ISSN 1748-7595 (Online)

Kent Business School

Working Paper Series

Dichotomous Choice Contingent Valuation with 'Dont Know' Responses and Misreporting

Iain Fraser
Kent Business School

Kelvin Balcombe
University of Reading

University of Kent

Working Paper No.151
October 2007

Dichotomous Choice Contingent Valuation with 'Dont Know' Responses and Misreporting.[1]

May 2007

Kelvin Balcombe[2]

and

Iain Fraser [3]

Address for Correspondence:

Dept of Agricultural and Food Economics

University of Reading

Earley Gate

P.O. Box 237

Reading RG6 6AR

UK

K.G.Balcombe@reading.ac.uk.

Phone: +44 (0)207-59-42623

[1]We would like to thank Peter Groothuis and Mihail Samnaliev for providing the data used in this paper. This data set provided by Peter Groothuis was originally produced by Chambers and Whitehead (2003).

[2]Reader in Applied Econometrcis, Deparment of Agricultural and Food Economics, University of Reading, UK.

[3]Reader in Environmental Economics, Applied Economics and Business Management Group, Kent Business School, University of Kent, UK.

Dichotomous Choice Contingent Valuation with 'Dont Know' Responses and Misreporting.

Summary

A new approach is presented that simultaneously deals with Misreporting and Don't Know (DK) responses within a Dichotomous Choice contingent valuation framework. Utilising a modification of the standard Bayesian Probit framework, a Gibbs with Metropolis-Hastings algorithm is used to estimate the posterior densities for the parameters of interest. Several model specifications are applied to two CV data sets. The first is on Wolf Management Plans. The second on the US Fee Demonstration Program. In contrast to other studies we find that DKs are more likely to be from people who would be predicted to have a positive utility for the bid. Therefore, a DK is more likely to be a YES than a NO. We also find evidence of misreporting, primarily in favour of the NO option. Finally, our willingness-to-pay estimates are both less than and greater than those previously reported which reflects the fact that inclusion of DK responses has no *a priori* impact on the key parameters of interest in this literature.

Key Words: Contingent Valuation, Don't Knows, Uncertainty, Misreporting, Bayesian Probit

JEL: C25, C11, Q51

1. Introduction

The expressed preference (EP) literature has long recognised that psychological and cognitive factors play a role in determining the values that are elicited from individuals, and may also lead to uncertainty by the individual about their own choices/preferences (e.g. Kahnmand and Tversky, 2000, Li and Mattson, 1995, Samnaliev et al. 2006). Likewise, research has also identified that people may either intentionally or unintentionally report preferences that would not be 'revealed' in a real situation i.e., misreport (e.g., Balcombe et al. 2007).[4] The practical consequences of uncertainty and/or misreporting are that the willingness to pay (WTP) estimates derived from EP studies may be subject to hypothetical bias and therefore highly inaccurate.

In this paper we develop a new econometric approach for the treatment of uncertainty and misreporting for dichotomous choice contingent valuation data. To date most research has been directed to resolving the issue of uncertainty, either via the inclusion of the Don't Know (DK) option or through the use of certainty scales that ask respondents to rate the certainty of their response. Our approach is to assume that uncertainty may be one reason for misreporting, although that misreporting may also occur for other reasons. It is also based on the view that the selection of DK represents a failure of the individual to recognise their own preferences and, in this sense, constitutes a form of misreporting. However, we entertain the possibility that uncertainty may not just lead to the selection of a DK response but may lead to the falsely reported acceptance or rejection. Some may find our terminology problematic since the word 'misreporting' may harbour connotations that such responses are somehow deliberately misleading. However, our use of the word

[4] The concept of misreporting is based on the idea of misclassification introduced by Hausman et al. (1998) and employed by Caudhill and Mixon (2005).

‘misreporting’ is not meant to imply that respondents necessarily intend to deceive, though this is another potential reason for misreporting and one which can equally well be dealt with within the structure we adopt.

As Samnaliev et al. (2006) argue, there is no precise definition of uncertainty, meaning that it can emerge for many reasons. For example, Li and Mattsson (1995) argued that survey respondents will have incomplete knowledge and this gives rise to preference uncertainty. Alternatively, Meyerhoff and Liebe (2006) examine protest responses and there is no reason to suggest that some of these manifest themselves as DKs. Thus, our approach simultaneously allows for a probability of misreporting on behalf of respondents as well as capturing uncertainty. That is, it embodies the notion that somebody can report one thing, when their utility suggests that they should report another. Our approach allows for probabilities of replying DK or misreporting to depend on the expected levels of utilities of individual respondents. Hence, our models can investigate whether respondents are more likely to answer DK if they are predicted to be a YES or NO, and whether YES’s are more likely to misreport NO and vice versa. The extreme case where DKs can legitimately be pooled with YESs or NOs, a model specification examined in this literature, can also be tested. Therefore, models without misreporting or where DKs are pooled emerge as special cases. In addition, our model allows the misreporting probabilities or the probabilities of reporting DK, to depend on the expected level of utility of the respondent. Overall, our approach is both general and more flexible than models currently in the literature.

The Probit models introduced in this paper can be estimated using either a Classical or Bayesian approach. However, due to the necessary constraints on an number of probability parameters, a Bayesian approach is advantageous. The Bayesian approach also allows for an approach to inference that allows non-nested models to be compared using marginal likelihoods and associated Bayes Factors . In this paper we describe and show how a Bayesian approach can be implemented which adds to a small but growing number of Bayesian applications in the contingent valuation literature (e.g., Arana and Leon, 2005 and Leon and Leon, 2003).

The paper is structured as follows. Section 2 briefly reviews the antecedent literature in this area. The next section develops the theoretical framework and assumptions we employ to develop our model. models we propose that taken account of uncertainty and misreporting. In Section 4 we present the modified Probit we employ in our analysis, describing the various models that emerge from the structure presented as well as our estimation methodology. The models developed in the paper are then used to analyse the Contingent Valuation (CV) data sets of Chambers and Whitehead (2003) and Samnaliev et al. (2006) in Section 5. Our analysis presents the empirical results along with a discussion of the key issues identified. Finally in Section 6 we conclude.

2. Antecedent Literature

2.1. Uncertainty

The appropriate treatment of DKs, and uncertainty more generally, has been extensively debated in the EP literature since the NOAA panel's recommendation to include a DK (or opt out) option (i.e., Arrow et al. 1993). Despite the importance of this issue many dichotomous choice contingent valuation studies assume that respondents are certain about their responses such that DK responses, if included, can be discarded without question. However, the situation is changing. Li and Mattson (1995) and Ready et al. (1995) were the first to address the question of the role of uncertain values in WTP studies and subsequent authors (e.g. Alberini et al. 2003, Carson et al. 1998, Chambers and Whitehead, 2003, Champ et al. 1997, Loomis and Ekstrand, 1998, Van Kooten, 2001, Wang, 1997) have catalogued, modified and extended these approaches. Within this literature uncertainty has been be rationalised by a number of different mechanisms (six are examined in Shaikh et al. (2007)) which is a reflection of the complex meaning of uncertainty.

In general, it has become common practice to assess levels of uncertainty using either an explicit DK option as part of the WTP question or via the use of certainty scales that require respondents to score the certainty of responses after the WTP question has been answered. In an effort to determine the relative merits of these two approaches there are now a number of papers that examine both approaches simultaneously (e.g., Samnaliev et al. 2006, Whitehead and Cherry, 2006).

The approach presented in this paper is most closely associated with the inclusion of the DK option and how to deal with these responses.[5] For example, Li and Mattson (1995) simply down-weight the responses of people that are uncertain, while other approaches make the restrictive assumption that uncertainty arises only because options have similar levels of utility, and that respondents are only able to make choices if utility thresholds are exceeded (e.g. Alberini et al. 2003, Wang 1997). This rationale gives rise to ordered logit or probit specifications such as that used by Groothuis and Whitehead (2002). Other authors such as Carson et al. (1998) and Chambers and Whitehead (2003) employ a multinomial Logit treatment of DK's. As in Hanener and Adamowicz (1998), Chambers and Whitehead investigate whether DK's are more like NO's than YES's. These results suggest that a DK is more likely to be a NO than a YES. However, while Carson et al. suggest that a DK can be taken as a NO, Chambers and Whitehead reject this hypothesis. Unfortunately, the use

[5] A useful summary and comparison of the various certainty scale methods employed in the literature is presented by Shaikh et al. (2007).

of the multinomial Logit (in this context) departs from the utility maximisation framework, whereby respondents answer YES or NO according to preferences characterised by a unique set of utility parameters, unless a restriction can be made across the parameter sets that effectively pool DKs with YES or NO responses. Therefore, if the responses cannot be pooled, then WTP estimates must effectively ignore the DK responses if the multinomial model is used.

