me(WTP) is given by the following equality:

$$\int_0^{Me(EC)}\left\{\sum_{j=1}^{m} p_j \Phi\left[h_j^{1/2}\left(B_i - \alpha_j - \beta^{'} X_i\right)\right]\right\} dB = 0.5 \qquad (6)$$

The likelihood function across the sample under the mixture assumption is:

---

[1]The mean and the variance of the error terms are $E[\varepsilon_i] = \sum_{j=1}^{m} p_j \alpha_j$; $Var[\varepsilon_i] = \sum_{j=1}^{m} p_j^2 h_j^{-1}$ respectively.

$$p\left(y|\beta,\alpha,h,p,X\right)=\prod_{i=1}^{N} y_i \sum_{j=1}^{m} p_j \Phi\left[h_j^{1/2}\left(B_i-\alpha_j-\beta' X_i\right)\right]+$$

$$+(1-y_i)\left\{1-\sum_{j=1}^{m} p_j \Phi\left[h_j^{1/2}\left(B_i-\alpha_j-\beta' X_i\right)\right]\right\} \qquad (7)$$

Maximum likelihood estimators of the parameters can be obtained by accordingly optimizing this function using some non-linear method. Bateman *et al.* (2001) point out that dichotomous choice models require relatively larger samples than other elicitation methods to obtain comparable efficiency levels. In addition, asymptotic properties of ML estimators do not need to be maintained with small and finite samples[2]. Anderson and Richardson (1979) and Griffiths *et al.* (1987) found out relevant biases with numerical simulations of probit and logit models with small samples, while Copas (1988) utilize Taylor series expansions to define the bias obtained for a logit model with small samples.

Bayesian methods as developed by Chib (1992) and Albert and Chib (1993) are capable of providing exact inference with small samples. The prior distributions are defined as follows:

$$\beta \sim N\left(\underline{\beta}, \underline{H_\beta^{-1}}\right);$$

$$\alpha \sim N\left(\underline{\alpha}, \underline{H_\alpha^{-1}}\right)$$

$$\underline{s_j^2} h_j \sim \chi^2\left(\underline{\nu_j}\right)$$

$$p(h)=\prod_{j=1}^{m}\left[2^{\underline{\nu}_j/2}\Gamma\left(\underline{\nu}_j/2\right)\right]^{-1}\left(\underline{s}_j^2\right)^{\underline{\nu}_j/2} h_j^{(\underline{\nu}_j-2)/2}\exp\left(-0.5\underline{s}_j^2 h_j\right)$$

$$p(p) \sim \text{Beta(r)}$$

where $r = (r_1,...,r_m)$, $\alpha = (\alpha_1,...,\alpha_m)$ and H_α is an *m* x *m* matrix. The last two matrixes are positively defined. The posterior distribution is difficult to evaluate by conventional multiple integration methods. This problem can be bypassed by using a Gibbs sampling algorighm. This involves sampling from the conditional posterior distributions when they are known in a feasible form. Thus, even though WTP = (WTP_1, WTP_2, ..., WTP_n)' is not observed, it is possible its simulation from available information. Thus, given Y = (y_1, y_2, ..., y_n)' and θ = (β, α, h, p), the posterior distribution following data augmentation $\pi(\theta|Y,WTP)$ and the conditional density of the latent variable $f(WTP_i|Y,\theta)$ are known in a manageable form. The posterior distribution of θ is simulated by sampling from these conditional distributions. The latent variable is distributed as

$$WTP_i \sim N\left(x_i'\beta + \alpha' e_i, h_{L_i}^{-1}\right) \tag{8}$$

where L_i = (i:e_{ij}=1) and $P(L_i = j) \propto p_j \exp\left[-.5h_j^{1/2}\left(WTP_i - \alpha_j - x_i'\beta\right)^2\right]$. Thus, if we assume that the parameters are independent, the complete conditional density functions are as follows

$$f\left(WTP_i \middle| Y,\theta\right) = \begin{cases} \phi\left(WTP_i \middle| x_i'\beta + \alpha' e_i, h_{L_i}^{-1}\right) I\left[B_i,\infty\right] & if\, y_i = 1 \\ \phi\left(WTP_i \middle| x_i'\beta + \alpha' e_i, h_{L_i}^{-1}\right) I\left[-\infty, B_i\right] & if\, y_i = 0 \end{cases} \tag{9}$$

$$\pi\left(\beta \middle| Y, WTP, \alpha, h^{-1}, p\right) = \phi\left(\beta \middle| \overline{\beta}, \overline{H}_\beta^{-1}\right) \tag{10}$$