The paper which is nearest to our research is that of Caudhill and Groothuis (2005). They explain that there are observed choices from the contingent valuation experiment and that there are also a number of unobservable choices that relate to the true meaning of a DK. That is a DK might be a YES, NO or DK. Using a logit specification this implies that there are five alternatives and they specify a likelihood function as an estimation problem in missing data. In addition, they also examine various constraints that pertain to the DK response that allow them to test if DK responses can be pooled with YES or NO responses. The method we propose is similar in spirit in that we also examine if DK is really a YES, NO or DK. However, we simultaneously allow for misreporting which significantly complicates the model we present.

Finally, Svensson (2006) provides an interesting study of how various efforts to deal with uncertainty associated with hypothetical bias can in turn give rise to other forms of bias. It is shown that the exclusion of uncertain responses in CV of WTP for traffic risk reduction biased results in favour of older respondents. These findings provide support for the use of all data collected and not just that part of the data that appears to meet researchers requirements regarding certainty of responses.

2.2. Misreporting

Misreporting in EP studies can happen in two or more ways. First, there may be differences between stated and actual preferences (or intentions), where the respondent is aware of the difference. This type misreporting can be inferred to exist from the misreporting of past behaviour by individuals (e.g., Granberg and Holmberg, 1991). Second, respondents may be imperfect predictors of their own behaviour/preferences and misreport their preferences for this reason (and will therefore arguably report DK).

Work on a statistical approach to account for misreporting within the EP literature has been conducted by Caudill and Mixon (2005) and Balcombe, et al. (2007). These research papers share a lineage with the misclassification approach employed by Hausman et al. (1998) using a logit specification.[6] They key finding in this research is that there is evidence of misreporting. For example, Caudhill and Mixon estimate that actual undergraduate cheating in exams is probably 20 percent higher than findings based on direct questioning of students. The method employed in this research simply takes the basic logit model and modifies the likelihood function directly to take account of the possibility of misreporting. Balcombe et al. employ a similar method, although the way in which they modify the likelihood function of the logit is different from that of Caudhill and Mixon who follow Hausman et al. Furthermore, to overcome some of the econometric difficulties, in particular model selection and identification, that emerge within a Classical context they employed a Bayesian methodology. In their EP study of consumer food choice and the use of pesticides Balcombe et al. found strong empirical support for misreporting. This existence of misreporting resulted in significant downward revision in WTP estimates, of almost 30 percent, for food produced without the use of pesticides.

[6] The use of the Probit in this paper is not the first time that the Probit has been employed to consider issues of misclassification. For example, Leece (2000) employed a Probit to examine issues associated with household choice of mortgage type.

3. The Theoretical Framework and Assumptions

Utility U_i is represented as:

$$U_i = \alpha' z_i - b_i + v_i \tag{1}$$

where b_i is the bid level of respondent $i(i = 1..n)$, $\alpha' z_i$ is the mean WTP for a person with attributes z_i (including an intercept) and v_i is a mean zero normal variate with precision (inverse variance) θ. An individual that is fully aware of their preferences will accept a bid if $U_i > 0$ and otherwise reject. The model can be scaled to have unit variance/precision by taking

$$\begin{aligned}
\sqrt{\theta} U_i &= \sqrt{\theta}\alpha' z_i - \sqrt{\theta} b_i + \sqrt{\theta} v_i \\
u_i &= \mu' z_i - \sqrt{\theta} b_i + e_i \\
&= x_i' \beta + e_i
\end{aligned} \tag{2}$$

Denoting $f_N\left(x|\mu, \theta^{-1}\right)$ as the normal density with mean μ and variance θ^{-1} the error in the normalised utility function is *iid* normal:

$$e_i \sim f_N\left(e_i|0, 1\right) \tag{3}$$

This assumption gives a probit model that is identified through the variance being 1, where the estimate of WTP can be recovered as $\frac{\mu' z_i}{\sqrt{\theta}}$. If b_i is the logged bid then this represents the logged WTP, in which case the WTP estimate is $exp(\frac{\mu' z_i}{\sqrt{\theta}})$. Under these assumptions it follows that

$$u_i \sim f_N\left(u_i|x_i'\beta, 1\right) \tag{4}$$

$$u_i \sim \phi\left(u_i - x_i'\beta\right)$$

where $\phi(.)$ denotes a standard normal density function.

Next we define

$$y_i = 0, 1, \bullet$$

where 0 denotes a rejection of the bid, 1 for and acceptance and $\bullet$ for a DN. We also define an indicator variable

$$\delta_i = 1 \text{ if } u_i > 0 \text{ and zero otherwise.}$$

The indicator variable simply tells us whether an individual has positive or negative utility at a given bid, independent of misreporting. Thus, in the absence of both misreporting and uncertainty δ_i=y_i. Unlike a standard probit model, a distinction is made between δ_i and y_i. The variable δ_i can only take two values (not three) and δ_i may or may not be equal to y_i where y_i is zero or 1. Therefore, people may make choices that diverge from their preferences.

We now develop our model by attaching probabilities to each of the events and these probabilities are (as we demonstrate) estimable. First we define the following (where $P(a|b)$ defines the probability of a given b) :

$$P(\delta|y, x_i) = \Psi_{\delta|y,i} \tag{5}$$

$$P(y|\delta, x_i) = \Theta_{y|\delta,i} \tag{6}$$

$$P\left(e_i < x_i'\beta\right) = \Phi_i \tag{7}$$

$$P(\delta|x_i) = \Phi_i^{\delta}(1-\Phi_i)^{1-\delta} \tag{8}$$

$$P(y|x_i) = \Lambda_{y,i} \tag{9}$$

The following conditions hold axiomatically:

$$\begin{aligned} \Lambda_{y,i} &= \sum_{\delta=0}^{1} \Theta_{y|\delta,i} \Phi_i^{\delta} \left(1-\Phi_i\right)^{1-\delta} \\ \Psi_{\delta|y,i} &= \frac{\Theta_{y|\delta,i} \Phi_i^{\delta} \left(1-\Phi_i\right)^{1-\delta}}{\sum_{\delta=0}^{1} \Theta_{y|\delta,i} \Phi_i^{\delta} \left(1-\Phi_i\right)^{1-\delta}} \\ \Theta_{y|\delta,i} &= \frac{\Psi_{\delta|y,i} \Lambda_{y,i}}{\Phi_i^{\delta} \left(1-\Phi_i\right)^{1-\delta}} \end{aligned} \tag{10}$$

These conditions on the key model parameters $\{\Lambda_{y,i}\}, \{\Psi_{\delta|y,i}\}$ and $\{\Theta_{y|\delta,i}\}$ allow us to generalise the Probit model so that we can examine uncertainty and misreporting. The structure of the model above share simililarities to a specification based on mixtures, where the population is divided into groups according to δ. However, unlike a mixtures approach, the underlying utility parameters are homogeneous accross the population.

To operationalise the model we need to make a number of assumptions. Furthermore, as we explain there are number of different ways in which we can proceed and this gives rise to a number of different model specifications.

4. Modified Probit: Alternative Model Structures

We can parameterise our model in a number of ways which yields distinctly different specifications. Four restrictions on the parameters $\{\Lambda_{y,i}\}, \{\Psi_{\delta|y,i}\}$ and $\{\Theta_{y|\delta,i}\}$ (apart from adding up restrictions and those in equation set (10)) are required. Importantly, not all of the parameters can be simultaneously fixed across i. The choice of which parameters to fix (as well as parameters to be estimated), is a question of model choice. For example, $\Theta_{1|0}, \Theta_{0|1}, \Theta_{\bullet|0}$,and $\Theta_{\bullet|1}$ can be specified as the parameters that require estimation (over and above β), with $\Theta_{1|1}$ and $\Theta_{0|0}$ being determined by adding up. However, this choice is by no means unique. $\Psi_{\delta|y}$ (i.e. $\Psi_{\delta|y,i} = \Psi_{\delta|y}$ for all i) might instead be chosen as parameters to be estimated. However, one cannot, in general, simultaneously fix parameters such as $\Theta_{1|1}$ and $\Psi_{1|1}$ since there is an implied relationship between the two that requires at least one must be a function of variables in the utility function. To aid understanding we now present four unique models that result from this characterisation. As we discuss, one of the approaches is impractical because of various statistical issues.