---

[2] See Amemiya (1985) or Huber (1981) for a more detailed discusión of the asymptotic proporties

$$\pi\left(\alpha \middle| Y, WTP, \beta, h^{-1}, p\right) = \phi\left(\alpha \middle| \bar{\alpha}, \bar{H}_{\alpha}^{-1}\right) \qquad \text{s.t. } \alpha_1 < \alpha_2 < \ldots < \alpha_m{}^3 \tag{11}$$

$$f\left(\bar{s}_j^2 h_j \middle| Y, WTP, \beta, \alpha, p\right) = f_{\chi^2}\left(\bar{s}_j^2 h_j \middle| \bar{\upsilon}_j\right) \tag{12}$$

$$f\left(p \middle| Y, WTP, \beta, \alpha, h^{-1}\right) = f_{Beta}\left(p \middle| r_1 + T_1, r_2 + T_2 \ldots, r_m + T_m\right) \tag{13}$$

where $\phi(.)$ I[a,b] is the truncated normal density function in [a,b], $T_j = \sum_{i=1}^{N} e_{ij}$ and f_{χ^2}, f_{Beta} represent the chi-squared and Beta density functions respectively. It can be shown that the posterior means are $\bar{\beta} = \bar{H}_{\beta}^{-1}\left(\bar{H}_{\beta}\underline{\beta} + \sum_{i=1}^{N} h_{L_i} x_i WTP_i\right)$ $\bar{\alpha} = \bar{H}_{\alpha}^{-1}\left[\bar{H}_{\alpha}\underline{\alpha} + \sum_{i=1}^{N} e_i\left(WTP_i - x_i^{'}\beta\right)\right]$ respectively. The variance-covariance matrixes are as follows: $\bar{H}_{\beta} = \left(\underline{H}_{\beta} + \sum_{i=1}^{N} h_{L_i} x_i x_i^{'}\right)$ and $\bar{H}_{\alpha} = \left(\underline{H}_{\alpha} + \sum_{i=1}^{N} e_i e_i^{'}\right)$. In addition, $\bar{s}_j^2 = \underline{s}_j^2 + \sum_i e_{ij}(WTP_i - \alpha_j - x_i^{'}\beta)^2$ y $\bar{\upsilon}_j = \underline{\upsilon}_j + T_j$.

Taking the starting value θ, $\theta^{(0)} = (\beta^{(0)}, \alpha^{(0)}, h^{(0)}, p^{(0)})$, the Gibbs sampling algorithm obtains iterated samples from each of the posterior conditional distributions. The algorithm is carried on *t* times leading to the simulated vector $(WTP^{(t)}, \beta^{(t)}, \alpha^{(t)}, h^{-1(t)}, p^{(t)})$ obtained from the joint distribution $(WTP, \beta, \alpha, h^{-1}, p)|Y$. These series of algorithms of size *t* are repeated over H times, leading to H values for each parameter which are simulated from the posterior distribution, i.e. $[WTP_h^{(t)}, \beta_h^{(t)}, \alpha_h^{(t)}, h_h^{-1(t)}, p_h^{(t)})]^{H}_{h=1}$. The moments of interest are obtained from these simulated values.

---

of ML estimators.

[3] This restriction guarantees the identification of the model (Geweke and Keane, 1999)

3. Monte Carlo experiment

The performance of the mixture distribution model can be compared with alternative parametric and semi-parametric models utilizing Monte Carlo simulation[4]. The objective is to measure the potential errors which can be obtained under alternative modeling approaches when there is misspecification in the model, i.e. when the estimated model does not conform with the model developing the empirical data. The crucial question is whether flexible approaches such as seminonparametric or mixture distribution models are capable of reducing these misspecification errors when the data are contaminated by heteroschedasticity or cross-sectional heterogeneity.