4.1. Model 1: The Fixed Probability of Misreporting

Model one is where:

- $\Theta_{1|0}, \Theta_{0|1}, \Theta_{\bullet|0}, \Theta_{\bullet|1}$ are fixed probabilities (and therefore $\Theta_{1|1}, \Theta_{0|0}$ by adding up)

This model is in the spirit of Hausman et al. (1998). Under these restrictions, the remainder of the parameters can be calculated as:

$$\begin{aligned} \Theta_{1|1} &= 1 - \Theta_{0|1} - \Theta_{\bullet|1} \\ \Theta_{0|0} &= 1 - \Theta_{0|0} - \Theta_{\bullet|0} \end{aligned} \tag{11}$$

$$\begin{aligned} \Lambda_{1,i} &= \Theta_{1|0}\left(1 - \Phi_i\right) + \Theta_{1|1}\Phi_i \\ \Lambda_{0,i} &= \Theta_{0|0}\left(1 - \Phi_i\right) + \Theta_{0|1}\Phi_i \\ \Lambda_{\bullet,i} &= \Theta_{\bullet|0}\left(1 - \Phi_i\right) + \Theta_{\bullet|1}\Phi_i \end{aligned} \tag{12}$$

$$\begin{aligned} \Psi_{1|1,i} &= \frac{\Theta_{1|1}\Phi_i}{\Theta_{1|0}\left(1 - \Phi_i\right) + \Theta_{1|1}\Phi_i}; \quad \Psi_{0|1,i} = 1 - \Psi_{1|1,i} \\ \Psi_{0|0,i} &= \frac{\Theta_{0|0}\left(1 - \Phi_i\right)}{\Theta_{0|0}\left(1 - \Phi_i\right) + \Theta_{0|1}\Phi_i}; \quad \Psi_{1|0,i} = 1 - \Psi_{0|0,i} \\ \Psi_{1|\bullet,i} &= \frac{\Theta_{\bullet|1}\Phi_i}{\Theta_{\bullet|0}\left(1 - \Phi_i\right) + \Theta_{\bullet|1}\Phi_i}; \quad \Psi_{0|\bullet,i} = 1 - \Psi_{1|\bullet,i} \end{aligned} \tag{13}$$

This model is quite tractable, since their need be no restrictions on the parameters $\Theta_{y|\delta}$ other than that they lie between zero and one.

4.2. Can other probabilities be treated as estimable parameters?

An alternative approach to treating misreporting and DK's is to assume that given a respondent has replied YES, NO or DK, then they have a constant probability of having either positive or negative utility, that depends only on what their response was. However, these restrictions do not fully define the model. A fourth restriction is required. Examples, of potential models of this sort are:

- $\Psi_{1|1}, \Psi_{1|0}$ and $\Psi_{1|\bullet}$ are fixed probabilities (and therefore $\Psi_{0|1}, \Psi_{0|0}$ and $\Psi_{0|\bullet}$ by adding up)
- $\Lambda_{\bullet,i}$ is either also fixed, or a known function of the data.

Alternatively, an identified model could be obtained by fixing

- $\Psi_{0|1}, \Psi_{1|0}, \Theta_{\bullet|0}, \Theta_{\bullet|1}$

Unfortunately, these approaches do not prove to be practical. While the remaining parameters can easily be calculated, feasible values for $\Psi_{\delta|y}$ can easily lead to the other parameters in the model being outside the unit interval unless they are highly constrained. For example, under either set of assumptions above:

$$\begin{aligned} \Lambda_{0,i} &= \frac{1}{\left(\Psi_{1|1} - \Psi_{1|0}\right)} \left[\Psi_{1|\bullet}\Lambda_{\bullet,i} - \Psi_{1|1}\left(\Lambda_{\bullet,i} - 1\right) - \Phi_i\right] \\ \Lambda_{1,i} &= \frac{1}{\left(\Psi_{0|0} - \Psi_{0|1}\right)} \left(\Psi_{0|\bullet}\Lambda_{\bullet,i} - \Psi_{0|0}\left(\Lambda_{\bullet,i} - 1\right) - \left(1 - \Phi_i\right)\right) \end{aligned} \tag{14}$$

Since $\Lambda_{0,i}$ or $\Lambda_{1,i}$ must be bounded between 0 and 1, then if Φ_i is either very small or very large for any individual in the sample, then $\Psi_{1|1}$ or $\Psi_{0|0}$ must be large in order for $\Lambda_{0,i}$ or $\Lambda_{1,i}$ to maintain their bounds. Thus, for any data set $\Psi_{1|1}$ and $\Psi_{0|0}$ need to depend on Φ_i such that $\Psi_{1|1} \to 1$ as $\Phi_i \to 1$ and $\Psi_{0|0} \to 1$ as $\Phi_i \to 0$. A similar problem arises if the parameter set becomes $\left\{\Psi_{0|1}, \Psi_{1|0}, \Theta_{\bullet|1}, \Theta_{\bullet|0}\right\}$. In short fixing $\Psi_{0|1}, \Psi_{1|0}$ and $\Psi_{1|1}$ is an impractical option. It might be possible to functionalise $\Psi_{1|1,i}$ in a different way from equation (13). However, we propose a more straight forward approach below.

4.3. Model 2: Variable Probabilities for Don't Knows.

Should the $\Theta_{y|\delta,i}$ be constants? Arguably, the probability that a respondent with positive utility for a bid replies DK is likely to increase as the bid increases. That is, if a respondent is expected to have a very large positive utility, they may also be more aware that their preference is for the affirmative. This type of reasoning gives rise to alternative model specifications, the simplest being as follows

$$\begin{aligned} \Theta_{\bullet|1,i} &= \Theta^*_{\bullet|1}\,(1-\Phi_i)^{\rho} \\ \Theta_{\bullet|0,i} &= \Theta^*_{\bullet|0}\Phi_i^{\rho} \end{aligned} \tag{15}$$

where ρ is an additional parameter. In order for the probabilities $\Theta_{\bullet|1,i}$ and $\Theta_{\bullet|0,i}$ to be globally bounded on the unit interval, $\Theta^*_{\bullet|1}$ and $\Theta^*_{\bullet|0,i}$ also need to be on the unit interval and $\rho > 0$. This extension creates no significant problems for estimation, although the nature of the sampler requires an additional step (for β) and the estimation of an additional parameter.

4.4. Model 3: Variable Probabilities of Misreporting

A similar type of adjustment as in equation (15) can also be made for the misreporting parameters $\Theta_{0|1,i}$ and $\Theta_{1|0,i}$. Arguably, the probability of falsely reporting NO given a positive utility might decrease as the expected level of utility increases. This argument is perhaps less compelling than for the case of the DKs. If misreporting is deliberate, then there is no reason to believe that the probability that the person will report falsely will change with their expected utility. However, if misreporting is due to uncertainty on the part of the respondent about their own preferences, then $\Theta_{0|1,i}$ and $\Theta_{1|0,i}$ are likely to depend on Φ_i. For example, somebody is more likely to misreport that they would accept a bid, when they have negative utility, the higher their (negative) level of utility u_i. This supports a third specification where, in addition to the conditions in equation (15) above:

$$\Theta_{0|1,i} = \Theta^*_{0|1}(1-\Phi_i)^{\rho} \quad (16)$$

$$\Theta_{1|0,i} = \Theta^*_{1|0}\Phi_i^{\rho}$$

A further expansion of the model would be to differentiate between ρ across the different probability parameters. However, these extensions are not investigated here. Importantly, as the parameter ρ converges to zero for both Models 2 and 3, the constant probability model (Model 1) is supported, suggesting that individuals' misreporting or DK responses are independent of their level of utility. Such a circumstance would occur if people were ignoring the bid information.

Finally, other models/hypotheses are of interest. In particular, the hypothesis that $\Theta_{0|1,i} = \Theta_{1|0,i} = 0$. If this hypothesis holds then there is no support for misreporting in the model. Also, we will be interested in whether DKs should be treated as YES or NO (i.e., pooling of DKs), and finally, the difference between results for those models that incorporate DKs and a model which simply does not use DKs. How we examine these various model specification is explained in Section 4.6.