Let us assume that WTP for a change from q^0 to q^1 is as follows:

$$\mathrm{WTP}_i = x_i'\beta + \varepsilon_i\,; \qquad \text{with Var}\,(\mathrm{WTP}_i) = \sigma^2 \tag{14}$$

where variables in x_i and parameters $\{\beta, \sigma\}$are known from an empirical experiment. Consider that there are only two covariates as follows:

$$x_1 \sim iid\ N(0,1) \text{ and } x_2 \sim \chi^2_5 .$$

Other specifications for the covariates were also considered but did not produce relevant changes in the results. The true values of the parameters were assumed from empirical data as $\beta_1 = 1500, \beta_2 = 0.005 \text{ y } \sigma^2 = 1985$. Thus, the deterministic part of true WTP which responds to observed heterogeneity is known, and there is only need to specify the stochastic part i.e. the distribution F_{ε}, which can be influenced by unobserved heterogeneity across the sample.

Non-market valuation data can be generated under alternative assumptions. Let us consider six alternative distribution structures as specified in Table 1: i) homocedastic logistic ii) normal, ii) Normal or probit heterocedastic, iii) heterocedastic logit, iv) heterocedastic probit, v) weak preference heterogeneity (winners and indifferent) and vi) strong preference heterogeneity (winners, losers and indifferent). The latter two assumptions reflect different positions about the value of the environmental good, which can respond to different preferences or ethical beliefs.

In order to measure the sensitivity of the results to the sample size, the data development process involves generating 300 samples of sizes 150 and 800 for the true and estimated WTP according to the following steps:

i) Take a sample of size 1000 of the covariates x_i and generate WTP from equation (14) assuming the true parameters $\{\beta, \sigma\}$ and fit a normal distribution to these WTP data. A four bid vector design (B_1, B_2, B_3, B_4) is generated by calculating the percentiles of the inverse $F^{-1}(\varphi/\varphi+1)$, where F is the fitted normal distribution with the estimated mean and variance, and $\varphi=1,..,4$.

[4] The computer codes in GAUSS program utilized for estimation and simulations are available from

ii) Generate a sample for the explanatory variables x_i for i=1,…n_J (n_J =150, 800). These samples (one for each sample size) are used to fix the deterministic part $x_i'\beta$ in equation (14).

iii) Generate 300 samples of the true WTP of size n_J =150, 800 by randomizing from each of the specification of the random part in Table 1 and assuming $WTP_i = x_i'\beta + \varepsilon_i$.

iv) For each sample size n_J generate a random sample of the bid vector (B_1, B_2, B_3, B_4) following a multinomial distribution MN(n_J, 0.25, 0.25, 0.25, 0.25).

v) The binary choice responses (y_i) are generated such that a yes is obtained if $WTP_i < B_i$ and no in the opposite case.

vi) The binary choice data is utilized to estimate models under the alternative distribution structures and sample sizes utilizing maximum likelihood for the rigid parametric logit, and the alternative flexible models of the semi-parametric (SNPDF) approach by Creel and Loomis (1997)[5], and the Bayesian mixture distribution (BMNP) approach[6]. The number of mixtures m for the BMNP is determined by grid search from 1 to 10 and choosing the model with the lowest Bayesian information criteria (BIC).

---

the authors upon request.

[5] The truncation point was chosen according to the lowest mean squared error.

[6] The starting values for the Gibbs sampling simulations were taken from maximum likelihood and the number of iterations was 12000. The computation of the posterior moments involved a burn-in period of 1500 draws.

vii) For each sample and estimated model we compute the mean squared error (MSE) as the goodness of fit statistics. The bias resulting from the difference between the estimated and true mean WTP is also generated, and therefore the variance for each estimated model.

Tables 2 to 4 present the simulation results assuming alternative specifications of the error term in the true model generating the data. The best number of mixtures for the BMNP varied according to the type of simulated data. For homocedastic data this was m=1, while for heterocedastic and heterogeneous data the m parameter varied between 3 and 4 in most cases. The MSE, bias and variance are all expressed as a percentage of the true WTP. Table 2 focuses on the data generated from homocedastic Logit and Probit models. The misspecification error according to the MSE is not large if a logit model is utilized to fit data generated from a normal distribution with constant variance. Horowitz (1993) showed that these differences are not significant at the 95% level. For this reason, we do not present here the estimation results with the normal distribution.

The SNPDF model reduces MSE both in large and small samples, but these reductions are due to the decrease in the estimated variance when the samples are small, and to the reduction in bias when they are large. This is also the case for the BMNP but with a larger reduction in MSE, particularly for small samples. Thus, the BMNP is largely more exact with small samples than the SNPDF model, either when the data is generated from logit or from probit. The reduction in MSE for the BMNP model is about 33% with respect to the rigid structure, while it is only of 15% and 5% for true logit and probit respectively with the SNPDF model.