4.5. Estimation Methodology

The parameters to be estimated are, in addition to β for each of the models 1,2 and 3 are:

$$\Omega_1 = \left\{\Theta_{1|0}, \Theta_{0|1}, \Theta_{\bullet|0}, \Theta_{\bullet|1}\right\} \quad (17)$$

$$\Omega_2 = \left\{\Theta_{1|0}, \Theta_{0|1}, \Theta^*_{\bullet|0}, \Theta^*_{\bullet|1}, \rho\right\} \quad (18)$$

$$\Omega_3 = \left\{\Theta^*_{1|0}, \Theta^*_{0|1}, \Theta^*_{\bullet|0}, \Theta^*_{\bullet|1}, \rho\right\} \quad (19)$$

Each of the three models can be estimated using a Bayesian or Classical approach. The classical approach would maximise the log-likelihood

$$\ln L = \sum_{y_i=0} \ln\left(\Lambda_{0,i}\right) + \sum_{y_i=1} \ln\left(\Lambda_{1,i}\right) + \sum_{y_i=\bullet} \ln\left(\Lambda_{\bullet,i}\right) \tag{20}$$

with respect the to the parameters in each model using the expressions in the previous section.

As we have already stated a Bayesian approach is employed here. A note on the priors that we use and the Bayesian algorithm we employ is presented in detail in the appendix to this paper. The algorithm is a Gibbs with Metropolis-Hastings algorithm. Only slight differences are made to the algorithm to estimate each of the models.

In general the Bayesian approach to estimation uses a latent variable approach. There are advantages in employing Bayesian methods as opposed to Classical in that all of the parameters in the model (with the exception of β) require inequality constraints, since they represent probabilities and are therefore between 0 and 1, and this is easily accommodated. The Bayesian approach to estimation is highly intuitive in the current context, since in order to model misreporting and/or DKs the latent variables that are generated within the algorithm can be derived using the probability calculations in the equation set (10). In addition, by introducing a rejection step within the algorithm we place an inequality restriction on the bid coefficient as it must be negative.

4.6 Hypothesis Testing

In order to evaluate the performance of the models the marginal likelihoods for each model were calculated. The marginal likelihood for a model M with parameters P is :

$$\int L_M\left(P\right) f_M\left(P\right) dP$$

where $L_M\left(\right)$ is the likelihood for model M and $f_M\left(P\right)$ are the priors. The ratio of two marginal likelihoods is referred to as a Bayes Factor. A Bayes factor represents the posterior odds in favour of the numerated model if the prior odds of each model are equal. The marginal likelihoods for each model are estimated using the Gelfand and Dey method (see Gelfand and Dey, (1994) or Koop (2003) pp.104-106, for details). The marginal likelihood for each of the models can be compared, with the best performing model being the one with the largest likelihood. This enables Bayesian hypothesis testing for the hypothesis that $\rho = 0$. We can also test hypothesis such as (in the case of model 1) $\Theta_{1|0} = \Theta_{0|1} = 0$, in which case there is no misreporting in the model.

4.6.1. DK meaning YES or NO?

Previous authors have tested for the "pooling hypothesis", that DK means NO or, alternatively, YES. In the multinomial Logit this takes the form of restricting across parameter sets (e.g., Caudhill and Groothuis, (2005)). In the context of the current model, the "pooling hypothesis" takes a different form. The concept that DK means YES (or NO) is not entirely unambiguous once misreporting exists. The best interpretation would be that:

$$\Psi_{1|\bullet,i} = \Psi_{1|1,i} \tag{21}$$

That is, the chance that a respondent has positive utility for a bid is the same for somebody that answers YES or DK. This is quite different from the hypothesis that people never report DK if it is NO. Taking again the case of Model 1:

$$\begin{aligned} \Theta_{\bullet|0} &= 0 \\ &\Rightarrow \Theta_{1|0} + \Theta_{0|0} = 1 \\ &\Rightarrow \Psi_{0|\bullet,i} = 0 \text{ and } \Psi_{1|\bullet,i} = 1 \end{aligned} \tag{22}$$

It would somewhat odd to impose this condition if $\Psi_{1|1,i} < 1$, whereby someone who replied DK is more likely to have positive utility for the bid than somebody who replied

YES! Under the constant probability case (Model 1), the hypothesis that $\Psi_{1|\bullet,i} = \Psi_{1|1,i}$ implies that:

$$\frac{\Theta_{1|1}\Phi_i}{\Theta_{1|0}(1-\Phi_i)+\Theta_{1|1}\Phi_i} = \frac{\Theta_{\bullet|1}\Phi_i}{\Theta_{\bullet|0}(1-\Phi_i)+\Theta_{\bullet|1}\Phi_i}$$

$$\Rightarrow \Theta_{1|1}\Theta_{\bullet|0} = \Theta_{\bullet|1}\Theta_{1|0} \quad (23)$$

In the absence of misreporting, this would imply that $\Theta_{\bullet|0} = 0$. However, with misreporting, clearly what is required for the last ratio to be positive to be preserved. Thus we have the two hypotheses:

DK as a YES

$$\Theta_{1|1}\Theta_{\bullet|0} = \Theta_{\bullet|1}\Theta_{1|0} \quad (24)$$

DK as a NO

$$\Theta_{0|0}\Theta_{\bullet|1} = \Theta_{\bullet|0}\Theta_{0|1} \quad (25)$$

In the case where the parameters $\Theta_{y|\delta}$ are not constants (as in Models 2 and 3), these restrictions can be imposed. However, it leads to a certain arbitrariness in the model, since one of the parameters $\Theta_{\bullet|0,i}$ or $\Theta_{\bullet|1,i}$ needs to be solved in terms of the others, thus they cannot both be set as in Model 2. For this reason, we only consider this test in the case of Model 1. These conditions can be imposed on the model and tested by using the marginal likelihood in the standard way. These calculations have been performed for all models estimated and the results are reported in the next section. Finally, although a testable hypothesis, it will be useful to consider the difference (in terms of estimated WTP) between models that incorporate DKs, and those that simply eliminate them from the sample. All these submodels are also considered in the next section.

4.7. Estimating WTPs

The WTP estimates can be calculated as described in Section 2, where $exp(\frac{\mu' z_i}{\sqrt{\theta}})$ can be used to compute a WTP estimate for an individual if the bid option is logged or $\frac{\mu' z_i}{\sqrt{\theta}}$ otherwise. For the sample as a whole, it is more informative to map the distribution of $exp\left(\frac{\mu' \bar{z}}{\sqrt{\theta}}\right)$ or $\frac{\mu' \bar{z}}{\sqrt{\theta}}$ where $\bar{z}$ is the sample mean. With no loss of generality, all elements of z_i other than the intercept can have a zero mean (by subtracting their means). In this case, the WTP estimate at the mean point is simply $exp(\frac{\mu_0}{\sqrt{\theta}})$ or $\frac{\mu_0}{\sqrt{\theta}}$ where μ_0 is the intercept. The denominator must be constrained to be positive unless researchers are prepared to admit that respondents have a negative utility for money. However, even very small values for $\sqrt{\theta}$ will lead to very large estimates of the WTP. Therefore, it may be sensible to place a rejection step within the sampler to impose the condition that $exp\left(\frac{\mu_0}{\sqrt{\theta}}\right)$ or $\frac{\mu_0}{\sqrt{\theta}}$ is less than the maximum bid within the sample. This has little or no effect on the resulting posterior median WTP, but a very large impact on the mean of the posterior. The posterior distribution of this quantity can be calculated using the posterior distributions for μ and $\sqrt{\theta}$ produced by the sampler. Also, providing the same constraints are put on all models, then by constraining $exp\left(\frac{\mu_0}{\sqrt{\theta}}\right)$ or $\frac{\mu_0}{\sqrt{\theta}}$ to be less than some value, creates no problems for the calculation of the marginal likelihoods.[7]

5. Model Applications

5.1. Data and Model Specification

In this Section we employ two data sets. The first is Chambers and Whitehead (2003) who estimated the WTP for a wolf management plan in Minnesota. This data set was also analysed by Caudhill and Groothuis (2005) in an analysis of DK responses. The

[7] Constraints on the magnitude of the WTP do create problems for testing the hypothesis that the either logged price or unlogged price should enter the model. This form of modification means that the integrating constant for the prior will change. For this reason we do not consider this issue in this paper.

results they present do not include all the explanatory variables used by Chambers and Whitehead and they pool the data from the two areas in which the survey was collected, St. Cloud and Ely. Chambers and Whitehead use the CV method where respondents are asked questions about their attitudes and "non use motives" in addition to age, sex etc. Importantly, they also include a DK response option in the CV and a large number of respondents answered DK. Chambers and Whitehead use a multinomial logit model to differentiate between DK and NO responses. Among other issues, they explore whether the DKs are equivalent to NOs as suggested by Carson et al. (1998). They reject this hypothesis. However, their results suggest that DKs are more similar to NOs than YESs.

The second data set we use was produced by Samnaliev et al. (2006). In this paper the authors examined WTP for access to public land as part of the US Fee Demonstration Program (FDP). Two versions of the survey were employed, one of which was a dichotomous choice CV with DK option. In our analysis we estimate almost the same WTP function, the only difference being that we did not have the Round variable which in the analysis presented by Samnaliev et al. was found to be statistically insignificant. Using a logit, Samnaliev et al. estimated a number models in which the DKs were dropped from the analysis, assumed to be YES responses and assumed to be NO responses. They found that when DK are dropped the resulting WTP is less than when they assumed YES and WTP is greater than when DK is assumed NO.

The results presented in this section used the algorithms described in detail in the appendix. These algorithms were tested using Monte Carlo data and we established that they accurately identified the various data generating processes examined. For both sets of data, the burn in phase was set to 2,000 iterations, followed by another 100,000 iterations in which every 10th observation was sampled (so as to decrease the dependence in the sequence). Convergence was monitored using visual plots of the sequences of values produced by the sampler and by modified t-tests for each of the parameters (allowing for the dependence in the series).

5.2. Marginal Likelihoods and Model Selection

We begin by examining the three model specifications outlined in Section 4. For both data sets and all models the marginal likelihoods are presented in Table 1.

{Approximate Position of Table 1}

The results in Table 1 clearly indicate that the models with constant probabilities for misreporting and DKs (Model 1) are preferred to all other model specifications. For example, the St. Cloud data Model 1 has a Bayes Factor relative to Model 2 of approximately $exp(191.56 - 189.5)$=7.8. You interpret this as indicating that Model 1 has posterior odds of nearly eight to one over Model 2 (where prior and posterior odds are equal). For the Ely data, Model 3 is the least supported, with posterior odds of nearly six to one. We also find the same result the FDP data although the difference between Model 1 and Model 3 is small. Importantly, for all data sets the marginal likelihoods support the use of the misreporting model over the model without misreporting $\Theta_{1|0} = \Theta_{0|1} = 0$. If this restriction is imposed then all data sets see large reductions in their marginal likelihoods relative to all three other models that incorporate misreporting. Notably, the fall in the marginal likelihood for the St.Cloud is larger than for the Ely and FDP data, suggesting that for some reason their is a greater degree of misreporting for this data set. With regard to the pooling tests, it can be seen that pooling DKs with either YESs or NOs in all data set reduces the marginal likelihoods, although pooling with the YESs yields a relative lower marginal likelihood for all data sets. Overall, we would reject the pooling of DKs with other responses. In summary, these results imply that while DKs should not be treated as YESs or NOs, they are more like a YES than a NO. This finding is in contrast with the previous findings of Chambers and Whitehead (2003) and Caudhill and Groothius (2005). Given these findings in our remaining analysis, we restrict our discussion to the results generated by Model 1 since it was the best performing model.

At this point it is also worth commenting on the estimates of the parameter ρ in Models 2 and 3, which were close to zero (around 0.01) in for both data sets. An examination of the posterior distribution for all of the ρ revealed it to be densely packed near 0 with the tail of the distribution around 0.05. The posterior distributions of ρ are consistent with the results for the marginal likelihoods which imply that respondents propensity to misreport or report DK is not related to the level of their expected utility. Our interpretation of this finding is that respondent's misreporting or reporting of DK is, therefore, not due to being indifferent between options and making a mistake. These results are more consistent with respondents making arbitrary/or predetermined choices that do not depend on the bid level. This finding tends to support the view that some respondents are not making informed choices as is required if the CV is to be meaningful.

5.3. WTP Function Parameter Estimates

The parameter estimates for the preferred models are presented in Table 2 Ely and St.Cloud, and for FPD in Table 3. The results in Table 2, correspond to the results in Table III of Chambers and Whitehead (2003) and the results in Table 3 with Table 3 in Samnaliev et al. (2006). In our analysis a $y = 1$ means YES, such that the signs of the coefficients will be roughly opposite to the NO results in Chambers and Whitehead. There is no reason to expect that the coefficient magnitudes should be the same for any of the data sets given that we employed a Probit framework rather a Logit. Moreover, in a

Bayesian analysis, standard deviations are not equivalent to standard errors in a Classical analysis.

{Approximate Position of Tables 2 and 3}

In general there is a strong correspondence between our results (i.e., parameter signs) and those reported in the earlier research. There are some differences regarding the statistical robustness. For example, for the ELY data we find that a minority of the variables have means that are larger in absolute value than the standard deviations. This holds for INCOME, ETHICAL, EDUCATION, AGE, and GENDER. Chambers and Whitehead (2003), in contrast, find that INCOME and GENDER are significant. For the St. Cloud data, exactly half of the variables have absolute posterior means larger than their standard deviations which is in keeping with Chambers and Whitehead. The FDP results in Table 3 our results are equivalent to Samnaliev et al. (2006) in terms of signs and in most cases similar in terms of statistical importance. We are not able to include the 'round' variable that is included in Samnaliev et al.since this was not included in the data set given to us. However, since this variable was insignificant in the analysis of Samnaliev et al. we would not expect it to have a substantive impact on our results.

Next we examine the misreporting parameters ($\Theta_{y|\delta}$) in Tables 2 and 3. In Table 2 we see that for $\Theta_{\bullet|1}$,which indicates DKs reported when a YES should be reported (i.e., have positive utility for the bid) that 40% and 25% of respondents report DK when they have a positive utility for the bid. In Table 3 for the FDP data we find 13%. This contrasts with only 13% and 16% In Table 2 and 10% in Table 3 of respondents reporting DK when they have a negative utility for the bid.

With regard to the other misreporting parameters, for St. Cloud, misreporting seems to be fairly evenly split, with around 13% to 16% of respondents reporting NO when they mean YES or YES when they mean NO. For the ELY data, there is more significant evidence that respondents with a positive utility for the bid often reply NO. 23% of respondents are expected to respond in this way, with only 5% reporting NO when they mean YES. Finally, for the FDP data we find quite high levels of misreporting at 37% saying NO when they mean YES and 33% the other way.

To explain the misreporting results in a different way we also present posterior densities for each of the data sets in Figures 1 (St Cloud), 2 (Ely) and 3 (FDP).

{Approximate Position of Figures 1, 2 and 3}

For example, in the top half of Figure 1 we can see the "NO when YES" ($\Theta_{0|1}$) density being packed towards zero, but the "YES when NO" ($\Theta_{1|0}$) having a symmetric distribution away from zero. [8] Overall Figures 1, 2 and 3 are consistent with the marginal likelihood results in Table 1. They confirm that a DK is more like a YES than a NO.

5.4. WTP Estimates

The final part of our analysis is a comparison of the WTP estimates for our preferred models and those reported in the original research. These results are summarised in Table 4.

{Approximate Posistion of Table 4}

Beginning with Chambers and Whitehead (2003) we can see that compared with a standard Bayesian Probit where the DKs have been eliminated from the sample and no misreporting is assumed there is little difference in the mean for St Cloud and a small

[8] We also estimate the models using the unlogged bid levels. The substantive findings regarding misreporting were unchanged.

mean for Ely. However, when we examine Model 1, our preferred model specification for both sets of data we see some more significant differences. For Model 1 we obtain mean/median WTPs of over $40 for St Cloud and around $4 median and $8 mean for ELY. The large difference between the mean and median for ELY are due to the highly skewed posterior which is illustrated in Figure 4 showing the distribution for the WTP in ELY. The increase in the estimated WTP for St Cloud region relative to the standard Probit is due to the fact that misreporting and DKs have largely been identified by the model as YESs. Since the model integrates this information into its estimation procedure, this increases the resulting WTP. If the mean WTPs are used, then the end result is a near doubling of the estimated WTPs for both regions relative to the findings of Chambers and Whitehead. However, if the medians are used, an increased estimate is only found for St. Cloud. In our opinion the median rather than mean estimates should be used due to the potentially volatile nature of the mean estimate when the bid coefficient can be close to zero. It is also the case that by taking account of the uncertainty in the data in this way the resulting standard deviations associated with the WTPs for Model 1 are quite large indicating that we need to treat our point estimates with a certainty degree of caution. Turning to the Samnaliev et al. (2006) data we find much less difference in the WTP results produced by each method. This greater degree of conformity is reflected in the WTP posterior distribution in Figure 4.