Table 3 shows the results under the true assumption of heterocedastic models. For the probit heterocedastic data we find that MSE raises for all estimated models above those obtained in Table 2 for the probit homocedastic data. These impacts on MSE are due to largest bias and become higher (40%) if we use the logit model to fit these data, while they stay much moderate for the alternative flexible models (10%) and with larger samples. It is interesting that with heterocedastic data there are not larger differences between the SNPDF and the BMNP models, since MSE are quite closed for all situations but for the small samples in the probit case.

In all flexible models with heterocedastic data, although MSE are rather close, the bias is always smaller with the BMNP model than with the SNPDF model. The logit model as applied to these data structure, although it shows up much larger MSEs than the flexible approaches, this is due to the largest variance. The bias is smaller, particularly with small samples. Thus, the BMNP model shows smaller bias than the SNPDF model, approaching the bias level obtained with the rigid logit model. On the other hand, when the flexible models are applied to the logit heterocedastic data, the results do not deviate as much from those obtained with the logit homocedastic data, with a slight improvement for the SNDPF model.

Table 4 presents the modeling results for the data generated under the assumption of weak and strong unobserved heterogeneity across the sample. In all models and with all sample sizes the MSEs rise above those obtained in previous results. This reflects the difficulty in modeling this type of data. The MSE is always larger under weak heterogeneity than under strong heterogeneity. This is clearly due to the larger bias and estimated variance obtained with the logit and SNPDF models applied to the former data sets. However, for the BMNP model the differences in MSE between both types of data are due to the larger variance under weak heterogeneity in the case of small samples and larger bias in the case of large samples.

The MSE obtained with the SNPDF is comparable to the logit model with small samples for both types of heterogeneity, with a small improvement with large samples. The BMNP model performs better than the competing models in both contexts of heterogeneity, both for small and large samples. It can be seen that these improvements in the MSE are more influenced by the reduction in biases with respect to the true WTP, although the variances are also smaller in three of the cases. The reduction in MSE is about 30% with respect to the SNPDF model in all situations, while the reduction in bias is even larger for small samples, but more moderate for large samples.

4. Conclusions

Dichotomous choice models have become a common approach to measure the value of non-market goods utilizing the CV method. The use of rigid parametric structures could lead to misspecification biases in welfare estimates. Flexible and seminonparametric approaches could be useful to check whether the data complies with rigid functional forms, and are capable of providing a more accurate representation of the empirical distribution given by sample responses. In this paper we have considered a Bayesian normal mixture distribution approach which approaches a semiparametric model as the number of distributions in the mixture is increased.

The mixture distribution model has the advantage of being able to represent unobserved heterogeneity by increasing the number of mixture distributions responding to different groups of individuals with different positions regarding the value of the environmental good. This also introduces flexibility in the modeling approach since it is capable of considering multimodal empirical distributions. In the limit, any distribution can be represented with an infinite number of mixtures.

The mixture model has been compared with the rigid logit model as well as with the seminonparametric approach by Creel and Loomis (1997). The results of the MC simulations for the alternative modeling approaches show that when the data is homocedastic both the SNPDF and the BMNP models are capable to improve performance against the logit model, with somewhat lower MSE for the BMNP model. This model overcomes SNPDF in small samples because the lower bias, while biases are comparable for large samples. However, the Bayesian mixture model does not show a great improvement against the competing modes with homocedastic data.

When the original data is characterized by heteroschedasticity, the simulation results show that the logit model presents the largest MSE because the large variance, while it is more accurate than the competing flexible models, which are more efficient. There are no large differences in MSE between the SNPDF and the BMNP. In general, the SNPDF slightly overcomes the BMNP for large samples, but the opposite is true for small samples. However, the BMNP model always produces lower bias, comparable to the rigid logit model in the case of large samples. In conclusion, the BMNP is more accurate than the SNPDF and more efficient than the logit model for heterocedastic data.

In the case of heterogeneous data the Bayesian mixture model overcomes their counterparts by a larger margin. This is the type of data which reflects a variety of preferences in the sample, with subjects who may be in favor of the policy option while others oppose to it. The results of the simulations show that the BMNP substantially reduces MSE against both logit and SMNP models, and this reduction is particularly due to the decrease in bias in the case of small samples and the fall in variance in the case of large samples. These results are valid for both types of heterogeneity (weak and strong), although the latter type of data leads to better performance of the models. The improvements in bias with the BMNP model are larger for small samples.