{Approximate Posiution of Figure 4}

6. Conclusion

In this paper we have introduced a new framework that simultaneously deals with misreporting and DK responses within a dichotomous choice CV framework. A Bayesian approach to estimation using a Gibbs with Metropolis-Hastings algorithm to estimate the posterior densities for the parameters of these models has been introduced and developed. The various models developed have been applied to two CV data set that have been used to publish research papers in the literature.

In accordance with previous studies in the literature we found strong evidence that respondents might reply YES when the may mean NO and vice versa. In two out of the three data sets we found evidence of misreporting, primarily in favour of the NO option and the resulting WTP estimates were substantively different in some cases to those previous reported. We also rejected the hypothesis that DKs could be pooled with YESs or NOs. However, in contrast to both papers and results previously presented (and others in the literature) we find that DKs are more likely to be from people who would be predicted to have a positive utility for the bid. Therefore, a DK, in this model structure, is more similar to a YES than a NO. This result is not without precedent in the literature. Indeed, Shaikh et al. (2007) present empirical evidence that there is no systematic reason to assume that including uncertainty within the analysis leads automatically to a reduction in WTP.

Interestingly we have found no evidence in favour of the hypothesis that DK responses or misreporting are a function of the expected utility level of participants. This gives rise to the conclusion that those reporting DK are not necessarily doing so because they are close to being both a YES and NO (zero utility). Clearly, these findings are data set specific and they need not apply to other data sets. Indeed, we expect that in many circumstances 'warm glow' effects within EP studies are likely to work in a reverse fashion to those found in this paper, though this supposition requires further research.

Finally, the procedures outlined here can, in principle, be applied in related contexts. For example, other forms of CV and Choice Experiments. In CV studies it is becoming quite common to use scales of uncertainty rather than a simply DK. With this type of data , probabilities of replying with different levels of uncertainty, given positive or negative levels of utility could be estimated using an extension of the models herein. It is also possible that these probabilities could be conditioned on the attitudinal variables. In the case of Choice Experiments it may well be possible to include an opt out or DK option using a.similar algorithm to that developed in this paper to investigate uncertainty.

Appendix

A1:Priors

Bayesian estimation and inference requires priors to be specified for the parameters. Inference using the marginal likelihood is priors requires that the priors are proper (integrate to 1). Our parameters are composed of β (the coefficient in the Probit) and Ω, which has three variants specified in Section 4 of the paper. The priors for β $(f(\beta))$ are normal with mean zero and variance V_0. The variance is specified as $(10^3 I)$ for the results in the paper. For the parmeters in Ω $f(\Omega)$ for Models 1,2 or 3 (and the subcases) the priors are set as Beta$(a_1,\ a_2)$. For the results in the paper the Beta priors Beta(1,2) for giving a slight penalty for larger values as they tend towards one. This is consistent with a prior belief that people do not misreport.

A2:The Algorithm (all quantities are as defined in the text)

Given a starting set of values

- 1. $\{\delta_i\}$ **draw:** Draw the indicator variables $\{\delta_i\}$ according using the probabilities $\left\{\Psi_{\delta|y,i}\right\}$
- 2. $\{u_i\}$ **draw:** Draw the latent variables $\{u_i\}$ from the truncated normal distribution with a mean $x_i'\beta$ and a unit variance
- 3 β **draw:**
  - 3.1. Draw the β^{prop} from a normal with mean $\left(V_0^{-1} + \sum x_i x_i'\right)^{-1} \sum x_i u_i$ and a variance $\left(V_0^{-1} + \sum x_i x_i'\right)^{-1}$ (where V_0^{-1} is set a priori)
  - 3.2.
  - For model 1, accept β^{prop} with probability one.

– For models *m=2 and 3* accept β^{prop} with probability (where $I\left(\left\{\Theta^{prop}_{y_i|\delta_i,i}\right\} \in (0,1)\right)$ denotes an indicator variable)

$$\alpha = \min\left(1, \frac{f\left(\Omega^{prop}_m\right)\prod_{i=1}^{n}\Theta^{prop}_{y_i|\delta_i,i}}{f\left(\Omega_m\right)\prod_{i=1}^{n}\Theta_{y_i|\delta_i,i}}\right) \times I\left(\left\{\Theta^{prop}_{y_i|\delta_i,i}\right\} \in (0,1)\right) \tag{26}$$

or else stick with old β.

- 4. Ω_m **draw:** Generate new parameters

$$\Omega^{prop}_m = \Omega_m + v \tag{27}$$

where v is a innovation with a symmetric distribution and accept with probability

$$\alpha = \min\left(1, \frac{f\left(\Omega^{prop}_m\right)\prod_{i=1}^{n}\Theta^{prop}_{y_i|\delta_i,i}}{f\left(\Omega_m\right)\prod_{i=1}^{n}\Theta_{y_i|\delta_i,i}}\right) I\left(\left\{\Theta^{prop}_{y_i|\delta_i,i}\right\} \in (0,1)\right) \tag{28}$$

else, stick with old Ω_m.(record Ω_m)

- Return to step 1.

At each iteration the parameters are recorded and are then used to map the posterior distributions as in Section 5 of the paper.

A3: Deriving the Posterior Distributions

The Posterior for the parameters and Latent Data where Y denotes all the data and $f(\beta)$ and $f(\Omega)$ (we do not notationally distinguish the different models below by subscripting Ω) are independent priors:

$$f\left(\{u_i\}, \{\delta_i\}, \beta, \Omega | Y\right) \propto f\left(Y, \{u_i\}, \{\delta_i\} | \beta, \Omega\right) f(\beta) f(\Omega) \tag{29}$$

A.3.1 The conditional posterior for the latent data can be factored as:

$$f(\{u_i\}, \{\delta_i\} | \beta, \Omega, Y) = f(\{u_i\} | \{\delta_i\}, \beta) f(\{\delta_i\} | \beta, \Omega, Y) \tag{30}$$

The first term is:

$$f(\{u_i\} | \{\delta_i\}, \beta) = \Pi_{i=1}^{n} f(u_i | \delta_i, \beta) \tag{31}$$

where $f(u_i | \delta_i, \beta)$ is a normal with mean $x_i'\beta$ and variance 1, truncated below zero if δ_i is positive and above zero otherwise. The second term is:

$$\begin{aligned} f(\{\delta_i\} | \beta, \Omega_M, Y) &= \Pi_{i=1}^{n} f(\delta_i | \beta, \Omega, y_i) \\ f(\delta_i | \beta, \Omega_M, y_i) &= \Psi_{\delta_i | y_i} \end{aligned} \tag{32}$$

(which are calculated as in (10)). Therefore, the latent data can be conditionally generated by generating δ_i using $\Psi_{\delta | y_i}$ and u_i from its truncated normal.

A.3.2. The conditional distribution for Ω

In making a conditioning statement such as $f(\Omega | \beta, \{u_i\}, \{\delta_i\}, Y)$ then since the value of δ_i is known with probability one given u_i :

$$f(\Omega | \beta, \{u_i\}, \{\delta_i\}, Y) = f(\Omega | \beta, \{u_i\} Y) \tag{33}$$

Accordingly, the posterior for Ω is:

$$\begin{aligned} f(\Omega | \beta, \{u_i\}, Y) &\propto f(Y, \{u_i\} | \beta, \Omega_M) f(\Omega) \quad (34) \\ &= f(Y | \{u_i\} \beta, \Omega_M) f(\Omega) f(\{u_i\} | \beta, \Omega) \end{aligned}$$

where, only the sign of $\{u_i\}$ matters in the first term. Therefore:

$$\begin{aligned} f(Y | \{u_i\}, \beta, \Omega) &= f(Y | \{\delta_i\}, \beta, \Omega_M) \\ &= \Pi_{i=1}^{n} \Theta_{y_i | \delta_i} f(\Omega) \end{aligned} \tag{35}$$

and the second term is not dependent on Ω

$$f\left(\{u_i\}|\beta,\Omega_M\right)=f\left(\{u_i\}|\beta\right) \tag{36}$$

Therefore:

$$f\left(\Omega|\beta,\{u_i,\delta_i\},Y\right)\propto \Pi_{i=1}^{n}\Theta_{y_i|\delta_i}f\left(\Omega\right)$$

For this posterior a Metropolis Hasting step can be used for Ω.

A.3.3. The Posterior for β

$$\begin{aligned} f\left(\beta|\Omega,\{u_i\},Y\right) &\propto f\left(Y,\{u_i\}|\beta,\Omega\right)f\left(\beta\right) \\ &\propto f\left(Y|\{u_i\},\beta,\Omega\right)f\left(\{u_i\}|\beta,\Omega\right)f\left(\beta\right) \end{aligned} \tag{37}$$

The first term

$$f\left(Y|\{u_i\},\beta,\Omega\right)=f\left(Y|\{\delta_i\},\beta,\Omega\right)=f\left(\Omega\right)\Pi_{i=1}^{n}\Theta_{y_i|\delta_i} \tag{38}$$

whereas in the second

$$f\left(\{u_i\}|\beta,\Omega\right)=f\left(\{u_i\}|\beta\right) \tag{39}$$

is as in the standard normal linear model (with $\{u_i\}$ being the dependent variable). In model 1 $\Theta_{y_i|\delta_i}$ are not dependent on β. Therefore:

$$f\left(\beta|\Omega,\{u_i\},Y\right)\propto f\left(\{u_i\}|\beta,\Omega\right)f\left(\beta\right) \tag{40}$$

and the posterior for β would then be normally distributed. However, in the case of models 2 and 3.

$$f\left(Y|\{\delta_i\},\beta,\Omega\right)\propto f\left(\Omega\right)\Pi_{i=1}^{n}\Theta^{*}_{y_i|\delta_i}\Phi_i^{\rho} \tag{41}$$

and since Φ_i^{ρ} is dependent on on β this proportionality must be accounted for.

References

[1] Alberini, A., Boyle, K. and Welsh, M. 2003. Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty. *Journal of Environmental Economics and Management* **45**: 40-62.

[2] Arana, J.E. and Leon, C.J. 2005. Flexible Mixture Distribution Modeling of Dichotomous Choice Contingent Valuation With Heterogeneity, *Journal of Environmental Economics and Management*, **50(1)**: 170-188.

[3] Arrow K., Solow, R., Portney, P., Leamer, E., Radner, R. and Schuman, H. 1993. Report of the NOAA Panel on Contingent Valuation.

[4] Balcombe K., Bailey, A., Chalak, A. and Fraser, I. 2007. Bayesian Estimation of Willingness to Pay where respondents misreport their preferences. *Oxford Bulletin of Economics and Statistics* 69 (3) 413-437.

[5] Carson, R. T., Hanemann, W. M., Kopp, R.J., Krosnick, J.A., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Conaway, M. and Martin, K. 1998. Referendum Design and Contingent Valuation: The NOAA Panel's No-Vote Recommendation, *Review of Economics and Statistics* **80**: 484–487.

[6] Caudhill, S.B. and Groothuis, P.A. 2005. Modeling Hidden Alternatives in Random Utility Models: An Application to "Don't Know" Responses in Contingent Valuation. *Land Economics*, **81(3)**: 445-454.

[7] Caudill, S. B. and Mixon, F. G. 2005. Analysing discrete responses: a logit model based on misclassified data, *Oxford Bulletin of Economics and Statistics*, **67**: 105–113.

[8] Chambers, C.M. and Whitehead, J. C. 2003. A contingent Valuation Estimate of the Benefits of Wolves in Minnesota. *Environmental and Resource Economics*, **26**: 249-267.

[9] Champ P.A., Bishop, R.C., Brown T.C. and McCollum D.W. 1997. Using donation mechanisms to value non-use benefits from public goods. *Journal of Environmental Economics and Management*, **33**: 151-162.

[10] Gelfand, A. and Dey, D. 1994. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society Series* B, **56**: 501-504.

[11] Granberg, D. and Holmberg, S. 1991. Self-Reported Turnout and Voter Validation, *American Journal of Political Science*, **35 (2)**: 448-459.

[12] Groothuis P.A. and Whitehead J.C. 2002. Does don't know mean no? Analysis of don't know responses in dichotomous choice contingent valuation questions. *Applied Economics*, **34**: 1935-1940.

[13] Haener, M. K. and W. L. Adamowicz 1998. Analysis of "Don't Know" Responses to Referendum Contingent Valuation Questions, *Agricultural and Resource Economics Review*, **27**: 218–230.

[14] Hausman, J. A., Abrevaya, J. and Scott-Morton, F. M. 1998. Misclassification of the dependent variable in a discrete-response setting, *Journal of Econometrics*, **87**: 239–269.

[15] Kahneman D. and Tversky A. 2000. *Choices Values and Frames*, Cambridge University Press, New York.

[16] Koop G. 2003. *Bayesian Econometrics.* Wiley Publishers, West Sussex, England.

[17] Leece, D. 2000. Household Choice of Fixed Versus Floating Rate Debt: A Binomial Probit Model with Correction for Classification Error, *Oxford Bulletin of Economics and Statistics*, **62**: 61-82.

[18] Leon, R. and Leon, C.J. 2003. Single or Double Bounded Contingent Valuation? A Bayesian Test, *Scottish Journal of Political Economy,* **50 (2)**: 174-188.

[19] Li, C. and Mattsson, L. 1995. Discrete Choice Under Preference Uncertainty: an Improved Structural Model for Contingent Valuation, *Journal of Environmental Economics and Management*, **28**: 256-269.

[20] Loomis, J. and Ekstrand, E. 1998. Alternative Approaches for Incorporating Respondent Uncertainty when Estimating Willingness to Pay: The Case of the Mexican Spotted Owl, *Ecological Economics*, **27**: 29-41.

[21] Meyerhoff, J and Liebe, U. 2006. Protest beliefs in contingent valuation: Explaining their motivation, *Ecological Economics,* **57**: 583– 594

[22] Ready, R.C., Whitehead, J.C. and Bloomquist, G.C. 1995. Contingent valuation when respondents are ambivalent, *Journal of Environmental Economics Management*, **29**: 181–196.

[23] Samnaliev, M. Stevens, H. and Moore, T. 2006. A comparison of alternative certainty calibration techniques in contingent valuation, *Ecological Economics*, **57:** 507-519.

[24] Shaikh S.L., Sun L.G. and van Kooten, C. 2007. Treating respondent uncertainty in contingent valuation : A comparison of empirical treatments. *Ecological Economics*, (forthcoming).

[25] Svensson, M. 2006. Why so (Un)certain? Calibration in Contingent Valuation Using the Certainty Approach, Working Paper No. 6, Department of Business, Economics, Statistics and Informatics, Orebro University, Sweden.

[26] van Kooten, G.C., E. Krcmar and E.H. Bulte 2001. Preference Uncertainty in Nonmarket Valuation: A Fuzzy Approach, *American Journal of Agricultural Economics,* **83**: 487-500.

[27] Wang, H. 1997. Treatment of 'don't know' responses in contingent valuation sur-

vey: a random valuation model, *Journal of Environmental Economics and Management*, **32**: 219–232.

[28] Whitehead, J.C. and Cherry, T.L. 2006. Willingness to Pay for a Green Energy Program: A Comparison of Ex-Ante and Ex-Post Hypothetical Bias Mitigation Approaches, Memo, Department of Economics, Appapachian State University, Boone, NC.