References

Albert, J.H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polichotomous Response Data". *Journal of American Statistical Association*, 88, pp. 669-679.

Amemiya, T. (1980). *Advances Econometrics*. Harvard University Press. Cambridge.

An, M.Y. (2000). "A Semiparametric Distribution for Willingness to Pay and Statistical Inference with Dichotomous Choice Contingent valution". *American Journal of Agriculture Economics*, 82, 487-500.

Anderson, E.B. and Richardson (1979). "Logistic Discrimination and Bias Correction in Maximun Likelihood Estimation". *Technometrics*, 21, 71-78.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. and Schuman, H. (1993). "Report of the National Oceanic and Athmospheric Administration Panel on Contingent Valuation". *Federal Register* 58, pp. 4602-4614.

Bateman, I.J., Langford, I.H., Jones, A.P. and Kerr, G.N. (2001). "Bound and Path Effects in Double and Triple Bounded Dichotomous Choice Contingent Valuation". *Resource and Energy Economics*, 23(3), 191-213

Bishop, R.C. and Heberlein, T.A. (1979). "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?. *American Journal of Agriculture Economics* 61:5, 926-930.

Cameron, T.A. (1988). "A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression". *Journal of Environmental and Management*, 15, pp 355-379.

Carson, R.T., Wilks, L. and Imber I. (1994). "Valuing Preservation of Australia's Kadaku Conservation Zone". *Oxford Economic Papers*, 46, 727-749.

Chen, H.Z. and Randall, A. (1997). "Semi-nonoparametric Estimation Of Binary Response Models with an Application to Natural Resource Valuation". *Journal Of Econometrics*, 76, pp-323-40.

Chib, S. (1992). "Bayes Inference in the Tobit Censored Regression Model". *Journal of Econometrics*, 51, pp. 79-99.

Copas, J.B. (1988). "Binary Regresión Models for Contaminated Data". *Journal of the Royal Statistical Society*, B, 50:2, 225-265.

Creel, M. and Loomis, J. (1997). "Semi-Nonparametric Distribution-Free Dichotomous Choice Contingent Valuation". *Journal of Environmental Economics and Management*, 32, 341-358.

Cooper, J.C. (2002). “Flexible Functional Form Estimation of Willingness to Pay Using Dichotomous Choice Data”. *Journal of Environmental Economics and Management*, 43, iss. 2, pp. 267-279

Cosslett, S.R. (1983). “Distribution Free Maximum Likelihood Estimator of he Binary Choice Model”. *Econometrica*, 51:3, 765-782.

Geweke, J. and Keane, M. (1999). “Mixture of Normals Probit Models” en C. Hsio, Lahiri, K, Lee, L.F. y Pesaran, M.H.(eds) *Analysis of Panels and Limited Dependent Variables: A Volume in Honor of G.S. Maddala*. Cambridge University Press. Cambridge.

Griffiths, W.H., Hill, R.C. and Pope, P.J. (1987). “Small Sample Properties of Probit Models Estimators”. *Journal of the American Statistical Association*, 82, pp 929-937.

Haab, T.C. (1999). “Nonparticipation or Misspecification?. The Impacts of Nonparticipation on Dichotomous Choice Contingent Valuation”. *Environmental and Resource Economics*, 14, 443-461.

Hanemann, M.W. (1984). “Welfare Evaluations in contingent valuation experiments with discrete responses”. *American Journal of Agricultural Economics*, 66. pp 103-118

Hanemann, M.W. and Kanninen B. (1999). “The Statistical Analysis of Discrete-Response CV Data”. Working Paper Nº 798. Department of Economics, Berkeley.

Horowitz, J.L. (1993). “Semiparametric and onparametric Estimation of Quantal Response Model” in *Hadbook of Statistics*, vol 11, G.S. Maddala, C.R. Rao and H.D. Vinod. Elsevier Science Publisher, B.V.

Huber, P.J. (1981). *Robust Statistics*. New York. John Wiley and Sons.

Huttala, A. (2000). “Binary Choice Valuation Studies with Heterogeneous Preferences Regarding the Program Being Valued”. *Environmental and Resource Economics*, 16, 263-279.

Kristöm, B. (1997). “Spike Models in Contingent Valuation”. *American Journal of Agriculture Economics*, 79, 1013-1023.