Table 1. Marginal Likelihoods

	St. Cloud	Ely	FDP
Model 1	-189.5	159.02	-632.85
Model 2	-191.56	-160.78	-635.13
Model 3	-191.32	-163.63	-633.15
Model 1 with $\Theta_{1\|0} = \Theta_{0\|1} = 0$	-199.06	-164.73	-638.71
Model 1. Pooling DKs and NOs	-203.51	-165.31	-641.64
Model 1. Pooling DKs and YESs	-192.39	-164.78	-637.23

Table 2. WTP Function - Wolf Management Plan

	St Cloud		Ely	
Param	Mean	St dev	Mean	St dev
Intercept	00.17	2.93	-1.18	2.92
Price	-3.90	1.32	-1.72	1.03
Income	-0.15	0.11	0.02	0.09
Plan trip	3.40	1.99	4.80	2.26
Altruism	3.46	2.43	2.14	2.47
Bequest	2.46	2.38	3.27	2.48
Exist	0.88	2.51	3.77	2.26
Ethical	3.26	2.23	0.96	2.09
Education	1.08	0.61	0.10	0.44
Age	-0.07	0.08	-0.07	0.07
Gender	-3.47	1.9	-1.60	2.60
Child	0.44	1.56	-2.02	1.92
$\Theta_{1\|0}$	0.13	0.05	0.05	0.03
$\Theta_{0\|1}$	0.16	0.05	0.23	0.10
$\Theta_{\bullet\|0}$	0.13	0.05	0.16	0.04
$\Theta_{\bullet\|1}$	0.40	0.05	0.25	0.07

Table 3. WTP Function - FDP Data

Param	Mean	St dev
Intercept	3.53	1.96
Price	-.88	0.42
Income	0.75	0.53
Visits	-3.26	0.56
Size of Household	-0.34	0.51
State	0.25	0.84
Age	-0.54×10^{-1}	1.57×10^{-1}
Age^2	1.09×10^{-3}	1.30×10^{-3}
Urban	0.68	2.00
$\Theta_{1\|0}$	0.33	0.04
$\Theta_{0\|1}$	0.37	0.07
$\Theta_{\bullet\|0}$	0.10	0.02
$\Theta_{\bullet\|1}$	0.13	0.03

Table 4. Comparisons of WTP

		Previous Literature*	Bayes Probit -Ignoring DKs, No Misreport	Model 1
St-Cloud	Mean	21.49	20.34	43.97
	Median		20.32	40.92
	(Stdv)	(5.52)**	(4.04)	(18.61)
Ely	Mean	4.77	2.58	8.70
	Median		2.32	3.99
	(Stdv)	(2.52)**	(1.76)	(12.98)
FDP	Mean	4.32	3.26	3.99
	Median	.	3.22	1.12
	(Stdv)	(0.54)***	(0.66)	(4.11)

* From Chambers and Whitehead (2003) and Samneliev et al. (2006)

** These are standard errors rather than standard deviations

*** These are approximate standard errors computed from CIs

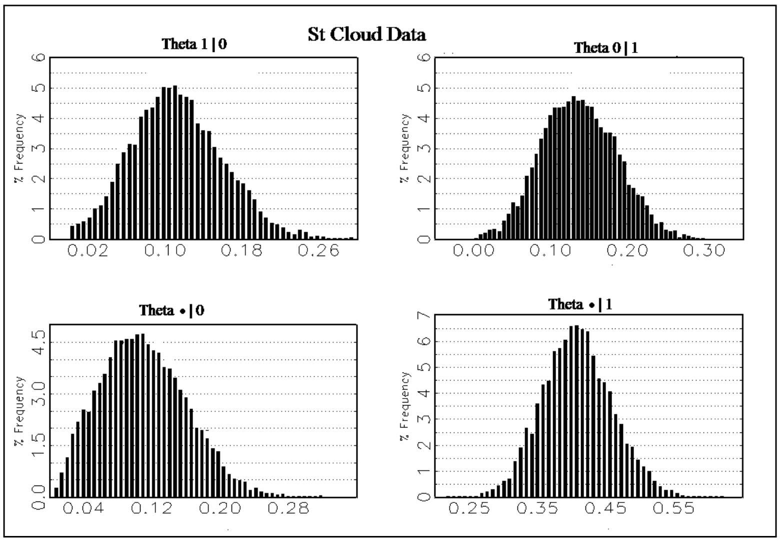

Figure 1: Distribution of Misreporting Parameters, St Cloud

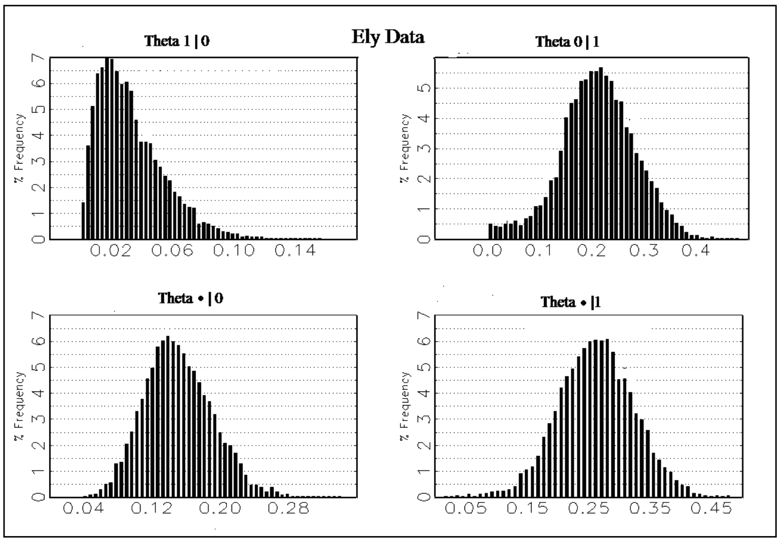

Figure 2: Distributions of Misreporting Parameters, Ely

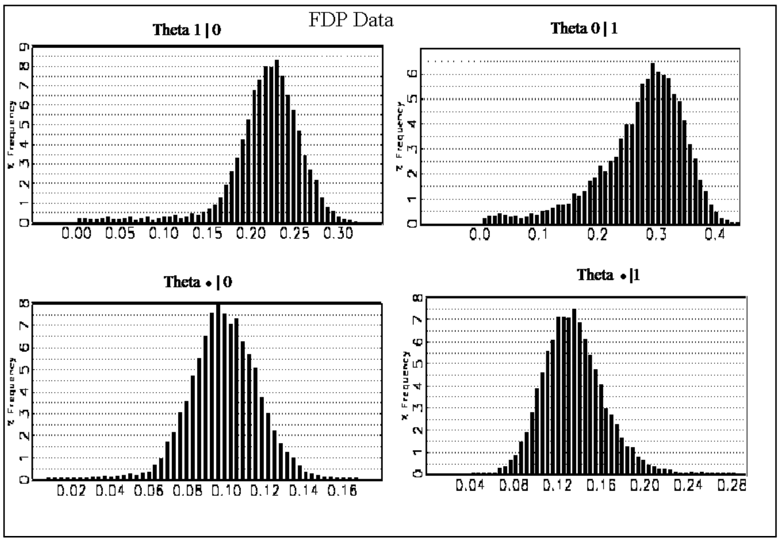

Figure 3: Distribution of Misreporting Parameters FDP Data.

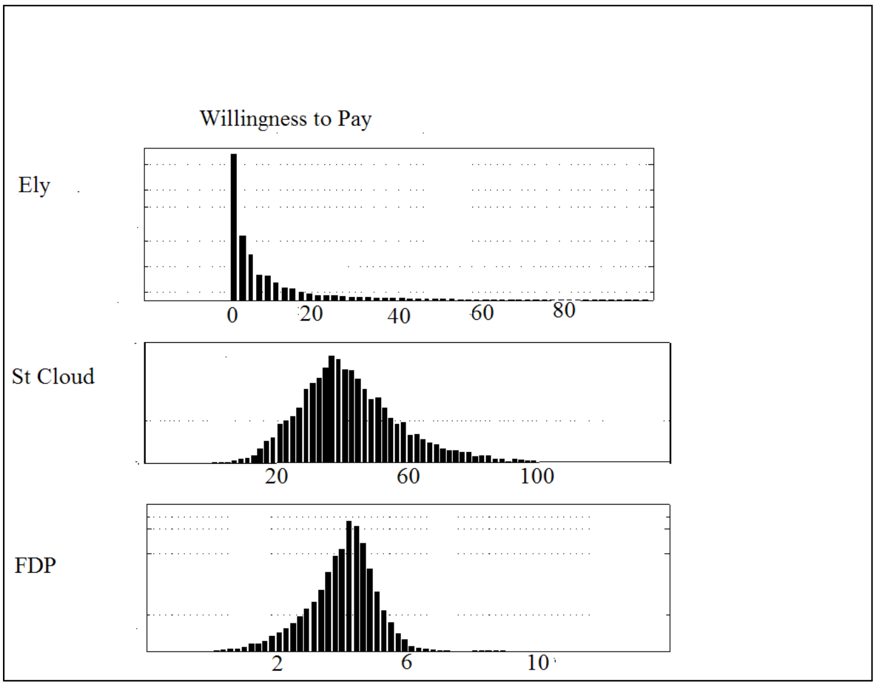

Figure 4: Distributions of Willingness to Pay

University of Kent

http://www.kent.ac.uk/kbs/research-information/index.htm