Kristöm, B. (1991). “A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies”. *Land Economics*, 66, 135-139.

Lewbel, A. and McFadden, D. (1997) “Estimating Features of a Distribution from Binomial Data”. Working paper. University of California. Berkeley.

Li, C.Z. (1996). “Semiparametric Estimation of the Binary Chice Model for Contingent Valuation”. *Land Economics*, 72 (4), pp. 462-473.

Li, C.Z. and Mattsson, L. (1995). “Discrete Choice under Preference Uncertainty: An Improved Model for Contingent Valuation”. *Journal of Environmental Economics and Management*, 28, 256-269.

McFadden, D, (1994). “Contingent valuation and Social Choice”. *American Journal of Agricultural Economics*, 76, November, 689-708.

Titterington, D.M., Smith, A.F.M and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.

Yatchew, A. and Griliches, Z. (1984). “Specification Error in Probit Models”. *Review of Economics and Statistics*, 66, 134-139.

Table 1. Specification of the error terms in true WTP models

MODEL	Distribution	Especification of ε_i
Model 1	Logistic (Logit)	ε_i ~ iid Logistic
Model 2	Normal (Probit)	$\varepsilon_i \sim$ iid $N(0,\sigma^2)$
Model 3	Logit Heterocedastic	$\varepsilon_i = v_i \times \left(1 + \left\lvert x_i'\beta \right\rvert\right)$, where v_i ~ iid Logistic
Model 4	Probit Heterocedastic	$\varepsilon_i = v_i \times \left(1 + \left\lvert x_i'\beta \right\rvert\right)$, where $v_i \sim$ iid $N(0,\sigma^2)$
Model 5	Weak heterogeneity (favour and indifferent)	$\varepsilon_i = p \times v_i$, where $v_i \sim$ iid $N(0,\sigma^2)$ and p=0.7
Model 6	Strong heterogeneity (favour, indifferent and against)	$\varepsilon_i = p_1 \times v_{1i} + p_2 \times v_{2i}$, where p_1=0.5 ; p_2= 0.2; $v_{1i} \sim$ iid $N(0,\sigma^2)$ and $v_{2i} \sim$ iid $N(-2\,x_i'\beta_1,\sigma^2)^{(*)}$;

(*) with β_1 = (1300, 0.003)

Table 2. Monte Carlo ressults assuming rigid structures for true F_{ε}

True models		Logit		Probit	
Estimated models		*n=150*	*n=800*	*n=150*	*n=800*
Logit	MSE	0.74	0.40	0.68	0.44
	Bias	0.19	-0.56	0.18	0.53
	Variance	0.70	0.08	0.65	0.16
SNPDF	MSE	0.62	0.48	0.65	0.42
	Bias	-0.39	-0.34	-0.43	-0.28
	Variance	0.46	0.36	0.47	0.35
BMNP	MSE	0.50	0.39	0.43	0.31
	Bias	-0.21	0.33	-0.27	0.25
	Variance	0.45	0.28	0.36	0.25

Table 3. Monte Carlo ressults assuming heteroschedascity for true F_ε

True models		Heteroc. Logit		Heteroc. Probit	
Estimated models		*n=150*	*n=800*	*n=150*	*n=800*
Logit	MSE	0.93	0.66	1.23	0.74
	Bias	0.14	-0.28	0.22	-0.22
	Variance	0.90	0.58	1.18	0.69
SNPDF	MSE	0.58	0.36	0.72	0.46
	Bias	0.38	0.42	-0.69	-0.43
	Variance	0.44	0.18	0.25	0.27
BMNP	MSE	0.50	0.39	0.53	0.48
	Bias	0.31	0.24	0.34	0.28
	Variance	0.40	0.32	0.41	0.40

Table 4. Monte Carlo ressults with heterogeneity for true F_{ε}

True models		Weak heterog.		Strong heterog.	
Estimated models		*n=150*	*n=800*	*n=150*	*n=800*
Logit	MSE	1.47	1.01	0.99	0.90
	Bias	0.60	0.50	-0.37	0.46
	Variance	1.11	0.76	0.85	0.68
SNPDF	MSE	1.41	0.87	0.95	0.84
	Bias	-0.62	-0.47	0.38	-0.32
	Variance	1.02	0.65	0.80	0.73
BMNP	MSE	1.15	0.67	0.65	0.51
	Bias	-0.23	0.41	-0.20	0.25
	Variance	1.09	0.50	0.61	0.45